Books to inspire budding scientists p. 1190

The access-versus-evidence trade-off in drug approval p. 1205

A molten iron exoplanet orbiting a nearby star p. 1271

Science

$15
3 DECEMBER 2021
science.org

AAAS

FOREST DEATH

Climate change fuels debate over Germany's woodlands p. 1184

You heard the message.

We've told you before that NEB® offers a broad portfolio of reagents for purification, quantitation, detection, synthesis and manipulation of RNA. But did you know that these products are available from bench-scale to commercial-scale to enable both academic and industrial needs? Further, we provide these products at quality levels that support vaccine and diagnostic manufacturing. Experience improved performance and increased yields, enabled by our expertise in enzymology.

RNA purification: Extract up to 100 µg of high quality, total RNA from a variety of sample types with the Monarch® Total RNA Miniprep Kit. Monarch RNA Cleanup Kits can quickly and easily clean up and concentrate RNA in just minutes, with no carryover contamination.

RNA detection: Optimize your RT-qPCR across a variety of sample types with Luna®. High-concentration mixes and kits optimized for multiplexing enable sensitive detection of SARS-CoV-2. Simple, one-step solutions for LAMP and RT-LAMP are also available.

RNA-seq: NEBNext® kits are available for RNA library preparation, rRNA depletion and poly(A) mRNA isolation. Save time with streamlined workflows, reduced hands-on time and automation compatibility.

RNA synthesis: Synthesize high-quality RNA with reagents designed to simplify your workflow, including HiScribe™ IVT kits and capping reagents. GMP-grade* reagents are available for mRNA synthesis of therapeutics and vaccines.

Find more details on products available, request samples, and access helpful RNA-related resources at **www.neb.com/RNA2021**.

*"GMP-grade" is a branding term NEB uses to describe reagents manufactured at our Rowley, MA facility, where we utilize procedures and process controls to manufacture reagents in compliance with ISO 9001 and ISO 13485 quality management system standards. NEB does not manufacture or sell products known as Active Pharmaceutical Ingredients (APIs), nor do we manufacture products in compliance with all of the Current Good Manufacturing Practice regulations.

One or more of these products are covered by patents, trademarks and/or copyrights owned or controlled by New England Biolabs, Inc. For more information, please email us at gbd@neb.com. The use of these products may require you to obtain additional third party intellectual property rights for certain applications.

© Copyright 2021, New England Biolabs, Inc.; all rights reserved.

NEW ENGLAND BioLabs® Inc.

be INSPIRED
drive DISCOVERY
stay GENUINE

CONTENTS

3 DECEMBER 2021 • VOLUME 374 • ISSUE 6572

NEWS

IN BRIEF

1176 News at a glance

IN DEPTH

1178 Startling new variant raises urgent questions
Omicron's many mutations look troubling but understanding its danger will take time
By K. Kupferschmidt

1179 Where did 'weird' omicron come from?

1180 Poor trials of health steps are worse than none, scientists say
Others say small COVID-19 studies accumulate into a clear picture over time
By C. O'Grady
PODCAST

1181 China tightens its grip on Hong Kong universities
As dissent is stifled, a wave of departures at all levels is "mainlandizing" the higher education system *By D. Normile*

1182 Deep brain stimulation takes new aim at depression
Case studies spotlight personalized approaches to tweaking brain circuits
By K. Servick

FEATURES

1184 Forest fight
Germany invented "scientific" forestry. But a huge dieback triggered by climate change has ignited a fierce debate over how the nation should manage its trees *By G. Popkin*

INSIGHTS

BOOKS *ET AL.*

1190 Books for young scientists and engineers

PERSPECTIVES

1196 Short-circuiting respiration
Fumarate siphons electrons to keep metabolism running
By S. C. Baksh and L. W. S. Finley
REPORT p. 1227

1197 Reactivation of Cumbre Vieja volcano
A long-quiescent volcano's behavior requires rethinking about forecasting and hazards *By M.-A. Longpré*
PODCAST

1199 Directing carboxylic acid dehydrogenation
A palladium ligand can activate carbon-hydrogen bonds yet avoid product olefin reactions
By Y. Iwabuchi
REPORT p. 1281

1200 Programming a quantum phase of matter
The ability to measure long-range entanglement may enable robust quantum memory *By S. D. Bartlett*
RESEARCH ARTICLES pp. 1237 & 1242

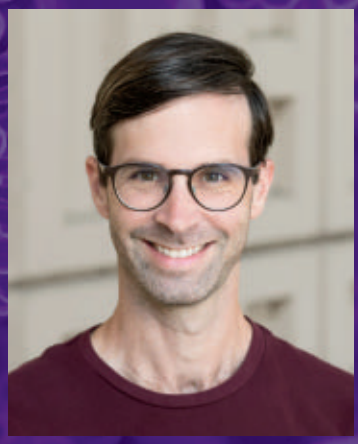
1201 Room-temperature mid-infrared detectors
Two independent groups designed nanoantennas for detecting mid-infrared light *By R. Gordon*
REPORTS pp. 1264 & 1268

1202 Dissecting the early COVID-19 cases in Wuhan
Elucidating the origin of the pandemic requires understanding of the Wuhan outbreak
By M. Worobey

CREDITS: (TOP TO BOTTOM) HANNAH AGOSTA; RAISSA MATHURA

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 you research essay today.

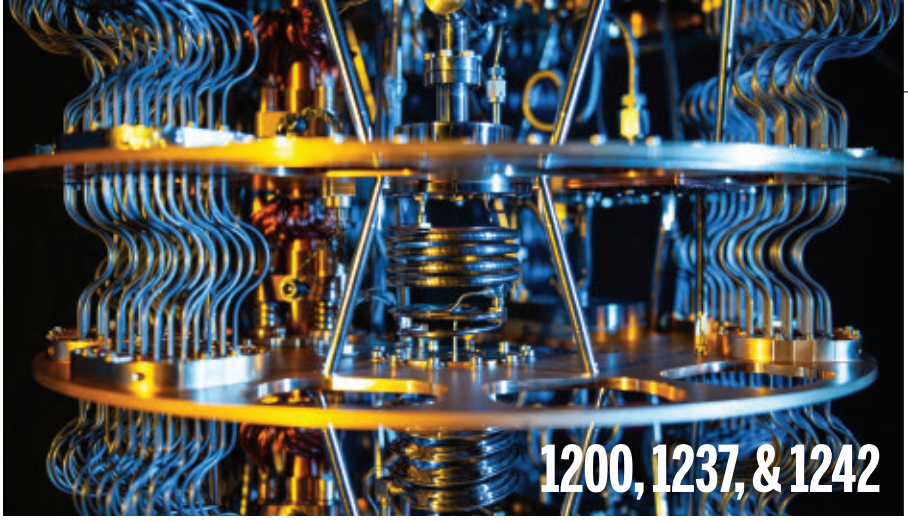


Eran Blacher, Ph.D.
2021 Winner

NOSTER | Science
MICROBIOME
PRIZE

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

Sponsored by Noster, Inc



Topologically ordered states are created in an array of Rydberg atoms in optical tweezers and on a superconducting chip, cooled in a dilution fridge.

POLICY FORUM

1205 Challenges in confirming drug effectiveness after early approval
Reform requires clarity about whether, when, and how meaningful postapproval trials are possible
By H. F. Lynch and C. T. Robertson

LETTERS

1208 Wildlife collisions put a dent in road safety
By F. Ascensão et al.

1208 Rethink roads through the Chaco Serrano forest
By Guadalupe Peralta et al.

1209 Sri Lanka's hasty agrochemical ban
By T. S. Priyadarshana

1209 Technical Comment abstracts

1209 Errata

RESEARCH

IN BRIEF

1210 From *Science* and other journals

REVIEW

1213 Developmental biology
The primitive streak and cellular principles of building an amniote body through gastrulation *G. Sheng* et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABG1727

RESEARCH ARTICLES

1214 Coronavirus
mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern *R. R. Goel* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABM0829

1215 Mechanical bonds
Active mechanisorption driven by pumping cassettes *L. Feng* et al.

1221 Human genomics
Genetic and functional evidence links a missense variant in *B4GALT1* to lower LDL and fibrinogen *M. E. Montasser* et al.

1227 Metabolism
Fumarate is a terminal electron acceptor in the mammalian electron transport chain *J. B. Spinelli* et al.
PERSPECTIVE p. 1196

Topological matter
1237 Realizing topologically ordered states on a quantum processor *K. J. Satzinger* et al.

1242 Probing topological spin liquids on a programmable quantum simulator *G. Semeghini* et al.
PERSPECTIVE p. 1200

1247 Plant science
Ground tissue circuitry regulates organ complexity in maize and *Setaria* *C. Ortiz-Ramírez* et al.

REPORTS

1252 Molecular biology
Error-prone, stress-induced 3′ flap–based Okazaki fragment maturation supports cell survival *H. Sun* et al.

1258 Organic chemistry
A biomimetic S_H2 cross-coupling mechanism for quaternary sp^3-carbon formation *W. Liu* et al.

Nanophotonics
1264 Continuous-wave frequency upconversion with a molecular optomechanical nanocavity *W. Chen* et al.

1268 Detecting mid-infrared light by molecular frequency upconversion in dual-wavelength nanoantennas *A. Xomalis* et al.
PERSPECTIVE p. 1201

1271 Exoplanets
GJ 367b: A dense, ultrashort-period sub-Earth planet transiting a nearby red dwarf star *K. W. F. Lam* et al.

1275 Carbon cycle
Strong Southern Ocean carbon uptake evident in airborne observations *M. C. Long* et al.

1281 Organic chemistry
Ligand-controlled divergent dehydrogenative reactions of carboxylic acids via C–H activation *Z. Wang* et al.
PERSPECTIVE p. 1199

DEPARTMENTS

1300 Editorial
Hubble's successor, at last
By Ellen Stofan

1290 Working Life
My weekly assignment
By Pijar Religia

ON THE COVER

In Germany, a changing climate has led to bark beetle outbreaks that have ravaged Norway spruce forests, such as this one in the Bavarian Forest National Park. The damage has helped fuel a fierce debate over the future of forest management. See page 1184. *Photo: Ingo Arndt/Minden Pictures*

Science Staff 1174
New Products 1286
Science Careers 1287

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.

PHOTO: GOOGLE QUANTUM AI

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*
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)**
Mark Smyth, *QIMR Berghofer*
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *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*

EDITORIAL

Hubble's successor, at last

Government investment in big, innovative science changes the world—it has allowed us to witness the collision of black holes, create a Global Positioning System that we can carry in our pockets, and unlock the human genome. Government investments in big research and infrastructure yield ever bigger breakthroughs. The James Webb Space Telescope (JWST), scheduled for liftoff on 22 December, represents a major international commitment. It will more than return its cost in extraordinary discoveries, new technologies, and inspiration to fuel a more diverse science and technology workforce. In short, like most big science infrastructure projects, it's a capital investment in a smarter and brighter future.

JWST has been in development since the mid-1990s when the astronomy and astrophysics community began thinking about the next great observatory to follow the Hubble Space Telescope. The new telescope would use a 6.6-m segmented primary mirror to detect near- and mid-infrared wavelengths of radiation. The large size of the mirror, and its even larger sunshade, require it to be tightly folded to fit inside its rocket fairing for launch. The configuration is so intricate that deployment will take about 1 month after a 30-day, 1.5-million-km journey to the second Sun–Earth Lagrange point. Science operations should begin about 4 months later, but the highly anticipated first images could come sooner.

JWST will rival Hubble in terms of revolutionary data. The Hubble telescope helped scientists confirm the existence of supermassive black holes, determine the age of the Universe (approximately 13.7 billion years), and realize that galaxies formed in the Universe much earlier than expected, among other discoveries. It also transmitted breathtaking images of an otherwise unimaginable universe. The JWST's instruments will study the structure of the first galaxies from 100 to 250 million years after the Big Bang, expand our understanding of the evolution of stars and protoplanetary systems, and probe questions such as the role of black holes in the early Universe. Some of the most exciting results will come from observing the atmospheres of planets around other stars and assessing their potential to harbor life. Given the large number of exoplanets identified over the past decade, JWST will have a rich set of targets to explore, giving us new insight into the ubiquity of environments that could harbor life beyond Earth. The Decadal Survey on Astronomy and Astrophysics released last month confirmed that the science that the JWST will do is still the highest priority.

Cost overruns and schedule slips of JWST have generated headlines, echoing Hubble. But nearly doubling in cost to $9.7 billion and launching years past its originally scheduled date overshadow innovative features behind the bugs: technologies that will enable future telescopes to better block out light and operate at ultracold temperatures. Developing, building, and testing the telescope pushed what we knew how to do. The hard lessons learned about program management will benefit future missions. There are still numerous risks—among them over 300 steps necessary to deploy the mirrors and sunshade in deep space, far beyond the reach of human hands to repair (at least any time soon). But big breakthroughs involve big risk. If you fail, you learn, and then try again.

> "Peering back to the origins of the Universe will reveal answers to questions we never thought to ask..."

The telescope will see first light in a very different time from when it was originally conceived—so much the better. The astronomy and astrophysics community has led the NASA community to remove bias from research proposal selection. When the Space Telescope Science Institute established double-blind review for observation time on Hubble, it more than doubled the number of successful first-time proposers and reduced the difference in success rate for women compared to men from over 6% to less than 1%. This will continue with JWST, creating a more level and diverse playing field and, as research shows, an even stronger and smarter scientific field.

The JWST will lift off on an Ariane 5 rocket from the European Space Agency's French Guiana launch site. The hopes of the astronomy and astrophysics community, and the eyes of the world, will lift with it. Peering back to the origins of the Universe will reveal answers to questions we never thought to ask, and studying planets around other stars will revolutionize the way we think about the evolution of planets, including our own. Big science doesn't just answer questions, it asks them—and sends humanity down new paths it could have never conceived.

–Ellen Stofan

Ellen Stofan is the under secretary for Science and Research, Smithsonian Institution, Washington, DC, USA, and is a member of the Space Telescope Institute Council. stofane@si.edu

PHOTO: JIM PRESTON/SMITHSONIAN NATIONAL AIR AND SPACE MUSEUM

10.1126/science.abn3947

NEWS

> "It's an unfortunate historic record, but it's a record."
>
> **Brown University physician Eli Adashi**, in *Nature Biotechnology*, arguing that medical information from the babies born from He Jiankui's human gene-editing experiments in China should be preserved as a record of his unethical conduct.

IN BRIEF

Edited by **Jeffrey Brainard**

COVID-19

Antiviral pill loses some luster

The drug company Merck & Co. last week reported its COVID-19 antiviral pill works less well than preliminary data suggested. In a press release, Merck announced that molnupiravir, which is designed to be taken within days of symptom onset, slices the risk of hospitalization and death by 30%, based on an analysis of the 1433 participants enrolled in its clinical trial. That's down from the 50% reduction Merck announced on 1 October based on an interim analysis of the first 762 participants. The difference between interim and final results appears largely due to lower risk of hospitalization or death in the placebo group, with the rate dropping from 14.1% in the interim analysis to 9.7% in the final one. Among participants getting the antiviral, the chance of hospitalization or death went from 7.3% to 6.8% when all the volunteers were studied. This week, a U.S. advisory committee narrowly recommended its use, suggesting restrictions on who should have access given concerns about side effects.

Vaccine network to expand

CORONAVIRUS | The world's largest COVID-19 vaccine testing network announced last week it will expand to support development of vaccines against other diseases that could cause a global health crisis. The Coalition for Epidemic Preparedness Innovations (CEPI), a partnership of public and private donors, last year launched a multicontinent laboratory network that aimed to hasten COVID-19 vaccine development by standardizing the reagents and protocols used in preclinical and early clinical testing of vaccines. Thirty developers of COVID-19 vaccines are using the network to assess whether their candidate can advance to late-stage clinical trials. Now, CEPI says the network will assess vaccines against other diseases that could cause an epidemic or pandemic, including Lassa fever, Middle East respiratory syndrome, Nipah, chikungunya, and Rift Valley fever. It will also prepare for "disease X"—the as-yet unknown pathogen that might one day cause the next pandemic. The expansion aims to draw in laboratories in regions where the network currently has no presence, including Africa, Oceania, and South America.

Germany IDs science priorities

LEADERSHIP | Germany's new coalition government—which will replace that of retiring Chancellor Angela Merkel—last week named a research minister and laid out policy priorities for science, including strengthening universities and boosting research spending to 3.5% of gross domestic product. In their agreement to form a government, the coalition—which includes the Social Democratic Party, the Greens, and the Free Democratic Party (FDP)—also outlines plans to protect biodiversity and reduce Germany's greenhouse gas emissions enough to meet a global 1.5°C warming limit. The pact does not mention the regulation of genetically modified crops, a point of disagreement between the Greens and FDP. The designated research and education minister is Bettina Stark-Watzinger, an FDP member who led a banking and finance research institute before being elected to parliament. Social Democrat Olaf Scholz is expected to become chancellor next week, after the parties approve the coalition agreement.

House science chair to retire

LEADERSHIP | U.S. Representative Eddie Bernice Johnson (D–TX), who championed increased federal support for research and expanding opportunities in science for women and people of color during her 3 decades in Washington, D.C., plans to retire from Congress at the end of 2022. The first Black woman to chair the committee, Johnson has teamed with the panel's top Republican, Representative Frank Lucas (OK), to stake out a centrist position on issues ranging from combating climate change and stamping out research misconduct to addressing China's growing scientific prowess.

Africa to join black hole viewing

ASTRONOMY | The Event Horizon Telescope (EHT), a global array of radio telescopes that in 2019 produced the first image of a black hole, will sharpen its vision with the addition in Namibia of the African Millimetre Telescope, the first such instrument on the continent. The project was approved this week after Radboud University in the Netherlands agreed to support the project with €1 million per year for operations and €1.9 million toward its expected €10 million construction costs. The team is negotiating with other institutions to secure the remaining funding. EHT achieves high

The design of an African telescope will be based on one on a peak in France.

PHOTO: IRAM

Floodwater filled St. Mark's Square in Venice, Italy, in November 2019 after the highest tide in 50 years.

PHOTO: FILIPPO MONTEFORTE/AFP/GETTY IMAGES

CLIMATE CHANGE

Barriers starve Venice's wetlands of sediments that sustain them

Flood barriers in Venice, Italy, may cause important salt marshes in its lagoon to drown by starving them of sediment they need to keep up with sea level rise, a study has found. To combat a growing incidence of major flooding, Venice spent billions of euros to build three barriers that can temporarily close the lagoon off from the Mediterranean Sea. But the barriers also block stormwater that moves sediment to the marshes, and after they began operating in October 2020, the rate of sedimentation of the marshes fell by 25%, Davide Tognin, a Ph.D. student at the University of Padua, and colleagues write in this week in *Nature Geoscience*. That's worrisome because the marshes provide habitats for wildlife, store carbon, and help clean nutrient pollution from the water. Adjusting when the barriers are deployed could allow more storm-carried sediment to reach the marshes, the authors say.

resolution by combining data from widely space radio dishes—from France to Hawaii and Greenland to Antarctica—into a virtual Earth-size dish. The 15-meter dish in Africa will help the project pursue its goal of capturing a movie of black hole activity. Namibian astronomers will gain opportunities to join the black hole project and carry out their own observations.

Geneticists curb use of 'race'

GENETICS | Human geneticists have mostly abandoned the word "race"—which refers to socially determined categories rather than biological groups—to describe populations in their papers, a study of a leading genetics journal has found. Examining 11,635 papers published from 1949 to 2018 by *The American Journal of Human Genetics,* researchers found that the use of "race" decreased from 22% of papers in the first decade to 5% in the last. During the same period, the journal's authors increasingly employed terms—"ethnicity," "ancestry," and location-based names such as "African," "Asian," and "European"—perceived as more useful for describing human biological variation. Such changes may reflect a growing consensus among the journal's authors that "race" is "a social category with biological consequences," the team writes this week in the journal. Acknowledging that earlier geneticists helped shape the racial categories still used today, the American Society of Human Genetics, which publishes the journal, this week announced a project to explore past injustices perpetrated through genetics.

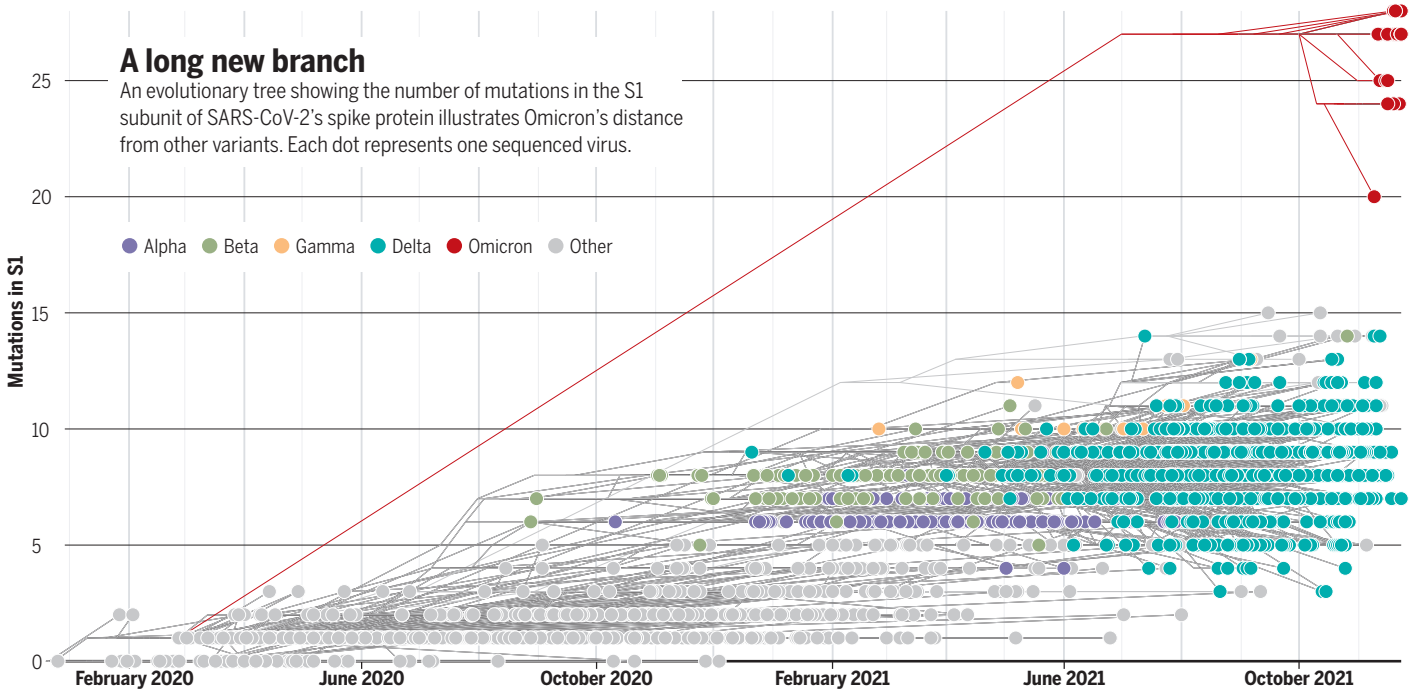
Einstein's goofs fetch $15 million

HISTORY OF SCIENCE | One person's trash is another's treasure. In 1914, Albert Einstein apparently meant to discard a sheaf of not-quite-right calculations he used to test his then-incomplete theory of gravity, the general theory of relativity. Last week, the documents sold for $15 million, Christie's auction house announced. They include 26 pages in Einstein's hand, 25 in that of Michele Besso—a Swiss-Italian engineer whom Einstein called "the best sounding board in Europe"—and three written by both. In those pages, the two struggle to explain why the axis of Mercury's elliptical orbit shifts forward on each circuit like the hand of a clock. In 1915, Einstein succeeded in using general relativity to explain how the Sun's gravity is responsible for the peculiarities of Mercury's orbit.

Warming trims Chile hydropower

CLIMATE CHANGE | A decline in water flow expected from Andes snowmelt, caused by warming global temperatures, has made a hydroelectric plant planned in Chile financially unsustainable, a U.S. utility told a bankruptcy court last month. The AES Corporation, sponsor of the Alto Maipo project, had based its viability on a 59-year record of water flow levels, Bloomberg Quint reported. But diminished flow in 2021 would have yielded only about half the electricity generation expected. On 17 November, AES asked the court for permission to restructure the project's finances so it can be completed and begin operations in 2022 using the lower flow.

IN DEPTH



COVID-19

Startling new variant raises urgent questions

Omicron's many mutations look troubling but understanding its danger will take time

By **Kai Kupferschmidt**

At 7:30 a.m. on 24 November, Kristian Andersen, an infectious disease researcher at Scripps Research, received an instant message on Slack: "This variant is completely insane." Molecular evolutionary biologist Andrew Rambaut of the University of Edinburgh was reacting to a set of new SARS-CoV-2 genome sequences shared on the global platform GISAID. Three came from samples collected in Botswana on 11 November that were sequenced by researchers there; one was picked up a week later in a traveler from South Africa to Hong Kong.

Andersen looked at the data and then replied: "Holy shit—that is quite something. The length of that branch …" A few minutes later he added: "Just had a look at the list of mutations—so nuts."

They were talking about what is now called Omicron, a new variant of concern, and the long branch Andersen noticed refers to its distance to every other known virus on SARS-CoV-2's evolutionary tree. The variant seemed to have picked up dozens of mutations, many of them known to be important in evading immunity or increasing transmissibility, with no intermediate sequences in the database of millions of viral genomes. On 23 November, after spotting the odd sequences in the GISAID database, Tom Peacock, a virologist at Imperial College London, had already posted his own verdict on GitHub: "This could be of real concern."

Now, once again, the world is watching as researchers work nights and weekends to learn what a new variant has in store for humanity. Is Omicron more infectious? More deadly? Is it better at reinfecting recovered people? How well does it evade vaccine-induced immunity? And where did it come from? Finding out will take time, warns Jeremy Farrar, head of the Wellcome Trust: "I'm afraid patience is crucial."

Researchers in South Africa were already on the trail of this new variant. Several teams were independently trying to figure out why cases were spiking in Gauteng, a northern province that includes Johannesburg and Pretoria. And a private lab called Lancet Laboratories had noticed that routine polymerase chain reaction (PCR) tests for SARS-CoV-2 were failing to detect a key target, the *S* gene, in many samples, a phenomenon previously seen with Alpha, another variant of concern. When Lancet sequenced eight of these viruses, it found out why: The genome was so heavily mutated that the test missed the gene.

Lancet shared the genomes with the Network for Genomics Surveillance in South Africa (NGS-SA), which called an urgent meeting on 23 November. "We were shocked by the number of mutations," says Tulio de Oliveira, a virologist at the University of KwaZulu-Natal and NGS-SA's principal investigator. After the meeting, de Oliveira says, he called South Africa's director general of health and "asked him to inform the minister and president that a potential new variant was emerging." The team sequenced another 100 randomly selected sequences from Gauteng in the next 24 hours. All showed the same pattern. After informing the government, de Oliveira and his colleagues presented their evidence at a press conference on the morning of 25 November. On 26 November, the World Health Organization (WHO) designated the virus a "variant of concern" and christened it Omicron.

GRAPHIC: NEXTSTRAIN.ORG. ADAPTED BY N. DESAI/*SCIENCE*

(Variant names follow the Greek alphabet, but WHO skipped the letters Nu and Xi, it said, "because Nu is too easily confounded with 'new' and Xi was not used because it is a common surname.")

One reason for concern about Omicron is that sequenced samples indicate it has rapidly replaced other variants in South Africa. But that picture might be skewed. For one, sequencing might have been focused on possible cases of the new variant in recent days, which could make it appear more frequent than it is. PCR data provide broader coverage and a less biased view, but there, too, samples with the *S* gene failure indicate a rapid rise of Omicron.

The rising frequency could still be due in part to chance. In San Diego, a series of superspreading events at a university resulted in an explosion of one particular strain of SARS-CoV-2 earlier this year, Andersen says: "It was thousands of cases and they were all the same virus." But the virus wasn't notably more infectious. South Africa has seen relatively few cases recently, so a series of superspreading events could have led to the rapid increase of Omicron. "I suspect that a lot of that signal is explained by that and I desperately hope so," Andersen says. Based on a comparison of different Omicron genomes, Andersen estimates the virus emerged sometime around late September or early October, which suggests it might be spreading more slowly than it appears to have.

The other reason to be concerned is Omicron's confusing genome. Its spike protein, which latches on to receptors on human cells, has 30 amino acid differences from that of the original virus from Wuhan, China. In addition, amino acids have disappeared in three places and new ones appeared in one place. Many of the changes are around the receptor-binding domain, the part of the protein that makes contact with the human cell. "That is very troubling," Farrar says. Structural biology mapping in 2020 showed some of these changes made the virus bind to the receptor much better.

It's hard to tell how infectious a virus is based on mutations alone, says Aris Katzourakis, an evolutionary biologist at the University of Oxford. "But if we were looking out for mutations that do affect transmissibility, it's got all of them," he says.

The sequence also suggests the virus could excel at evading human antibodies, says Jesse Bloom, an evolutionary biologist at the Fred Hutchinson Cancer Research Center. The human immune system produces a host of different antibodies that can neutralize SARS-CoV-2, but many of the most important ones fall into three categories that each target a slightly different site on the spike protein of the virus, simply called 1, 2, and 3. A mutation called E484K has long been worrying because it changes the shape of the site that class 2 antibodies recognize, making them less potent. Omicron carries a mutation called E484A in this site and similar changes in the sites for the other two classes of antibodies.

"It seems clear to me that our antibodies and the spike protein are sort of in a genetic arms race," says virologist Paul Bieniasz of the Rockefeller University. "A week ago, I was more confident that antibodies were going to come out on top, and I'm a little less confident now." Bloom thinks people who

Where did 'weird' Omicron come from?

The unsettling new SARS-CoV-2 variant did not develop out of one of the earlier variants of concern, such as Alpha or Delta, but instead appears to have evolved in parallel—and in the dark. Omicron is so different from the millions of SARS-CoV-2 genomes that have been shared publicly that pinpointing its closest relative is difficult, says Emma Hodcroft, a virologist at the University of Bern. It likely diverged early from other strains, she suggests. "I would say it goes back to mid-2020."

That raises the question of where Omicron's predecessors lurked for more than a year. Scientists see essentially three possible explanations: The virus could have circulated and evolved in a population with little surveillance and sequencing. It could have gestated in a chronically infected COVID-19 patient. Or it might have evolved in a nonhuman species, from which it spilled back into humans recently.

Christian Drosten, a virologist at Charité University Hospital in Berlin, favors the first possibility. "I assume this evolved not in South Africa, where a lot of sequencing is going on, but somewhere else in southern Africa during the winter wave," he says. "There were a lot of infections going on for a long time and for this kind of virus to evolve you really need a huge evolutionary pressure."

But Andrew Rambaut of the University of Edinburgh can't see how the virus could have stayed hidden in a group of people for so long. "I'm not sure there's really anywhere in the world that is isolated enough for this sort of virus to transmit for that length of time without it emerging in various places," he says.

Instead, Rambaut and others prefer option two. When Alpha was discovered in late 2020, that variant also appeared to have acquired numerous mutations all at once, leading researchers to postulate it came from a chronically infected COVID-19 patient. Sequencing of SARS-CoV-2 samples from some chronically infected patients bolstered the idea by showing their viruses developed some of the same mutations seen in Alpha and other variants.

Drosten says experience with chronic infections of influenza and other viruses in immunosuppressed patients argues against this hypothesis. Variants able to elude the immune system do develop in such people, but they come with a host of other changes that make them less able to transmit from person to person. "These viruses have very low fitness out in the real world," he says. That's because the mutations allowing a virus to survive in one individual over time may be very different from those needed to best spread from one person to the next.

Jessica Metcalfe, an evolutionary biologist at the Institute for Advanced Study in Berlin, isn't so sure. "I think one reason that this virus has done so well is that better binding to ACE2 [its receptor on human cells] helps for both within-host spread and between-host spread," she says. Still, Metcalfe agrees with Drosten that Omicron likely circulated and evolved in a hidden population.

Some scientists suspect the virus was hidden in animals rather than people. "The genome is just so weird," says Kristian Andersen of Scripps Research, pointing to its medley of mutations, many of which have not been seen before in other variants.

Aris Katzourakis, an evolutionary biologist at the University of Oxford, is skeptical of a nonhuman origin. "I'd start worrying about animal reservoirs more if we were succeeding in suppressing the virus, and then I could see it as somewhere it might hide," he says.

It is also unclear whether the poor vaccination rates in Africa played a role in Omicron's emergence, as public health officials concerned about vaccine equity have suggested. "The idea that if we had vaccinated more in Africa, we wouldn't have this: I'd like that to be true, but we have literally no way of knowing," Katzourakis says. —*K.K.*

recovered from COVID-19 or were vaccinated are unlikely to completely lose their ability to neutralize the virus. "But I would expect, based on this particular combination of mutations, that the drop in neutralization is larger than for all the other major variants."

Experiments in the laboratory will have to show whether he is right. Alex Sigal, an infectious disease researcher at the Africa Health Research Institute, says he received swabs with Omicron on 24 November and has started to grow the virus. Producing enough of it to test against sera from vaccinated and recovered individuals will take a week or two, he says. Other researchers will test viruses genetically engineered to carry just the spike protein of Omicron, a process that is faster than growing the variant itself, but a bit further removed from what happens in real life.

As such studies take place, it's crucial to closely monitor the pandemic, Farrar says. "Do you see cases increasing not just in South Africa, but the broader South African region?" The same applies to the rest of the world. The virus has already been picked up in more than a dozen countries, many of which are not high on the list of the most connected places to Johannesburg. "My feeling is that Omicron has likely spread to many more places where it will soon be detected," says Oxford epidemiologist Moritz Kraemer. Epidemiologists will also watch for changes in disease severity—how many people are hospitalized and die. All that will take time.

In the meantime, the European Union, the United States, and many other countries have restricted travel to and from southern Africa in a bid to protect themselves. Travel restrictions are unlikely to stop the variant, Farrar says, but they can buy some time. "The question is what you then do with the time."

But the restrictions could, ironically, hamper science. "It is really difficult to get the needed reagents in a plane as the last 10 planes that I tried to get our reagents on got canceled," de Oliveira says. And the economic and social cost could be a disincentive to report new variants: "I've heard through the grapevine that countries didn't push sequences out very quickly [in the past] because they were worried about travel bans," says Emma Hodcroft, a virologist at the University of Bern. "This is the opposite of what we want."

Such considerations did not stop the South African researchers, de Oliveira says. "We do risk a massive backlash in case [Omicron] does not cause a massive wave of infection and can be controlled," he wrote in a message. "But this is a risk that I am comfortable to live with as the pandemic has caused so many deaths and suffering. [Our] hope is that our early identification will help the world." ■

Passengers ride a subway in Denmark, where an inconclusive trial on mask wearing took place.

PHOTO: RITZAU SCANPIX DENMARK/REUTERS

COVID-19

Poor trials of health steps are worse than none, scientists say

Others say small COVID-19 studies accumulate into a clear picture over time

By **Cathleen O'Grady**

In November 2020, researchers in Denmark published inconclusive results of a randomized trial to probe whether mask requirements protect against COVID-19. Although nearly 5000 people took part in the DANMASK study, it was too small, and its time frame too short, to answer the question. News reports, however, described the study results as questioning the efficacy of masks, and a June analysis noted that it fed into antimasking misinformation campaigns.

Now, another group argues such small, weak trials of public health measures can do more harm than good. Writing in *Trials* last month, the group argues such studies waste funding and time, and can give a dangerous appearance of certainty. Much research "does not lead anywhere that is useful," says co-author Noah Haber, an independent study design specialist. "It's noise at best, and it's misinformation at worst, because it looks like information." Other researchers, however, think any evidence is better than none.

Henning Bundgaard, the lead DANMASK author, could not be reached for comment. However, in August in the *Annals of Internal Medicine*, Bundgaard and co-authors defended their sample size and methods, and said the study has been misinterpreted.

There's little doubt that small, weak studies have proliferated during the pandemic, which raised pressure to test interventions fast. A February *Nature Communications* analysis of 686 clinical COVID-19 studies found they had poorer quality methods than a matched group of trials with similar study designs. A *BMJ* review last month of 72 studies found hand washing and wearing masks reduced risk of SARS-CoV-2 infection, but it also said most of the studies had moderate to serious weaknesses. And a study led by Haber, posted as a preprint in January, found that only one of 36 studies of COVID-19 policies met four criteria that would make results useful for policymakers, such as tracking outcomes for long enough that policy measures had time to influence local infection rates.

Atle Fretheim, a health services researcher at the Norwegian Institute of Public Health, thinks Haber and his team are wrong to dismiss small studies completely. Because relatively few people in a country or region are infected at any one time, impractically large sample sizes would be needed to study public health measures with certainty, he says.

In March, Fretheim argued in *Trials* that studies like DANMASK can contribute to evidence when taken together. "If face masks work, more or less, in all the 20 small, stupid trials that have been done, then we

can probably agree that they seem to work."

Small studies can also play a crucial role in preventing waste, adds Manoj Lalu, an evidence quality researcher at the Ottawa Hospital Research Institute. They allow researchers to road-test crucial questions such as how to recruit participants or keep consistency across study sites before embarking on a study involving thousands of people.

But hoping that someone will pick up the baton after a small trial is not a strong justification, says McGill University bioethicist Jonathan Kimmelman. Testing involving humans is only ethical when it can inform an important decision—like whether to launch a large trial following a pilot study, or whether to roll out vaccination after a large trial. Running a small study without the intent to take it further comes with a range of harms, he says: "You're wasting volunteers' time, there's an opportunity cost, you're putting out shit information that can only be misinterpreted by people who don't know how to interpret it, and it's only going to take time away from the people who do know how to interpret it."

Misinformation is a critical problem, Haber and his co-authors say. Any small, inconclusive study opens the door to misinterpretation, says Emily Smith, a George Washington University epidemiologist and a co-author of Haber's. "We live in the real world."

There's also the danger that policymakers will base decisions on single flawed studies, says Kelly Cobey, a metascientist at the University of Ottawa Heart Institute. Syntheses that pull together large bodies of evidence, like meta-analyses, aren't the answer either: "If the quality of the individual studies is poor, the quality of the synthesis is poor."

Haber and his co-authors agree that health measures such as wearing masks or school closures are difficult to study. That's because of the complexities of studying human behavior—such as whether people wear masks routinely and correctly—and because small effects require huge populations, Haber says. A high-quality trial on mask wearing is possible, he says, noting a massive trial in Bangladesh, released as a preprint in May and published in *Science* today. The trial, which had a large sample and careful design, reported small benefits. But it, too, has been criticized for its analysis methods and for overstating its findings.

The pandemic has shown both "the best and the worst" of science, Kimmelman says. Vaccine trials and the large-scale Recovery trial of treatments have been "unbelievable," he says. But questions about mask wearing or the ideal timing of booster shots haven't had the same resources or attention: "We've discovered how piss-poor we are at answering certain kinds of questions that are of paramount importance for public health." ■

ACADEMIC FREEDOM

China tightens its grip on Hong Kong universities

As dissent is stifled, a wave of departures at all levels is "mainlandizing" the higher education system

By **Dennis Normile**

When prodemocracy demonstrations erupted in Hong Kong in 2019, its publicly funded universities were hotbeds of unrest. A year later, five university presidents signed a statement supporting a law that would make such protests difficult if not impossible. Two did not sign the document, which endorsed the new National Security Law Beijing was about to pass: Wei Shyy of the Hong Kong University of Science and Technology (HKUST) and Kuo Way of the City University of Hong Kong. Both universities have now announced their presidents will step down.

Kuo's departure is timed to the end of his third 5-year contract in 2023, by which time he'll be 72, dampening speculations about political pressure. But Shyy, 66, will resign in October 2022, a year before his contract ends. "Everybody is wondering what's his rationale for stepping down a year early," says Carsten Holz, a development economist at HKUST.

> **"Everybody is wondering what the future holds for the city and the universities."**
>
> A Hong Kong academic

Many other Hong Kong academics are leaving, too. The city's universities, among the best in Asia and long known for their academic freedom, are seeing a wave of resignations and departures as China tightens its grip in what one academic calls the "mainlandizing" of the universities. Three deans at the University of Hong Kong (HKU), including Gabriel Leung, who played an important role in Hong Kong's successful fight against COVID-19, recently announced they are leaving early or not renewing contracts.

Although numbers are hard to come by, sources within the city's universities—few of whom want to be named—say significant numbers of native Hong Kong scholars and Westerners at lower levels are leaving as well, troubled by declining academic freedom, attacks on those considered disloyal by pro-Beijing local media, "patriotic education" being imposed on their school-age children, and investigations under the National Security Law, which passed in June 2020. Whereas Hong Kong's Basic Law, in effect since the territory passed from the United Kingdom to China in 1997, guarantees its universities autonomy and academic freedom, the new law called on the government to strengthen its guidance and supervision of the universities. It gave authorities new powers to punish vaguely defined offenses such as "secession" and "subversion."

Some caution against reading too much into the upheaval. "There will be adjustments among Hong Kong universities," given the changing political climate, says astronomer Sun Kwok, a former dean of science at HKU. But he believes China will recognize that HKU's "diverse, multicultural academic staff" is a strength that will help advance the overall level of scholarship in China and the region.

Others are gloomy. "It's sad for the universities and sad for the city," says one tenured Western academic. "Three years ago I thought I would be here until I retire," he says, but now he is keeping an eye out for opportunities to move. Another Western scholar predicts the trend will turn the city's universities into "ciphers of the mainland China institutions."

Critics say the shake-up at HKU, the territory's oldest and most prestigious university, has been driven by Xiang Zhang, a physicist who became president in July 2018. Zhang is the first HKU head to have been born and educated through the undergraduate level in China. He later worked at the mechanical engineering department at the University of California, Berkeley, and is a naturalized U.S. citizen. HKU recently announced that his 5-year contract will be renewed in 2023.

Among top-level researchers parting ways with HKU are the former dean of science, U.K. ecologist Matthew Evans, who stepped down in October to take a position at United Arab Emirates University. He declined to comment for this article. The dean of social sciences, psychologist William

A student passes a "Lennon wall" filled with protest messages—including one that says "Vow to express freedom"—at the University of Hong Kong in the summer of 2020. The university has since removed all protest signs.

Hayward, will leave at the end of his current contract in August 2022. Hayward, a New Zealand native, did not respond to an email seeking comment.

On 21 November, Gabriel Leung, HKU's highly respected dean of medicine, announced he will step down next summer as well, even though his contract runs until July 2023. Since his appointment in 2013, Leung oversaw the faculty's rise from 36th to 20th place in the *Times Higher Education*'s "clinical and health" rankings. A native Hong Konger, he was also one of the architects of Hong Kong's "suppress and lift" strategy that alternately toughened and relaxed restrictions on social activities to help keep COVID-19 infections at a very low level.

Leung, 49, will become executive director for charities and community of the Hong Kong Jockey Club, overseeing the club's Charities Trust, one of the world's biggest philanthropic organizations. (Leung has a long-standing interest in social issues and is also a governor of the United Kingdom's Wellcome Trust, the world's largest philanthropic funder of medical research.) Instead of answering questions from *Science*, Leung shared the email he sent to HKU staff and students announcing his departure "with a heavy heart." He explained he would dedicate "the next stage of my work life addressing the ultimate challenge of the 'unethical epidemic of inequalities.'"

Asked to respond to rumors that deans were told in advance their contracts would not be renewed, an HKU spokesperson wrote in an email to *Science* that "Deans vacate their posts upon contract completion" but can continue as members of the faculty. The spokesperson added: "It is our standard practice to conduct international searches for faculty deans."

So far, few Westerners have been appointed to newly opened positions, however. Instead, they are attracting interest from mainland-born Chinese scholars, some currently in the United States where they feel increasingly unwelcome due to the U.S. Department of Justice's China Initiative. Although both HKUST and City University have said they will search internationally for new presidents as well, critics believe candidates will have to be acceptable to the Chinese government—and that their political views will permeate the universities they will lead.

The changing political winds are influencing career decisions by junior academics as well, Holz and others say. Holz, who has been outspoken about threats to academic freedom in Hong Kong, says he has felt "no pressure because of what I've written. There's been no interfering in my teaching or research." But he acknowledges that not all of his colleagues have been so fortunate. Two HKUST scholars left Hong Kong abruptly earlier this year after pro-Beijing newspapers attacked them for comments made at conferences, he says. (The two could not be reached for confirmation.) These incidents leave "an underlying anger and tenseness in people," Holz says.

As another Hong Kong academic puts it, "Everybody is wondering what the future holds for the city and the universities." ■

NEUROSCIENCE

Deep brain stimulation takes new aim at depression

Case studies spotlight personalized approaches to tweaking brain circuits

By **Kelly Servick**

For patients whose depression resists treatment with drugs and electroconvulsive therapy, surgically implanted wires that stimulate the brain might bring relief. But in recent years, two randomized, controlled trials of this approach, known as deep brain stimulation (DBS), were halted after underwhelming results in interim analyses. "It was like the air was let out of the room," Sameer Sheth, a neurosurgeon at Baylor College of Medicine, says of those results. "It was a big let-down."

Now, researchers are testing more sophisticated, personalized DBS techniques they hope will yield stronger results. The tests to date have involved just one or a few patients, far from proof of effectiveness. But researchers hope they'll inform larger studies that finally cement the effectiveness of DBS in depression. "With all these irons in the fire … we will hopefully build up enough understanding and evidence," says Sheth, an author of a case study published last week.

DBS is already approved in the United States to treat epilepsy, obsessive compulsive disorder, and movement disorders such as Parkinson's disease. The hope is that it could also shift patterns of abnormal activity in neural circuits that may drive depression symptoms. Early studies without control groups yielded promising results, but two randomized, controlled trials, sponsored by the medical device companies Medtronic and St. Jude Medical, Inc. (later acquired by Abbott Laboratories) did not show significant benefits after several months of DBS, teams reported in 2015 and 2017.

Long-term follow-up of participants has revived some optimism. For example, many people in the 30-participant Medtronic trial

PHOTO: LIAU CHUNG-REN/ZUMAPRESS/NEWSCOM

improved over 1 year or more—beyond the timeline of the initial study, says Stanford University psychiatrist Mahendra Bhati, a co-investigator. In October, he and colleagues published a follow-up study of eight trial patients, most of whom continue to use their implant about 10 years later. About one-half have had at least a 50% improvement over their pretreatment score on a depression scale.

Many researchers say that because depression can manifest differently in different brains, personalizing the treatment could make it more effective. Several teams are now exploring more precise stimulation approaches informed by individual brain anatomy and neural recordings.

Helen Mayberg, a neurologist at the Icahn School of Medicine at Mount Sinai who has pioneered the use of DBS in depression, has led one effort. She and colleagues previously identified an area near the front of the brain called the subcallosal cingulate (SCC) as a hub in a network linked to the negative mood component of depression. The halted randomized St. Jude Medical trial, known as BROADEN, targeted the SCC. Mayberg suspects subtle differences in where surgeons placed the electrodes account for much of the variation in patient outcomes in that trial. So her group and others now use an MRI approach called tractography to visualize locations of nerve bundles and target electrodes more precisely.

Mayberg is also tracking how the brain responds to this more precise stimulation, hoping to identify neural responses that predict whether and when a person's symptoms will improve. Last month in *Translational Psychiatry*, she and colleagues describe a change in activity in the SCC during brain stimulation in the operating room that correlates with treatment response in the week after surgery in seven studied patients.

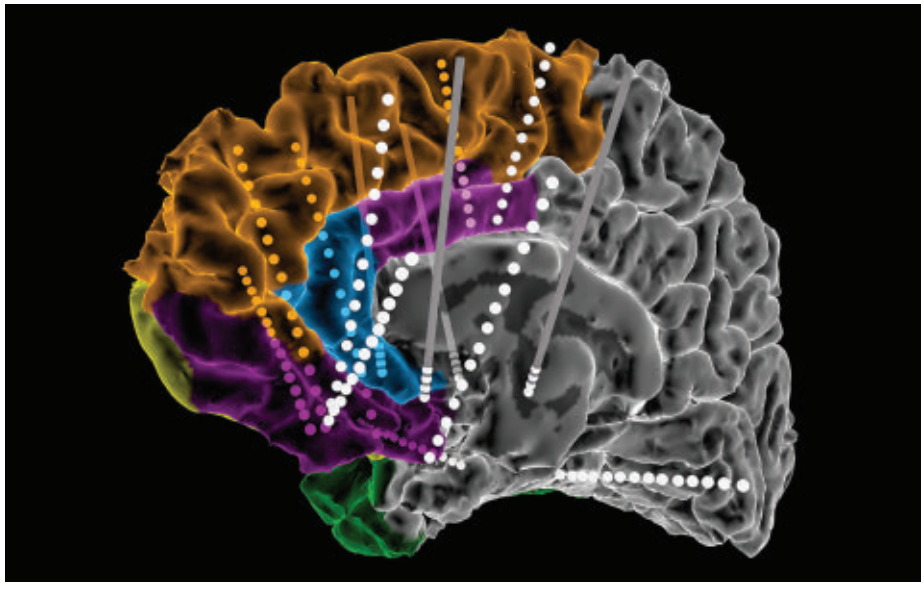
Neuropsychologist Isidoor Bergfeld and colleagues at Amsterdam University Medical Center are testing tractography-guided DBS implants in another part of the brain implicated in depression known as the superolateral branch of the medial forebrain bundle. Bergfeld aims to recruit 24 patients and expects results in 2 to 3 years. He hopes data from his group's trial and from a similar ongoing study in Germany can combine to support regulatory approval in the European Union for depression.

Other teams have more extensive personalization in mind. Psychiatrist Katherine Scangos of the University of California, San Francisco, and colleagues are using an individual's neural data to decide which target region to stimulate and exactly when to deliver the stimulation. Before inserting DBS electrodes, the team implants a separate set of electrodes that can record from and stimulate tissue across diverse regions of the brain. Surgeons already rely on the approach, known as stereotactic electroencephalography (sEEG), to identify the source of seizures before epilepsy surgery. Now, Scangos's team has used it to stimulate mood-related regions and select a target where stimulation reliably relieves depression symptoms.

For the first patient, of a planned 12 in an ongoing trial, the ideal stimulation target seemed to be the ventral capsule/ventral striatum (VC/VS), already an established target in DBS depression trials. The team also found that activity in a second site in the patient's brain, the amygdala, strongly predicted periods of severe symptoms. So the patient's implant—a device called NeuroPace, which is approved to detect and stop seizures—was programmed to stimulate the VC/VS only in the presence of those amygdala activity patterns. "Every time that train sort of veers off the tracks, we push it back on," Scangos says. That first patient has felt significant relief from her symptoms for more than 1 year since the surgery, Scangos and her colleagues reported in October in *Nature Medicine*.

Other researchers are waiting for more evidence that this automated, "closed-loop" approach to the timing of stimulation is necessary. "With severe depression, it's not like you're alternating between really depressed and totally normal … over moments or hours," Sheth says. His team is using sEEG electrodes to find the best settings for continuous DBS in each patient. During a 10-day hospital stay, the researchers use sEEG readouts to identify patterns of brain activity associated with a positive mood and track how those readouts change in response to DBS stimulation from electrodes in both the SCC and VC/VS. A computer algorithm then identifies the ideal frequency, amplitude, and other settings for improving mood.

Data from the first participant in an ongoing trial, published on 22 November in *Biological Psychiatry*, show his symptoms improved in the first few months after surgery—and then steadily worsened during a blinded discontinuation phase, when stimulation at some of the sites was reduced gradually at a time point unknown to the patient. (Researchers then turned the implant back to full power.)



Recording sites on implanted electrodes (dots) defined optimal settings for a stimulation device (gray lines).

IMAGE: RAISSA MATHURA

Mayberg cautions that to be used widely in the clinic, DBS will have to be cost effective and simple enough "that your average, competent neurosurgeon and depression expert can … implement it."

Sheth says that implanting the set of sEEG electrodes to guide treatment is likely too invasive and labor-intensive to be feasible for every patient, though it might offer a chance to adjust DBS in patients whose symptoms aren't responding. He also hopes researchers can develop less invasive methods to inform stimulation settings.

"We're the first to admit that time will tell whether this approach or variants of it can successfully be applied," Sheth says. But the variety of ambitious DBS approaches being tested is a good sign, he says. "I'm just happy that groups are still fighting to help all these patients." ■

FOREST FIGHT

Germany invented "scientific" forestry. But a huge dieback triggered by climate change has ignited a fierce debate over how the nation should manage its trees

By **Gabriel Popkin**, *in Schwenda, Germany*

Dead conifers in a forest near Königshain, Germany. Policies and tradition often discourage leaving dead trees in place.

PHOTO: FLORIAN GAERTNER/PHOTOTHEK VIA GETTY IMAGES

Last summer, Friederike and Jörg von Beyme stood on a bramble-covered, Sun-blasted slope outside this small town in eastern Germany. Just 4 years ago, the hillside, part of a nearly 500-hectare forest the couple bought in 2002, was green and shady, covered in tall, neatly arranged Norway spruce trees the couple planned to cut and sell.

During January 2018, however, a powerful storm felled many of the trees. Then, over the next 3 years, a record drought hit Germany and much of Central Europe, stressing the spruces that still stood. The back-to-back disasters enabled bark-boring beetles that had been munching on dead trees to jump to drought-weakened ones. Beetle populations exploded. In just 3 weeks, towering spruces that had seemed healthy were dead.

The von Beymes salvaged what they could, rushing to log and sell the dead and diseased trees. But thousands of other forest owners did the same, causing the timber market to collapse. The couple's piles of logs were worth less than what it had cost to cut and stack them. Now, they don't expect to earn a profit from logging spruces for 20 years. "We have a big forest now with big problems," Jörg von Beyme says.

The von Beymes are far from alone. Since 2018, more than 300,000 hectares of Germany's trees—more than 2.5% of the country's total forest area—have died because of beetles and drought fueled by a warming climate. The massive dieback has shocked the public. And it has raised hard questions about how a country renowned for inventing "scientific" forestry more than 3 centuries ago should manage forests so they can continue to produce wood and protect ecosystems in the face of destablizing climate shifts.

Everyone agrees that new approaches are needed, but no one, it seems, can agree on what those should be. Some advocates want Germany's government and forest industry to stop promoting the widespread planting of commercially valuable trees such as Norway spruces, and instead encourage landowners to allow forests to regenerate on their own. Others say that to meet economic, environmental, and climate goals, Germany must double down on tree planting—but using more resilient varieties, including some barely known in Germany today.

The stakes are high: Germany's forest products sector generates some €170 billion annually and employs more than 1.1 million people. If its wood supplies dwindle, pressure could grow to log forests elsewhere around the world. Declining forests could also imperil efforts to replace building materials that generate huge emissions of green-

Forest researchers Pierre Ibisch (left) and Jeanette Blumröder check a data logger in a pine forest that burned in 2018 and is now being allowed to naturally regenerate.

house gases, such as concrete and steel, with potentially climate-friendlier wood.

The disagreements are often fierce, with the opposing sides trading insults in the media and even holding competing forest summits. "The intensity of the debate," says ecologist Christopher Reyer of the Potsdam Institute for Climate Impact Research, "is surprising for everybody."

IT'S NO EXAGGERATION to say modern industrial forestry was invented in Germany. In the early 1700s, mining official Hans Carl von Carlowitz, who lived not far from where the von Beymes live today, became alarmed by devastating timber shortages caused by demand from mining and smelting. In response, he penned a 1713 treatise proposing that forests be managed sustainably. Wood harvests should be limited to what the land could produce, von Carlowitz wrote, and trees should be assiduously replanted to ensure a future supply. (Of course, Indigenous people around the world had been applying similar ideas for millennia.)

German forests started to recover as landowners adopted the approach. And Germany's scientific approach to forestry—planting fast-growing species in neat rows, perfectly spaced for maximum timber production—became an international model. After World War II, with Germany in ruins and Allied nations demanding shipments of timber for reparations, foresters doubled down on von Carlowitz's vision. Areas where deciduous trees such as beech and oak would have grown naturally were planted in monocultures of fast-growing evergreen spruce and pine. The trees were so essential to Germany's economy that they became known as the *brotbaums* or "bread trees."

For decades, the program looked like a stunning success: Even as West Germany experienced its *Wirtschaftswunder* (economic miracle) starting in the 1950s, timber stocks increased. By the early 21st century, the total amount of wood in German forests had reached a volume probably not seen since the Middle Ages. Today, nearly one-third of Germany is forested.

But many of those forests are far from natural. Norway spruce alone, for example, accounts for one-quarter of the trees—and more than half the timber harvest. The shallow-rooted species naturally grows in high latitudes or on cold mountainsides. But in Germany, as well as in the Czech Republic, Austria, and elsewhere, foresters planted it throughout low-lying and far warmer regions. The monocultures nurtured only a fraction of the biodiversity found in native deciduous forests, but as long as there was enough rain and temperatures stayed cool enough, the spruces thrived.

In recent years, however, global warming has begun to disrupt long-standing weather patterns, delivering extremes these forests hadn't experienced. The unprecedented drought that began in 2018 was especially devastating for Germany's spruce plantations. The combination of extreme summer heat and a lack of precipitation set off a deadly chain reaction. Soils dried out to a depth of 2 meters. The water-starved spruces could no longer produce the tough gooey resin that helps protect them against insects, leaving them open to attack by bark beetles, which normally feed on dead or dying trees. Beetle populations swelled—one adult can produce hundreds of offspring in a season—and overwhelmed whole forests, turning them from green to ghostly gray.

The destruction hit hardest in Germany, the Czech Republic, and Austria. Forests in France, Poland, Switzerland, Slovakia, and Italy also took hits. Across Central Europe, some 300 million cubic meters of wood was damaged, according to forest scientist Andreas Bolte of the Thünen Institute, the German government's forest research agency.

FOR MANY forest owners, and for ordinary Germans for whom a wander in the woods is a favorite pastime and an essential part of their cultural identity, the dead trees delivered a huge shock. In a 2019 speech, former Chancellor Angela Merkel soberly recounted the "very, very large forest damage" that had affected "thousands of forest owners." The dismay has helped fuel an intense political and scientific struggle over the future of Germany's forests.

PHOTO: LENA MUCHA

All sides agree the recent die-off highlights the climate change threat. "It's kind of an early warning, … a signal of what may still come," says forest researcher Gert-Jan Nabuurs of Wageningen University & Research. The future, he says, "is worrying."

Most also agree that existing monocultures, so important to European forestry's past, cannot ensure its future. "It's a clear signal to the wood industry that you have to change the utilization from Norway spruce to other species," Bolte says.

The consensus breaks down, however, when it comes to solutions. For some, the dieback offers a rare chance to dramatically shift forest policy toward a more hands-off approach. Allowing devastated forests to naturally regrow, the thinking goes, could revitalize ecosystems and start to reverse centuries of biodiversity decline.

One leading proponent of this view is Peter Wohlleben, a prominent author and forester. In books and media appearances, he describes natural forests as interconnected, cooperative communities. And he argues that Germany's vaunted scientific forestry, with its single-minded focus on maximizing timber production, disrupted those communities, creating simplified forests that are highly vulnerable to climate extremes.

Wohlleben and his allies are calling for a wholesale rethinking of plantations. "It's always better to let nature do the job," he says. "I don't know any place on Earth where a planted forest is better than a native forest."

Pierre Ibisch and Jeanette Blumröder, biologists at the Eberswalde University for Sustainable Development, agree. In August, as bursts of rain and gloom alternated with intense sunshine, they visited a fire-scarred research plot 1 hour's drive from Berlin that they believe could help prove the point.

Just a few years ago, the plot—part of a forest owned by the small town of Treuenbrietzen—was covered by Scotch pines, a common plantation species in regions with sandy soils. In the hot, dry summer of 2018, however, fires torched some 400 hectares of the pine forest, closing highways and forcing hundreds of people to flee their homes; smoke even reached Berlin. In the past, such large fires were almost unheard of in mild Central Europe.

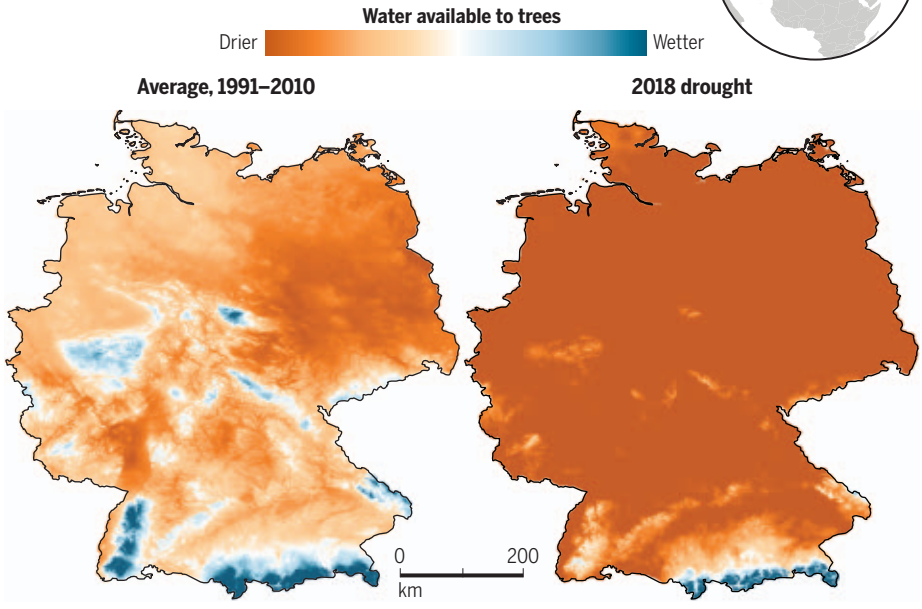
In this plot, charred trees were removed, replaced by newly planted pines. But the drought, which continued through 2020, killed many of the puny seedlings, Blumröder pointed out as she surveyed the site. And even the survivors were struggling to keep up with fast-growing poplar saplings, some already 3 meters tall, that had sprouted on their own. The poplars' vigor indicates that replanting is not necessary, Blumröder and Ibisch argue. "The problem is, foresters don't wait," Ibisch says. "They always say they think in long-term scales. But when calamity happens … they panic."

In some other burned plots, Ibisch and Blumröder persuaded Treuenbrietzen's forester to deviate from usual practices. On one tract, he left charred trunks standing and didn't replant, allowing forest succession to proceed on its own—a rare practice. In others, he cleared some of the snags and planted rows of oaks—which many researchers believe could be more resilient to future climate change—instead of pines.

In preliminary results, the new approaches are producing promising outcomes. In areas where some or all burned trees were left standing, for example, Ibisch and Blumröder have found more plant, fungus, and insect

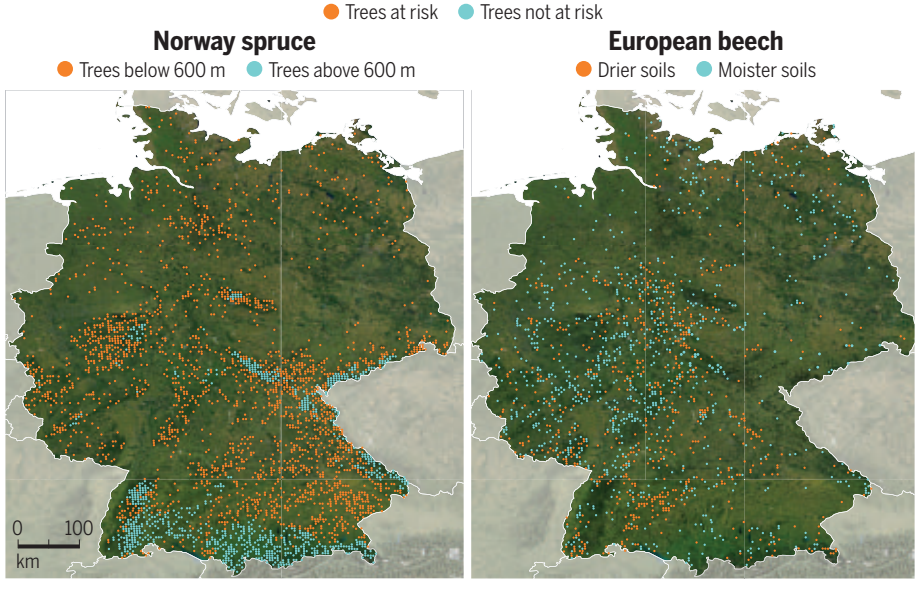
Dry spell

A record 3-year drought that began in 2018 (right) set off a cascade of tree stress, fires, and insect attacks that killed more than 2.5% of Germany's forests. The destruction highlights the threat posed by climate change, researchers say.



Stress maps

A rapidly shifting climate has made many of Germany's most important trees vulnerable to various threats, projections show. Droughts, which are predicted to become more severe, are expected to make Norway spruce growing in lower, warmer areas vulnerable to bark beetle attacks (left). A drier climate also threatens European beech trees growing in soils with less capacity to store water (right).



GRAPHIC: K. FRANKLIN/SCIENCE

Loggers remove spruce trees killed by drought and insects near Drübeck, Germany. Critics of such active management say forests should be left to rebound on their own.

species than in cleared tracts. Soil temperatures in the uncleared tracts are lower on hot days, and winds calmer, helping the soil retain moisture. Moss is beginning to cover the ground where fallen trees have started to rot, preventing erosion and stimulating the growth of underground soil fungal networks. The lesson for Germany's foresters, Blumröder believes, is that they should "step back, let the system do [its thing] first, and then learn from it."

In Harz National Park, which sits in mountains straddling the former border of East and West Germany, ecologist Gunter Karste with the Harz National Park Authority is also bucking tradition. Here, waves of bark beetles have killed more than 10,000 hectares of spruce stands. But research published by Karste and colleagues persuaded park managers to let the dead snags stand and hold off on replanting. Today, the lifeless gray, spire-like trunks are everywhere, surrounded by tangles of fallen trees, their airborne root systems still clinging futilely to soil. People now call the tracts the *Harzer Silberwald*, or Harz Silver Forest, Karste says.

Less than 3% of Germany's forests are currently managed like this, as strict nature preserves, but such practices could soon become more common. The German government has a goal to increase the figure to 5%, thanks in part to the ecological benefits Karste and others have documented. Although the dead trees "look awful the first 5 years," Karste says, what regrows is far more diverse and resilient than a plantation. Although still largely spruce, which thrives on cold mountainsides, the trees vary far more in size and age than do those in uniform, planted stands. That creates a greater variety of niches for wildlife, Karste notes. In the understory, wildflowers bloom and bees buzz; blueberries, mountain ash, birch, and other shrubs and small trees thrive. Meanwhile, owls, bats, and other species roost in dead trunk cavities. Karste says research suggests that "when you don't leave the dead trees, you lose 40% of the biology."

The more diverse, naturally regenerating forest will also likely cope better with future drought and pests, he says, because trees of different ages respond differently to such stresses, making it more likely that some will survive. If the park had simply cleared and replanted, he says, "then in 60 years you would again have a forest that's as interesting for the bark beetle as for the spruce forester."

THE IDEA OF LEAVING forests alone alarms other researchers. They argue the climate is changing so quickly that, without human help, even many native trees won't survive in places where they've long thrived.

"We have beeches dying now, we have maples dying … and pines that were considered pretty drought tolerant," says Henrik Hartmann, a plant scientist at the Max Planck Institute for Biogeochemistry. "It's not a spruce problem. It's a general forest problem." Recent modeling suggests more than half of Europe's forests are now vulnerable to insects, storms, fires, or a combination of these threats, Hartmann and colleagues reported earlier this year in *Nature Communications*.

To reduce the risks, some experts argue forest owners need to strategically plant new, more resilient tree varieties. Hints about strong candidates could come from a 250-hectare arboretum founded in the late 1800s in Wuppertal, a hilly town in western Germany. Here, collectors planted some 200 tree species from all over the world. More than 100 of those species are still growing, offering a rare opportunity to assess how the mature trees are handling climate change.

This fall, Leonore Gärtner, the state forester who now manages the area, strolled with her dog through a stand featuring some North American natives—Alaskan red cedar, incense cedar, and western hemlock—each with a number painted on the trunk. It looked more like the Olympic Peninsula of Washington state than Germany. But Gärtner was excited because the trees were thriving, even after 3 years of drought. "It's amazing," she said. "The trees are looking good, very healthy."

Gärtner believes the stand indicates foresters would do well by planting diverse mixes of commercially valuable species, increasing

PHOTO: LENA MUCHA

the likelihood that at least some will survive to harvest age in a changing climate.

Others are exploring variations on this approach. For example, Nabuurs is co-leading a project that will plant native trees that haven't been heavily used in forestry, such as linden and sweet chestnut, at 11 sites across Europe and assess their resilience to climate shifts. Hartmann, meanwhile, urges researchers to exploit the genetic diversity hidden within European tree species. Pines, for example, grow across much of the continent, and trees from hotter, drier areas—such as southern Europe—might have already evolved resistance to conditions forecast for Germany and other more northerly nations.

Hartmann cautions against immediately replanting dead forests with trees that have grown well in the past, instead urging foresters to first consult climate models that predict which tree species might fare best in the future. "We should not just blindly start reforesting sites that have been disturbed," he says. "We could, by doing this, create the next disaster."

WIDELY IMPLEMENTING new forestry techniques will require changes in government policy and buy-in from foresters and landowners. Germany's agriculture ministry has already met the dieback with an unprecedented aid program, showering forest owners with €1.5 billion to help them remove dead trees and replant. Those receiving funds must plant a mix of species, the ministry has said, though owners not taking funds can still plant monocultures. And for the first time, the government has made funds available to forest owners who want to allow their woods to regenerate naturally.

Last week, Germany's newly elected government went further, saying it intends to amend federal law to increase native forests, end logging in publicly owned old-growth beech stands, and promote other policies advocated by environmentalists.

The next step is largely up to the 2 million or so private landowners—individuals, families, and firms—who own about half the country's forests, and the cities and states that own most of the rest. And whereas environmentalists want more forests managed primarily for ecological values rather than timber, most forest owners, private or public, aim to make money from logging.

The von Beymes, for one, aren't keen on the hands-off approach. They see their denuded hillside, now thick with blackberries and grasses, not as a flourishing ecosystem, but a weedy, unprofitable mess. "That, to me, is not a forest," Jörg von Beyme says.

Most sawmills are designed for evergreen conifers and continue to demand them, he notes. That means that for now it is nearly impossible to sell species that come in naturally, such as poplars and birches, and even some new planted varieties that might do well in the future climate. The von Beymes also note that the commercially valuable deciduous trees they are growing in some forests—including oaks and beeches—can take 140 to 160 years to mature, compared with a mere 60 to 80 years for spruce. Moreover, they add, climate research indicates the cold- and moisture-loving beech "has no future" as a dominant species in their area.

That's why the von Beymes have planted some of their land with Douglas fir, a fast-growing conifer from North America. German foresters have been planting the species for nearly 2 centuries, but it is now gaining popularity because it's thought to be especially drought- and pest-resistant. Jörg von Beyme, for example, points to data from the Helmholtz Centre for Environmental Research suggesting Douglas fir can tolerate drier soils than spruce.

But some are skeptical of the tree's long-term future here. It's native to the rain-soaked Pacific Northwest, they note, a far cry from increasingly dry Central Europe. And mature Douglas firs planted decades ago at Burgholz are losing their needles, Gärtner says; some have even been attacked by bark beetles.

Bark beetles (bottom) carve galleries in trees (top) and kill many. Infestations are spreading in Germany because of drought and warmer temperatures.

PHOTOS: (TOP TO BOTTOM) LENA MUCHA; NIGEL CATTLIN/FLPA/MINDEN PICTURES

THE VON BEYMES won't know for decades whether the bet they've placed on their Douglas firs will pay off. In the meantime, the debate over Germany's forests continues to simmer. Earlier this year, Wohlleben organized a summit called Waldsterben 2.0 (Forest Death 2.0), at which scientists, activists, and officials from Germany's Green Party largely endorsed natural regeneration and criticized government officials for propping up the plantation system. Wohlleben says scientists from the government forest ministry declined to attend, but a ministry spokesperson says they never received invitations. The ministry held its own summit, where it announced new incentives for forest owners and a plan to compensate forest owners for using their forests to absorb and store carbon.

Some observers lament that the debate has become so polarized and are urging a middle path. "We don't have perfect solutions anymore," Reyer says. It is time to "stop pointing fingers at each other because it's not leading anywhere," Hartmann adds. Trees will still need to be planted, many argue, but more forested land should be left to nature.

One thing is clear: Germans will need to adapt to forests very different from the ones they've known. "This is disturbing for people," Hartmann says. "The forest of the future will not look like the one where I was walking with my grandpa." ■

Reporting for this story was supported, in part, by an Arthur F. Burns Fellowship.

INSIGHTS

BOOKS *et al.*

AAAS/SUBARU SB&F PRIZES FOR EXCELLENCE IN SCIENCE BOOKS

Books for young scientists and engineers

A timely vaccine explainer, a sumptuous feast of favorite foods, a handy guide to the night sky, and a number of charming titles featuring charismatic creatures beloved by children (think chickens, bears, and butterflies) are among this year's finalists for the AAAS/Subaru SB&F Prizes for Excellence in Science Books, awards sponsored by Subaru of America, Inc., and the American Association for the Advancement of Science (AAAS, the publisher of *Science*) that seek to encourage the publication of high-quality science titles for all ages. Read on for reviews of the finalists written by staff of the *Science* family of journals. **—Valerie Thompson**

MIDDLE GRADES SCIENCE BOOK

The Great Bear Rescue

Reviewed by **Sacha Vignieri**[1]

There are a great many species threatened with extinction today, and a challenge adults face is how to convey messages about these species to younger generations without passing along a sense of hopelessness. Author Sandra Markle takes on this challenge in *The Great Bear Rescue*, which describes efforts taken over the past few decades to save a small population of brown bears living in the Gobi Desert.

The Gobi Desert is one of the harshest regions on Earth, yet the bears that live there have adapted to survive in this unlikely environment, and the Mongolian people have long revered them. In *The Great Bear Rescue*, Markle describes some of the extreme challenges faced by the bears in a way that is clear and yet not overwhelming to young readers. These include the need to find water, a rare resource in the desert, and the native tubers that make up most of their diet. A newer, and especially dangerous, threat is the presence of illegal mining. Her prose is paired with compelling photographs of the bears, their environment, and the people working to save them.

There are only a handful of Gobi bears left—between 22 and 40, only eight of which are females. Such a dire situation may seem like a lost cause, but Markle's focus on researchers' efforts to understand these animals and local people's efforts to protect them leaves the reader feeling relatively optimistic.

I came away with the impression that Gobi bears now rely on the local people, who maintain feeding stations and are restoring essential plant resources, to survive. It made me sad to think that this resilient species might no longer be capable of sustaining itself. It was heartening, however, to see how local people's efforts to save and protect the bears connect the two populations in a way that is likely to improve the bears' chances of survival in our changing world.

The Great Bear Rescue: Saving the Gobi Bears, *Sandra Markle*, Millbrook Press, 2020, 40 pp.

The How and Wow of the Human Body

Reviewed by **Hilary Stewart**[2]

Full of vibrant illustrations, unexpected interviews with all your favorite organs, and answers to the most important questions (how many bathtubs could one person fill with spit

ILLUSTRATION: HANNAH AGOSTA

in a year?), *The How and Wow of the Human Body* takes readers on a journey through our various organs and appendages, starting with the head and concluding, appropriately, with the rear end.

Authors Mindy Thomas and Guy Raz both have a background in podcasting, and it shows in the way information is presented in the book. Dialogues, funny fictional ads, graphics, comics, and even biographies of different body parts can all be found, while QR codes at the end direct readers to the authors' favorite episodes of their *Wow in the World* podcast. There are also lots of fun challenges and examples of human record breakers dotted throughout. The authors' aim is to answer questions about how the human body works and to provide a barrage of fun and weird facts with which to wow one's friends.

At times, the book moves a little too quickly. I would have appreciated a more comprehensive overview of the body and less of a focus on the sensationally gross. But overall, it achieves what it sets out to do: inspire curiosity and imagination around the way our bodies work. What it lacks in depth it certainly makes up for in humor, a willingness to engage with irreverent questions ("Is it okay to eat my boogers?"), and its offering of helpful flowcharts (when trying to decide whether to pick at something on your skin, the answer is almost always no).

The How and Wow of the Human Body, *Mindy Thomas and Guy Raz, Illustrated by Jack Teagle*, Clarion Books, 2021, 192 pp.

There's No Ham in Hamburgers

Reviewed by **Trista Wagoner**[3]

There's No Ham in Hamburgers is organized around favorite American foods: hamburgers, French fries, pizza, ice cream, hot dogs, chicken fingers, peanut butter, cookies, chocolate, and breakfast cereals. Each chapter includes the origin of the food under consideration, as well as descriptions of how it is made, information about its nutritional value, and a recipe.

Scientific concepts appear in highlighted text boxes and include an explanation for why chips become stale and how milk chocolate requires overcoming the tendency for oil and water to separate. At least one recipe also doubles as an easy experiment, inviting readers to measure the temperature of ice before and after making ice cream. Careful readers will catch the author's observation that most of our favorite foods are unhealthy at the quantities at which we regularly consume them, but this qualification is easy to overlook.

A more culturally sensitive perspective would have improved the book. Zachman writes, for example, that the Aztecs "believed in human sacrifice" but fails to effectively contextualize this statement and gives short shrift to the experience of Indigenous societies, maintaining that Spanish explorers "discovered" potatoes, tomatoes, and other crops long cultivated by Native peoples.

Despite these shortcomings, readers of all ages are likely to learn something new about food. (For example, I learned that ancient Egyptians built incubators for chicken eggs and that mustard may have been the world's first condiment.)

There's No Ham in Hamburgers: Facts and Folklore About Our Favorite Foods, *Kim Zachman, Illustrated by Peter Donnelly*, Running Press Kids, 2021, 144 pp.

A Shot in the Arm!

Reviewed by **Seth Thomas Scanlon**[4]

As more children become eligible for COVID-19 vaccines, resources that help explain the "what" and "why" of vaccination to young minds are urgently needed. Don Brown's informative and inviting graphic novel *A Shot in the Arm!* arrives at just the right time.

Narrated by Lady Mary Wortley Montagu, who championed inoculation in 18th-century Britain, the book recounts efforts to understand and combat diseases such as smallpox, rabies, anthrax, cholera, and polio. Contributions by scientists including Antony van Leeuwenhoek, Edward Jenner, Robert Koch, Louis Pasteur, and Jonas Salk are recounted in an approachable and inspiring way. The author also does an excellent job describing the societal and political responses to vaccination, with special attention paid to antivaccine movements. Brown ends his tale with a timely and engaging depiction of the COVID-19 pandemic and the global effort to quickly develop vaccines.

Mini-primers on the science behind vaccines interspersed throughout the text may leave budding scientists wanting more. Additionally, the detailed timeline at the end is much appreciated, but the accompanying bibliography should have included more age-appropriate resources. But these are minor criticisms. Overall, I found the book to be informative, clear-eyed, and enjoyable.

J. N. Hall wrote in *Science* in 1894 that "[t]here is, indeed, a certain element of danger in vaccination, as in every other thing of established value, but it is strange that in the face of such evidence…there are still found those who deny the value of the most beneficent discovery ever made by man" (*1*). When it feels like the same battles are being fought again and again, take heart in this book's core message: that despite all the roadblocks, progress continues to be made in protecting humanity against infectious diseases, thanks to the curiosity, creativity, and passion of scientists around the globe.

REFERENCES AND NOTES

1. J. N. Hall, *Science* **23**, 72 (1894).

A Shot in the Arm!, *Don Brown*, Amulet Books, 2021, 144 pp.

CHILDREN'S SCIENCE PICTURE BOOK

Mimic Makers

Reviewed by **Kelly Servick**[5]

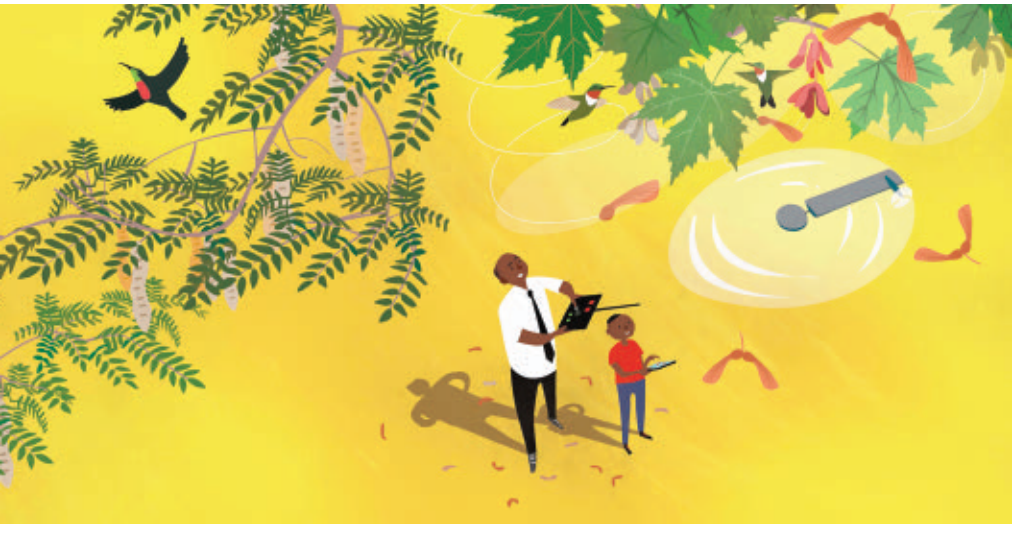
In the wrinkles of a leaf and along the fin of a whale, inspiration is hiding. That is the message in Kristen Nordstrom's *Mimic Makers*, a joyful parade of scientists and engineers who have looked to nature to solve problems.

Nordstrom describes how the tapered shape of a kingfisher's bill pointed engineer Eiji Nakatsu to a quieter, more efficient design for the Shinkansen bullet train and how the ridged back of a Namibian beetle led designer Pak Kitae to create a device that collects drinking water from morning

dew. Each example is vibrantly rendered in Paul Boston's illustrations, many of which combine close-ups of an invention with sweeping landscapes to suggest its impact on human life.

The book does not shy away from technical detail—young readers can flip to the glossary to unpack terms such as "photosynthesis" and "condense"—and some of the featured inventions may be new to adult readers, too. But Nordstrom does more than list off cool inventions. She depicts a diverse global community of inventors, all mentioned by name, who draw ideas from the native species and phenomena around them.

Mimic Makers is also a call to action: A final appendix offers advice for journaling, sketching, prototyping, and engineering nature-inspired products. A portion of the author's profits goes to the Biomimicry Youth Design Challenge, a competition that asks groups of middle- and high-schoolers to come up with nature-inspired solutions to the climate crisis. "It doesn't matter where you live—in a city, on a farm, or by the sea," affirms the book's inclusive final message. "Nature's secrets are waiting for you."



Mimic Makers: Biomimicry Inventors Inspired by Nature, *Kristen Nordstrom, Illustrated by Paul Boston*, Charlesbridge, 2021, 48 pp.

Chickenology

Reviewed by **Lauren Kmec**[6]

Which came first, the chicken or the egg? *Chickenology: The Ultimate Encyclopedia* provides a definitive answer to this age-old question (the egg!). Brimming with delightful illustrations, *Chickenology* begins with chicken "basics," including an overview of their life cycle and communication methods, as well as how to distinguish a hen from a rooster. Next comes an informative section on physical characteristics and how they vary from one breed to another. Plunging into a discussion of chicken intelligence, readers will discover that despite their small brains, these animals have an aptitude for logical reasoning—they can learn to count and, surprisingly, to play the piano.

Chickenology dedicates several pages to the egg: structure, size, and even culinary applications. Straying outside the avian world, a striking illustration depicts the eggs of different creatures, from delicate, bead-like moth eggs to dogfish egg cases adorned with cascading tendrils. The book also explores humanity's relationship to chickens throughout history and offers a primer on backyard chicken rearing. The latter includes tips for constructing a proper henhouse and transforming chicken droppings into fertilizer. *Chickenology* closes with a few short profiles of chickens from around the world. I was astonished to learn that the Ayam Cemani breed from Indonesia is entirely black inside and out—even its skeleton is black!

At more than 70 pages, *Chickenology* is longer than a traditional picture book, but the text and imagery are nicely balanced. It also contains some technical language, so parents should plan to read along with younger children. More-experienced readers will be able to tackle the book on their own and, because it is packed with engaging details, will be sure to return to it time and time again.

Chickenology: The Ultimate Encyclopedia, *Barbara Sandri and Francesco Giubbilini, Illustrated by Camilla Pintonato*, Princeton Architectural Press, 2021, 80 pp.

Abby Invents the Foldibot

Reviewed by **Jennifer Sills**[7]

We have machines that wash and dry our clothes, but we are on our own when it comes to the pesky task that follows: folding. Abby and Miko, the protagonists of *Abby Invents the Foldibot*, tackle this challenge with gusto, using steps surely modeled after author Arlyne Simon's own experience as an engineer. After identifying the need for a folding machine and receiving a vote of confidence from a parent, they set off on the path to invention.

First, the pair survey their neighbors to confirm demand for a folding machine. Then they design and hone a variety of models. Along the way, Abby and Miko face setbacks, rely on their ingenuity, and draw inspiration from everyday objects. Finally, they complete their "Foldibot" and—true engineers that they are—apply for a patent.

Together with the book's colorful depictions of an effortlessly diverse cast of characters, created by illustrator Diana Necşulescu, the story breaks down the opaque methods of discovery and will no doubt empower young readers to keep their eyes open for invention opportunities in their own lives. The final pages of the book provide questions to the reader, such as what they might have named their own folding machine and how they could improve its speed and design. Those hoping to recreate the Foldibot exactly, however, will be disappointed. Although the illustrations include a few diagrams and explanations, key components of this mythical machine remain unexplained. Here's hoping, however, that this entertaining little manual will inspire the next laundry innovation: a machine that places folded clothes in the proper drawer!

Abby Invents the Foldibot, *Arlyne Simon, Illustrated by Diana Necşulescu*, Abby Invents, 2021, 48 pp.

Monarch Butterflies

Reviewed by **Valerie Thompson**[8]

The monarch butterfly takes center stage in this text for grade-schoolers, which offers age-appropriate insights into the species' life cycle, habitat, and extraordinary migration. The book, written by former elementary school teacher Ann Hobbie, begins by walking readers through a monarch's first moments as a newly hatched caterpillar and describing its rapid growth and eventual

© 2021 PAUL BOSTON FROM *MIMIC MAKERS: BIOMIMICRY INVENTORS INSPIRED BY NATURE* BY KRISTEN NORDSTROM

metamorphosis. Simple questions ("How is the caterpillar changing?...What are they up to?") encourage young readers to engage more deeply with the material, while Hobbie's vivid descriptions ("It scoots out of its skin and pops off its head capsule like a face mask!") will either delight or horrify, depending on the audience.

Monarch migration comes next, with separate consideration of populations found east and west of the Rocky Mountains, which undertake markedly different journeys. Colorful visuals by illustrator Olga Baumert help to emphasize the rich habitats in which the butterflies live and through which they migrate, as well as the vibrant role they play in Mexican Día de los Muertos celebrations.

The book gently describes how habitat destruction and pesticides can harm butterflies before assuring readers that there are also ways to help them thrive. It closes with a glossary, instructions for rearing monarchs, and a note directing educators to complementary curriculum prepared by the Monarch Joint Venture program, a multi-institutional effort to protect monarch migration for which Hobbie serves as chair of the board.

A child who reads this text would be forgiven for mistakenly concluding that monarchs are only found in the Americas. (They are native to this region but now inhabit parts of Europe and various Pacific islands and nations.) Still, *Monarch Butterflies* would make an admirable addition to any grade-schooler's lepidoptery library.

Monarch Butterflies: Explore the Life Journey of One of the Winged Wonders of the World, *Ann Hobbie, Illustrated by Olga Baumert*, Storey Publishing, 2021, 48 pp.

HANDS-ON SCIENCE BOOK

Biology for Kids

Reviewed by **Michael A. Funk**[9]

In *Biology for Kids*, Liz Lee Heinecke profiles 25 biologists, describing their upbringing and personal lives, their professional contributions, and their legacies in a few paragraphs each. The list includes some of the usual suspects, such as Charles Darwin and Louis Pasteur, but a large portion are less-often-told stories of women and people of color who have made essential contributions to biology and medicine. These stories help explain, in very general terms, some of the fundamental concepts today's biologists use to identify different species, trace the spread of disease, and peer into the invisible world of microbes and molecules.

Each scientist's profile is accompanied by a beautiful illustration by artist Kelly Anne Dalton and is followed by an experiment or activity related to the subject's work. Biology is a broad field, and this book has an activity for every curious kid, from studying animal behavior, to growing plants from seed, to watching yeast take over a slice of bread. It will not spoil the fun, but some of the activities lack much in the way of explanation of the underlying concepts. Others require considerable preparation and cleanup, which may put off some kids (and parents). Fortunately, there are many options to pick from, and several require little more than observing what is happening in the world.

The short profiles in this book also raise a challenge for parents and older readers: find and read biographies of scientists. Many have led fascinating, unusual, and challenging lives that intersect with the science and culture of today in surprising ways.

ILLUSTRATION: OLGA BAUMERT

Biology for Kids: Science Experiments and Activities Inspired by Awesome Biologists, Past and Present, *Liz Lee Heinecke, Illustrated by Kelly Anne Dalton,* Quarry Books, 2021, 128 pp.

The Science and Technology of Marie Curie

Reviewed by **Adrian Cho**[10]

Marie Skłodowska Curie, who pioneered the study of radioactivity, has become a symbol of dogged determination. In Paris, she and her husband, Pierre, toiled endlessly in a shed filled with noxious chemicals to extract the tiniest amounts of radioactive radium from tons of the mineral pitchblende. But as Julie Knutson briskly recounts, Curie overcame numerous daunting challenges in her life.

Curie—the first woman to win a Nobel Prize and the only person to win a Nobel Prize in two different sciences—was born in 1867 in Russian-occupied Warsaw. She was not allowed to learn Polish in school and was largely educated by her parents. Sexism cost her opportunities and almost the 1903 Nobel Prize in Physics. In 1906, Pierre died in an accident, leaving her with two daughters, aged 1 and 8. Yet Curie persisted in her studies.

Knutson's book brims with sidebars on topics ranging from the structure of the atom to the bicycling craze of the 1890s. Many come with QR codes that link to additional websites and videos. Each of the seven chapters ends with suggestions for activities and experiments, such as instructions to fashion a nifty spectroscope from a cardboard tube and a compact disk. (Sorry, no experiments with radioactivity.)

Knutson addresses sexism head on and strives to encourage girls who are interested in science. However, she also notes that Curie's success did not instantly open doors for other women. Instead, Curie's contemporaries largely viewed her as a singular talent. By acknowledging this point, Knutson implicitly challenges her readers not to fall for the same lazy thinking.

The Science and Technology of Marie Curie, *Julie Knutson, Illustrated by Michelle Simpson*, Nomad Press, 2021, 128 pp.

Cardboard Box Engineering

Reviewed by **Marc S. Lavine**[11]

Have you ever seen a young child receive a gift and show far more interest in playing with the packaging than the toy or game it contained? In *Cardboard Box Engineering*, Jonathan Adolph provides more than 20 projects designed to entice, encourage, and inspire readers to create, tinker, and build.

The book is a perfect blend of information, illustration, and blueprints that, together, educate the reader about types of packaging materials, describe the tools and supplies that are needed for making the projects, and provide a detailed set of fabrication steps and photographs. Some projects, such as the three-wing boomerang, require little more than cardboard and glue, while others make use of common or

inexpensive objects such as screws, string, cork, toothpicks, and rubber bands.

The projects are grouped by topic and include sections on robotics and animatronics; harnessing energy; and aeronautics and nautical engineering. There are also quite a few historical spotlights that highlight people such as Sir George Cayley, Christian Ristow, and Stephanie Kwolek who helped to advance specific branches of science or engineering. While some of the projects should be good to go once built, others, such as the soaring airplane, offer creators the opportunity to tinker with the finished glider in order to optimize how it flies.

This is a fabulous book for at-home rainy-day projects, but it is also one that could aid educators in the classroom, where projects could be used to help teach concepts in science and engineering.

Cardboard Box Engineering: Cool, Inventive Projects for Tinkerers, Makers and Future Scientists, *Jonathan Adolph*, Storey Publishing, 2020, 176 pp.

Sky Gazing

Reviewed by **Keith T. Smith**[12]

Astronomy is a science that anyone can do—just step outside and look up at the night sky. This engaging book provides an accessible introduction to stargazing, aimed at children aged 9 to 14. It provides a wealth of information on the Sun, the Moon, and the planets in our Solar System and briefer coverage of stars, comets, and meteors.

The book is superbly illustrated with a colorful mix of photographs, diagrams, hand-drawn pictures, and a few comic strips. There are numerous practical activities to try, such as making a sundial from a stick or a pinhole camera from a shoe box.

Throughout the book, author Meg Thacher emphasizes objects that can be seen with the naked eye, occasionally supplemented with a basic pair of binoculars—no telescope is necessary. Light pollution and the weather can limit any budding observer, but Thacher offers practical tips on how to avoid or minimize their impact. Sky charts show the constellations and how to find targets in both the Northern and Southern Hemispheres. The descriptions are up to date with recent research and spacecraft results, although the focus is on NASA missions over those of other space agencies.

There are some niggles. Textual labels are sometimes incongruously small for the large diagrams they are attached to, and there are as many as six different fonts on the same page, which I found distracting. Some of the activities might not hold a child's attention, such as recording a twice-daily log of local weather conditions. I also spotted a few minor factual errors, but those were mostly oversimplifications, which are forgivable in a children's book. These are small issues that should not distract from an otherwise excellent introduction to astronomy.

Sky Gazing: A Guide to the Moon, Sun, Planets, Stars, Eclipses, and Constellations, *Meg Thacher*, Storey Publishing, 2020, 132 pp.

YOUNG ADULT SCIENCE BOOK

Chemistry for Breakfast

Reviewed by **Marc S. Lavine**[11]

From a young age, my kids learned not to ask me a science question unless they had time for a detailed answer. It always seemed better to me to give broad context to their inquiries, so they would ultimately have a better understanding of the way the world works. In the opening chapter of her book, Mai Thi Nguyen-Kim similarly cautions readers against seeking simple scientific answers. "This might sound arduous at first, but I promise that scientific thinking doesn't make the world drier," she writes. "[I]n fact, it makes it more colorful and literally full of wonder."

As the daughter of a chemist and a trained chemist herself, it is not surprising that Nguyen-Kim has chosen chemistry as the lens through which to frame many of the episodes in her daily life that are described in the book. Whether discussing nonstick pans or the formulation of toothpaste, she concisely describes the underlying science so that readers can appreciate the important role played by fluorine in both contexts.

One recurring subtheme is the importance of learning to distinguish between marketing hype and science. The notion that something "natural" is better than something "artificial," for example, is based on a false premise that often misleads consumers.

Nguyen-Kim brings an irreverence to her writing that makes each chapter easy to read. Even her discussions of the challenges of being an academic scientist, which might seem discouraging in a book trying to promote science, make sense in the broader storytelling. Her goal is not to convince readers to become scientists. Rather, she seeks to show how it is important to know enough science to be able to navigate the world as a critical thinker. One can appreciate the beauty and smell of a rose, but one can appreciate it even more by understanding the biology and chemistry that underlie these characteristics.

Chemistry for Breakfast: The Amazing Science of Everyday Life, *Mai Thi Nguyen-Kim, Translated by Sarah Pybus , Illustrated by Claire Lenkova*, Greystone Books, 2021, 240 pp.

Ms. Adventure

Reviewed by **Caroline Ash**[13]

Volcanologist Jess Phoenix thought she was writing a book about volcanoes, but *Ms. Adventure* might just as easily fall into the category of medical textbook. Phoenix, we learn, has survived a kidney infection on the flanks of a Hawaiian lava flow and a gashed thigh from falling into a Peruvian sewer and has undergone shoulder surgery in Mexico. Yet she has gamely plunged into one misadventure after another in the name of science.

Phoenix's circuitous route to a career in science started early, when she had to take medical leave from Smith College while unhappily studying English and history. This break proved fortuitous; during her time away, she took classes in geology and discovered her true passion. This book records her conversion to geology and her conviction that science should be accessible to all.

Phoenix's accounts are replete with memories and often headlined by accidents. She recalls the fun and trepidations of fieldwork as a young student—explaining in detail how to defecate in Death Valley undetected—and a 3-month stint volunteering at the Hawaiian Volcano Observatory. Here, she helped survey the enormity of Mauna Loa. The pinnacle of this adventure is a descent down the 1984 vent—a feat accomplished, for once, without mishap. From Mauna Loa, Phoenix teeters along fissures in the nearby erupting Kīlauea before

ILLUSTRATION: HANNAH BAILEY

plunging into the depths to sample a sea volcano for her master's thesis.

The book continues apace across South and Central America through a selection of close encounters with hazards, ranging from sewers and glacial moraines to drug traffickers and ending at her epiphany: election to the Explorers Club. It is in these hallowed halls that she decides to run for Congress in 2018. "I did it because that is what explorers do, after all," she writes in the book's closing pages. "We enter the unknown, seeking to illuminate the darkness and charting a course for all those who wish to find a way."

Ms. Adventure: My Wild Explorations in Science, Lava, and Life, *Jess Phoenix*, Timber Press, 2021, 272 pp.

Great Adaptations

Reviewed by **Laura M. Zahn**[14]

Earth teems with life-forms, including many distinctive creatures few of us will ever see in the wild, let alone study in the course of our careers. Vanderbilt biologist Kenneth Catania's book *Great Adaptations* recounts tales from the author's studies in animal biology across a menagerie of organisms, from star-nosed moles to tentacled snakes.

This engrossing read walks readers through Catania's life as a scientist. Key discoveries are juxtaposed with lessons about the scientific method, and QR codes sprinkled throughout the book link to short, engrossing movies highlighting some of the author's fascinating work. We learn, for example, that worms head to the soil's surface when encountering noises that likely are similar to those made by moles digging; a fact exploited by human "worm grunters," who collect them for fishing bait.

Catania does a good job of making it clear when his studies are not conclusive. He acknowledges, for example, that a macabre test he designs to assess cockroach "free will" is more for show than a true test of roach preference. And although he does not walk readers through his failures and disappointments, he makes it clear that there have been many dead ends and that his discoveries are sometime due to serendipity. His story also offers plenty of encouragement to those who may wish to follow a similar career trajectory.

Throughout the text, Catania reiterates how important it is to test our prior assumptions. Concluding that "one underappreciated key to making discoveries is keeping an open mind and not having too many preconceptions," he proves himself a font of useful knowledge that is applicable to one's broader life as well as to one's scientific endeavors.

Light-emitting diodes, powered by an eel's high-voltage attack, illuminate a fake alligator head.

PHOTO: KENNETH CATANIA

Great Adaptations: Star-Nosed Moles, Electric Eels, and Other Tales of Evolution's Mysteries Solved, *Kenneth Catania*, Princeton University Press, 2020, 224 pp.

The Loneliest Polar Bear

Reviewed by **Pamela J. Hines**[15]

Polar bears draw crowds at zoos. But how do they get there? What happens behind the scenes? Kale Williams takes on these questions and more with his wide-ranging and well-researched book *The Loneliest Polar Bear*. The book tells the story of Nora, a polar bear cub abandoned by her mother soon after birth at the Columbus Zoo in Ohio.

A team of zookeepers, who came to be known as the "Nora Moms," waited before deciding that Nora's survival was up to them. They invented an infant feeding formula based on what is known of polar bear diets in the wild, adjusted the temperature of the den from cozy to chilly as Nora grew into her fur, and introduced Nora to other polar bears, which have more social interaction, even in the wild, than one might think.

Along the way, Williams introduces readers to some of the Arctic research teams that scan the horizons for polar bears. We also learn about the small Alaskan village of Wales, where, in 1988, an Inupiat hunter fell into the den of Nora's grandmother, startling both bear and hunter and ending in the bear's death. (Although this is not how kills usually go, subsistence hunting has long sustained Indigenous communities.) The two cubs discovered in the den were rescued and transferred to a zoo in Anchorage: one went on to become Nora's father.

Today, both humans and polar bears in the Arctic struggle with the effects of climate change, as access to hunts diminishes and ice-sustained ranges shrink. Nora's story gives readers a glimpse into the science and history of both climate research and the various roles played by zoos, as well as how both polar bear and Native communities are struggling to navigate our changing world. ■

The Loneliest Polar Bear: A True Story of Survival and Peril on the Edge of a Warming World, *Kale Williams*, Crown, 2021, 288 pp.

10.1126/science.abn0866

[1]The reviewer is a deputy editor at *Science*. Email: svignier@aaas.org [2]The reviewer is a publications assistant at *Science*. Email: hstewart@science-int.co.uk [3]The reviewer is a managing editor at the *Science* journals. Email: twagoner@aaas.org [4]The reviewer is an associate editor at *Science*. Email: sscanlon@science-int.co.uk [5]The reviewer is an associate editor and news writer at *Science*. Email: kservick@aaas.org [6]The reviewer is a lead content production editor at *Science*. Email: lkmec@aaas.org [7]The reviewer is the Letters editor at *Science.* Email: jsills@aaas.org [8]The reviewer is a senior editor at *Science*. Email: vthompso@aaas.org [9]The reviewer is an associate editor at *Science*. Email: mfunk@aaas.org [10]The reviewer is a news writer at *Science*. Email: acho@aaas.org [11]The reviewer is a senior editor at *Science*. Email: mlavine@aaas.org [12]The reviewer is a senior editor at *Science*. Email: ksmith@science-int.co.uk [13]The reviewer is a senior editor at *Science*. Email: cash@science-int.co.uk [14]The reviewer is a senior editor at *Science*. Email: lzahn@aaas.org [15]The reviewer is a senior editor at *Science*. Email: phines@aaas.org

PERSPECTIVES

METABOLISM

Short-circuiting respiration

Fumarate siphons electrons to keep metabolism running

By **Sanjeethan C. Baksh** *and* **Lydia W. S. Finley**

The flow of electrons sustains all living organisms. Metabolic networks harness the chemical energy of nutrients by capturing high-energy electrons, which are then deposited onto the mitochondrial electron transport chain (ETC). Here, a series of protein complexes couple oxidation-reduction reactions to the generation of an electrochemical potential that drives synthesis of adenosine triphosphate, a common energy currency. Electron flow through the ETC is also critical for myriad other cellular functions, including biosynthetic and signaling processes (*1*). In metazoans, oxygen enables continual electron flow by serving as a terminal electron acceptor. Nevertheless, many cells exist in oxygen-poor environments while sustaining ETC-dependent reactions (*2*), raising the question of where electrons go when they cannot reduce molecular oxygen. On page 1227 of this issue, Spinelli *et al.* (*3*) identify the tricarboxylic acid (TCA) cycle metabolite fumarate as an alternative electron acceptor.

At the heart of the ETC is the electron carrier ubiquinone, which accepts electrons from inputs into the ETC, forming ubiquinol. Ubiquinol oxidation by complex III of the ETC regenerates the oxidized ubiquinone required to enable continual electron donation into the ETC. Thus, blocking ubiquinone regeneration by ablating complex III or its downstream partner, complex IV (which reduces molecular oxygen), should ultimately block any metabolic pathway that depends on depositing electrons into the ETC (*4*). However, Spinelli *et al.* found that genetic ablation of complex III or IV in human cancer cells did not abolish electron flow. This observation demonstrates that oxygen is not required for ETC function and suggests the existence of alternative terminal electron acceptors.

Both hypoxia and ETC inhibition were associated with increased amounts of succinate, a TCA cycle intermediate that conventionally deposits electrons to ubiquinone through complex II of the ETC, thereby forming fumarate. Under basal conditions, most succinate is generated by oxidative metabolism of substrates in the TCA cycle. Spinelli *et al.* demonstrated that during low oxygen or ETC inhibition, a portion of succinate was generated from fumarate, suggesting that fumarate reduction to succinate may provide a valve for excess electrons from the ETC. Notably, fumarate was not passively capturing rogue electrons leaking from the ETC; rather, fumarate reduction was catalyzed specifically by complex II itself (see the figure).

Spinelli *et al.* demonstrated that the ability of complex II to operate in reverse and reduce fumarate required increased concentrations of ubiquinol relative to ubiquinone. Under conditions of extremely impaired ubiquinone regeneration, the relatively unfavorable reaction of fumarate reduction by complex II becomes permissive, siphoning electrons from the ubiquinol pool on to fumarate. This is consistent with prior work demonstrating that either complex I or complex II, each of which deposits electrons into the ETC, can reverse directionality and accept electrons from ubiquinol during conditions of low oxygen or ETC inhibition (*5, 6*). By providing an alternative route for ubiquinol oxidation, reversed complex II activity enables oxidative metabolic networks to continue to deposit electrons into the ETC even when oxygen cannot serve as an electron acceptor. This work therefore calls into question the widely held notion that hypoxia will inhibit pathways that depend on ETC function and cautions that oxygen consumption does not always represent ETC activity. Intriguingly, as stem cells frequently reside within hypoxic niches (*7*), the work by Spinelli *et al.* raises the possibility that reverse complex II activity may support ETC-dependent cell-fate decisions, for example, by regulating succinate and nucleotide pools, thereby regulating gene expression (*7–10*).

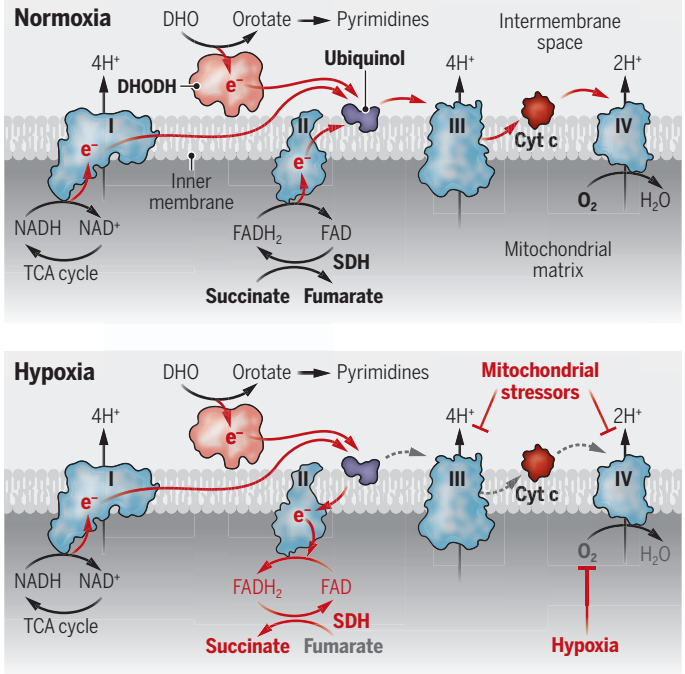
Spinelli *et al.* find that complex II reversal is not limited to extreme conditions of low oxygen or impaired ETC activity. Carbon tracing strategies in mice revealed a surprising heterogeneity in the degree to which tissues generate succinate from fumarate. Some tissues never demonstrated fumarate reduction; others reduced fumarate after physiological challenges such as exercise or hypoxia. Other tissues, such as kidney, liver, and brain showed constitutive production of succinate from fumarate regardless of oxygen availability. Although the mechanisms governing the ability of tissues to initiate complex II reversal and fumarate reduction remain to be elucidated, these results reveal the intrinsic diversity of ETC wiring across mammalian tissues.

What is the benefit of reversing complex II? One possibility is that this may allow tissues to cope with "reductive" stress—

Cell Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. Email: finleyl@mskcc.org

Electron flow

Under normoxia, electrons flow through ubiquinol to oxygen as a terminal electron acceptor. When oxygen reduction is limiting, electrons accumulate on ubiquinol, leading to reverse electron flow through SDH, generating succinate from fumarate.



Cyt c, cytochrome c; DHO, dihydroorotate; DHODH, dihydroorotate dehydrogenase; $FADH_2$, flavin adenine dinucleotide; NADH, nicotinamide adenine dinucleotide; SDH, succinate dehydrogenase; TCA, tricarboxylic acid.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

conditions in which the deposition of electrons into the ETC exceeds the capacity of the ETC to reduce oxygen. Absorbing a few electrons onto fumarate might enable cells to avoid uncontrolled electron leakage that generates damaging reactive oxygen species. Such scavenging could be particularly advantageous in tissues such as liver and kidney that are major sites of metabolite detoxification. Larger reverse fluxes through complex II might enable cells to sustain oxidative pathways such as amino acid and nucleotide biosynthesis and thus maintain cell proliferation under otherwise hostile conditions. Understanding the function of complex II reversal in normal physiology will thus require quantifying the net flux of fumarate reduction relative to succinate oxidation.

The ability of complex II to provide a valve for excess electrons also has important implications for pathological conditions such as ischemia and cancer. Complex II reversal has been linked to ischemia-reperfusion injury (*6*), but whether it is required for tissue adaptation to low oxygen is unknown. Likewise, tumor cells frequently experience hypoxic microenvironments in vivo. Given that complex II reversal supports the ETC to a lesser extent than oxygen, it is unknown whether complex II can meet the biosynthetic demands of tumor cells growing in hypoxic microenvironments. The potential of fumarate to absorb electrons to combat oxidative stress is especially intriguing in the context of human tumors with mutations in fumarate hydratase (FH). The current findings suggest that fumarate accumulation, characteristic of FH-mutant tumors (*7*), may provide a metabolic advantage under mitochondrial stress. Conversely, mutations in complex II are also observed in human tumors (*7*), raising the question of whether these cells engage alternative electron acceptors during hypoxic stress or whether the absence of fumarate reduction exposes a targetable weakness in these tumors. Counterintuitively, complex II and FH mutations would be predicted to have opposite effects on fumarate-mediated ubiquinol reduction, but both occur in renal cell carcinoma. These tumors may therefore represent a valuable system for future studies to understand the pathological implications of reverse complex II activity. ■

REFERENCES AND NOTES

1. J. B. Spinelli, M. C. Haigis, *Nat. Cell Biol.* **20**, 745 (2018).
2. P. Lee *et al.*, *Nat. Rev. Mol. Cell Biol.* **21**, 268 (2020).
3. J. B. Spinelli *et al.*, *Science* **374**, 1227 (2021).
4. I. Martínez-Reyes *et al.*, *Nature* **585**, 288 (2020).
5. E. L. Robb *et al.*, *J. Biol. Chem.* **293**, 9869 (2018).
6. E. T. Chouchani *et al.*, *Nature* **515**, 431 (2014).
7. S. C. Baksh, L. W. S. Finley, *Trends Cell Biol.* **31**, 24 (2021).
8. M. P. Rossmann *et al.*, *Science* **372**, 716 (2021).
9. C. Mao *et al.*, *Nature* **593**, 586 (2021).
10. R. M. White *et al.*, *Nature* **471**, 518 (2011).

10.1126/science.abm8098

VOLCANOLOGY

Reactivation of Cumbre Vieja volcano

A long-quiescent volcano's behavior requires rethinking about forecasting and hazards

By **Marc-Antoine Longpré**[1,2,3]

After 50 years of repose, Cumbre Vieja volcano—historically the most active of the Canary Islands—entered an eruptive episode on 19 September 2021, forcing the evacuation of ~6400 residents and destroying infrastructure worth more than 400 million euros. The volcanic unrest began in 2017 but accelerated only 8 days before the onset of the eruption. This behavior, characterized by comparatively protracted periods of quiescence and unrest, is at odds with global systematics for basaltic volcanoes (*1*).

Close monitoring of persistently active volcanoes, such as Kīlauea (Hawai'i) and Mount Etna (Sicily), allows for the identification of patterns in precursory unrest that help in the forecasting of eruptions (*1*). However, this strategy is not possible at volcanoes characterized by much longer quiescence periods, for which the most recent eruptive events predate the monitoring record. Cumbre Vieja volcano, on the island of La Palma in the Canary Archipelago, produced six eruptions between 1500 and 2020 CE, with repose periods ranging from 24 to 237 years (*2*). Before September 2021, it had last erupted in 1971, when a single seismic station installed on the island of Tenerife served the entire archipelago. In the past two decades, Spain's Instituto Geográfico Nacional (IGN) and the Instituto Volcanológico de Canarias (INVOLCAN) greatly expanded their monitoring networks on Cumbre Vieja and elsewhere in the Canary Islands, which allowed for the capture of the details of volcanic unrest in the run-up to the 2021 eruption. A preliminary evaluation of this unrest, and of the style and impact of the ensuing and ongoing eruption, offers valuable lessons for eruption forecasting, hazard assessment, and risk management in the Canaries and similar volcanic islands.

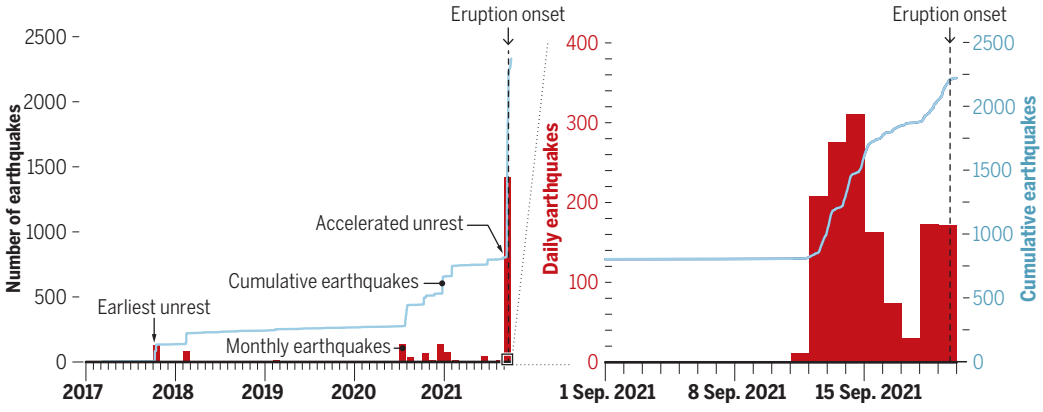
[1]School of Earth and Environmental Sciences, Queens College, City University of New York, Queens, NY, USA. [2]Earth and Environmental Sciences, The Graduate Center, City University of New York, New York, NY, USA. [3]Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. Email: mlongpre@qc.cuny.edu

The IGN earthquake catalog (*3*) indicates very low background seismicity beneath La Palma in the interval from 2000 to 2016, with only six low-magnitude events. In October 2017, however, an earthquake swarm was detected, comprising 128 events over 8 days, and a similar cluster of 84 earthquakes recurred in February 2018 (see the figure). Most of these events and subsequent pre-eruptive earthquakes ranged in magnitude from 1 to 2. In hindsight, these discrete, 15- to 30-km-deep seismic swarms likely mark the earliest sign of volcanic unrest (*4*, *5*)—although some have placed it even earlier (*6*)—and they may represent forerunner magma intrusions at mantle depth. Seismic activity returned to comparatively low levels for the following 2.5 years, with only 53 located events from March 2018 to June 2020, but picked up again in late July 2020. Six distinct swarms, ranging from 14 to 160 events, occurred between July 2020 and February 2021. Nevertheless, the next 6 months were relatively quiet, with minor clusters of 39 and 14 earthquakes occurring in June and August 2021, respectively. Notably, hypocenter modes show a deepening trend from October 2017 (20 to 25 km) to June 2021 (30 to 35 km) but return to 20 to 25 km in August 2021, possibly reflecting magma plumbing dynamics.

Beginning on 11 September 2021, the patterns in precursory unrest changed substantially. The number of detected earthquakes rapidly increased to several hundred daily, of which only a subset have been located. These events clustered at much shallower depths (<12 km) and were of greater mean magnitude than previous seismicity. Notably, and in contrast to earlier swarms, this activity was accompanied by marked ground deformation recorded by GPS and interferometric synthetic aperture radar (InSAR) (*7*), presumably caused by shallow magma migration. This accelerated run-up provided only 8 days of warning for an imminent eruption, which started at 14:10 UTC on 19 September 2021. Relocated earthquakes of the 11 to 19 September sequence reported by IGN show a remarkable shallowing trend and migra-

Cumbre Vieja reactivates

After 50 years of quiescence, sporadic and low-magnitude seismicity beneath the volcano started up in 2017. However, a sudden increase in the number of earthquakes occurred only 8 days before the eruption began. Data from (*3*).



tion to the northwest—toward the eventual eruption site.

Two eruptive fissures about 200-m long broke out on the northwestern flank of the ridge-shaped volcano, at 1 km above sea level and 2 km east of the village of El Paraíso. Timely evacuation of the residents living downslope of the vents proceeded smoothly. From the beginning, the eruption was both explosive and effusive, and pyroclastic activity of Strombolian style rapidly built a main cone and fed lava flows, initially traveling at 700 m/h (*7*). Lava bulldozed through densely settled areas and had destroyed ~2600 buildings, >70 km of road, and 2.3 km^2 of crops within 6 weeks. In comparison, the 2018 eruption of Kīlauea volcano destroyed 723 buildings over 3 months (*8*). By the end of the ninth day of eruptive activity, the lava reached the coast, adding ~0.4 km^2 of land leading up to the Atlantic Ocean (*7*). The rapid and steady growth of the lava flow field, reaching 9.9 km^2 by 3 November, along with the associated destruction were captured by Sentinel-2 satellite data and reported frequently by the Copernicus Emergency Mapping Service (EMS).

The substantial explosivity of the Cumbre Vieja eruption is somewhat surprising because the extent of this behavior was not clear from the historical records of previous Canary Island eruptions (*2*). Violent Strombolian explosions at the main vent produced sustained, ash-rich eruption columns rising 3 to 6 km above sea level (*9*). This led to widespread volcanic ash fall in various parts of the island, even reaching Tenerife, covering >65 km^2 with thicknesses greater than several millimeters according to Copernicus EMS data. Within 4 weeks, up to 60 cm of ash and lapilli had accumulated at proximal locations, ~1 km southwest of the new cone. Ash fall and accumulation caused sporadic closure of La Palma airport and the cancellation of more than 300 flights (*10*), negatively affecting tourism, a major economic sector of the island. In addition to volcanic ash, the eruption column contained copious amounts of gases. Emissions of sulfur dioxide ranged from 3200 to 53,600 metric tons per day (*7*). Sulfur dioxide clouds remained in the troposphere but drifted thousands of kilometers over parts of Europe, Africa, Asia, and the Atlantic Ocean (*9*, *11*).

Behind images of wreckage in the lava's path hides tragedy for thousands of islanders who have lost their homes. Fortunately, this natural disaster has yet to directly claim human lives, thanks to the coordinated emergency response of local authorities who were well prepared, having gained experience with volcanic crisis management during the 2011 to 2012 submarine eruption near the neighboring island of El Hierro.

With disaster come opportunities. Identifying and dissecting the reactivation of Cumbre Vieja, from its very inception, after five decades of quiescence has tremendous value. The preliminary analysis outlined here suggests that the eruption had been preparing for 4 years. However, unrest accelerated to the point where an eruption seemed likely only 8 days before magma broke the surface. This scenario contrasts with the precursory run-up to the 2011 to 2012 eruption that lasted 96 days and increased more progressively (*12*). More broadly, Canary Island volcanoes defy global relationships between repose and eruption run-up times, both of which are typically shorter at more commonly pictured basaltic volcanoes (*1*). This represents crucial information for eruption forecasting at quiescent volcanoes.

The 2021 eruption of Cumbre Vieja is also a telling example of how basaltic volcanoes may simultaneously produce lava flows and considerable explosive activity. This can cause wide dispersal of potentially harmful volcanic ash that goes well beyond the immediate vicinity of the vent. This presents an opportunity to reassess past eruptions in the Canaries along with related hazards.

Another bright light among dark ash clouds is the international scientific cooperation effort coordinated largely by INVOLCAN. Multi-disciplinary teams on and off the ground are sharing data in near-real time, particularly by means of Twitter (*13*), which allows for rapid knowledge buildup and efficient dissemination to the public.

The eruption currently shows no sign of waning, which is expected because historical eruptions have lasted between 3 weeks and 3 months (*2*). When it does end, the resilient people of La Palma will recover and rebuild their communities. But someday, perhaps in a future distant enough as to challenge collective memory, the volcano will erupt again, in a different location. A similar social dilemma arises as that which has been described at Kīlauea volcano (*8*): Will the stakeholders do the necessary long-term planning to permanently reduce risk associated to urban development on the flanks of Cumbre Vieja? ■

REFERENCES AND NOTES

1. L. Passarelli, E. E. Brodsky, *Geophys. J. Int.* **188**, 1025 (2012).
2. M.-A. Longpré, A. Felpeto, *J. Volcanol. Geotherm. Res.* **419**, 107363 (2021).
3. Instituto Geográfico Nacional, Catálogo de terremotos (2021); www.ign.es/web/ign/portal/sis-catalogo-terremotos.
4. P. A. Torres-González *et al.*, *J. Volcanol. Geotherm. Res.* **392**, 106757 (2020).
5. C. López *et al.*, in *EGU General Assembly Conference Abstracts* (2018); https://ui.adsabs.harvard.edu/abs/2018EGUGA..20.7694L.
6. J. Fernández *et al.*, *Sci. Rep.* **11**, 2540 (2021).
7. Global Volcanism Program, "Report on La Palma (Spain)" in *Weekly Volcanic Activity Reports, 8 September-2 November 2021*, S. K. Sennert, Ed. (Smithsonian Institution and US Geological Survey, 2021).
8. B. F. Houghton *et al.*, *Nat. Commun.* **12**, 1223 (2021).
9. Toulouse Volcanic Ash Advisory Centre, La Palma (2021); http://vaac.meteo.fr/volcanoes/la-palma/.
10. Flightradar24, La Palma Airport (2021); www.flightradar24.com/data/airports/spc.
11. Copernicus Sentinel-5P Mapping Portal, Maps of volcanic SO2 concentrations (2021); https://maps.s5p-pal.com/so2/.
12. C. López *et al.*, *Geophys. Res. Lett.* **39**, L13303 (2012).
13. R. Lacassin *et al.*, *Geosci. Commun.* **3**, 129 (2020).

10.1126/science.abm9423

GRAPHIC: N. DESAI/*SCIENCE*

ORGANIC CHEMISTRY

Directing carboxylic acid dehydrogenation

A palladium ligand can activate carbon-hydrogen bonds yet avoid product olefin reactions

By **Yoshiharu Iwabuchi**

Enzymes can activate functional groups on molecules that are normally unreactive. For example, methane monooxygenase activates carbon-hydrogen (C–H) bonds to form methanol (*1*). Chemists have been able to mimic this chemistry using transition metal compounds to activate C–H bonds in organic compounds (*2*) and convert them into value-added compounds. Targeting specific C–H bonds in alkyl chains is more challenging and has long been thought to require enzymes, but strategies have emerged to activate and target particular C–H bonds in the presence of other more reactive functional groups, often relying on the ligands on the metal centers (*3*). On page 1281 of this issue, Wang *et al.* (*4*) describe a catalytic system that efficiently activates the β-methylene C–H bonds of fatty acids to create unsaturated acids, a reaction that had previously been considered impossible through direct C–H activation methods because of the higher reactivity of the olefin products.

Chemical routes to activating C–H bonds must not only overcome their high bond strengths (enthalpies of ~100 ± 15 kcal/mol) but also target specific bonds and not activate weaker bonds in other functional groups of the molecule. One strategy has been to use palladium (Pd) to displace H and form Pd–C bond intermediates. Regioselectivity requires bringing the active Pd catalytic center into close proximity to the C–H bond. The ligand coordinating the Pd atom plays a substantial role, with an appropriate topology to simultaneously promote the bond formation with the C and to break the bond between C and H. For example, the β-methylene groups of alkyl amides can be targeted for arylation with Pd catalysts (*5*).

Graduate School of Pharmaceutical Sciences, Tohoku University, Aobayama, Sendai 980-8578, Japan. Email: iwabuchi@mail.pharm.tohoku.ac.jp

Wang *et al.* extended the above approach and developed a catalytic system that efficiently proceeds with the regioselective dehydrogenation of fatty acids. This reaction had not been realized because the C(sp^2)–H bond of the product of this reaction, α,β-unsaturated carboxylic acid, is more reactive toward Pd than is the C(sp^3)–H bond of the substrate carboxylic acid. The newly formed double bond would react with the regenerated catalytically active species immediately after the reaction proceeds to form a chelate and deactivate the catalyst.

The authors show that the undesired reaction can be inhibited with the right ligand. Tautomeric, bidentate pyridone ligands formed six-membered ring transition states but had limited dehydrogenation yields because of unwanted product reactivity. A detailed analysis of the reaction mechanism model by use of density functional theory calculations showed how the Pd ligand can be designed to inhibit the unwanted second, nonproductive C(sp^2)–H bond activation (see the figure, top). The energy of the transition state for the withdrawal of C(sp^2)–H hydrogen from the first product α,β-unsaturated carboxylate intermediate reveals why one class of ligands was more effective (see the figure, bottom). For the complex with the L8 ligand, which forms a five-membered ring transition state, this reaction is disfavored by 6.8 kcal/mol, whereas the L33 ligand that formed a six-membered ring transition state had no barrier. Wang *et al.* took advantage of this to develop another useful reaction with this system, that of the synthesis of γ-alkylidene butenolides.

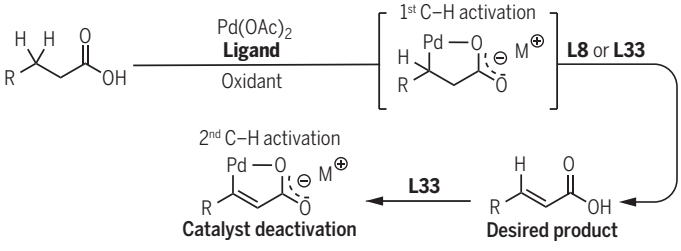
This process of reaction development by using experience, logical thinking, and computational chemistry exploited an unexplored chemical space, which nature has been refining through repeated evolution from chance mutation. Although enzymes acquired by living organisms and molecular catalysts created by chemists have different compositions and shapes, they can have many things in common. The α,β-dehydrogenation of carboxylic acid takes place as a central process in fatty acid metabolism by withdrawal of α-hydrogen from acyl–coenzyme A thioester and hydride transfer of β-hydrogen to flavin cofactor in the fatty acid hydrogenase (*6*). The approach of Wang *et al.* should strengthen the effective methodology of chemistry to make the impossible possible and lead to further development. ■

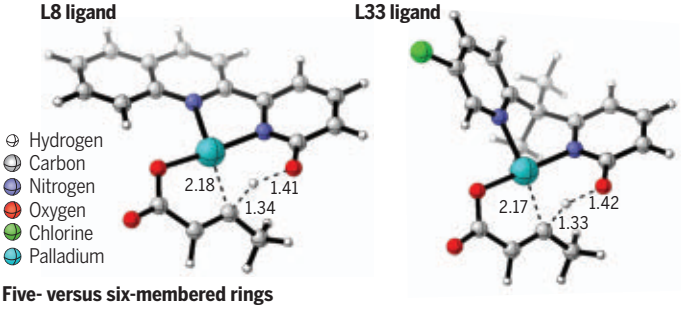
Stopping after the first bond

Wang *et al*. developed a palladium (Pd) catalyst that can activate the C(sp^3)–H bond of a carboxylic acid to form a double bond next to the –COOH group. An important feature in ligand design was avoiding attack on the remaining C(sp^2)–H bond, which would deactivate the catalyst and bind it to the product.

Active and overactive catalysts

The Pd catalysts with either the L8 or the L33 ligand activate the C–H bond to dehydrate the acid and form the double bond product. The ligand L33 attacked the remaining C–H bond, leading to catalyst deactivation and product binding, unlike the L8 ligand. Abbreviations: Ac, acetate; R, alkyl.





Five- versus six-membered rings

In the second, undesired activation step, the L8 catalyst would form a five-membered ring that is uphill in energy by almost 7 kcal/mol. The L33 catalyst forms a six-membered ring isoenergetic with the reactants and deactivates. Selected bond distances are in ångstroms.

REFERENCES AND NOTES

1. M. O. Ross, A. C. Rosenzweig, *J. Biol. Inorg. Chem.* **22**, 307 (2017).
2. R. H. Crabtree, *Chem. Rev.* **85**, 245 (1985).
3. T. W. Lyons, M. S. Sanford, *Chem. Rev.* **110**, 1147 (2010).
4. Z. Wang *et al.*, *Science* **374**, 1281 (2021).
5. G. Chen *et al.*, *Science* **353**, 1023 (2016).
6. C. Thorpe, J.-J. P. Kim, *FASEB J.* **9**, 718 (1995).

10.1126/science.abm4457

GRAPHIC: N. CARY/*SCIENCE* BASED ON Y. IWABUCHI

QUANTUM COMPUTING

Programming a quantum phase of matter

The ability to measure long-range entanglement may enable robust quantum memory

By **Stephen D. Bartlett**

At very low temperatures, some materials may condense into exotic phases of matter, where quantum entanglement becomes the dominant feature governing their behavior. These quantum phases—a world different from the ordinary states of solid, liquid, gas, and plasma—exhibit exotic properties such as quasiparticle excitations that interfere with each other in unusual ways. From an application point of view, these quantum phases may serve a key role in increasing the robustness of quantum memory devices—a key component in quantum computers. However, although theorists have predicted the existence of these quantum phases under a variety of conditions, quantum phases with long-range entanglement are extremely difficult to realize experimentally. On pages 1237 and 1242 of this issue, Satzinger *et al.* (*1*) and Semeghini *et al.* (*2*), respectively, provide direct observation of these phases and their key features by using a coupled superconducting circuit and an array of atoms.

There are many kinds of quantum phases, with superconductivity and Bose-Einstein condensates being two well-known examples. Motivated by fundamental questions in condensed matter physics, with potential implications for quantum information systems, researchers have been studying the pattern of long-range entanglement among spins. The quantum correlations among spins can be long range and "topological," meaning that the correlations are unchanged under continuous local deformations. For this reason, such exotic quantum phases of matter are said to be "topologically ordered."

Among the topologically ordered quantum phases, the most well studied are those that break time-reversal symmetry, meaning that they would behave differently if time ran backward. A key experimental signature of non–time-symmetric topological phases is their robust edge modes, which are persistent currents that run along the outer edge of the material. One such example is the fractional quantum Hall effect, which led to the discovery of a wealth of topological insulators and superconductors. Conversely, quantum computer architects are more interested in topologically ordered phases that are symmetric under time reversal because such phases can be used for error correction and therefore help protect quantum information from noise, perturbations, and other deleterious effects.

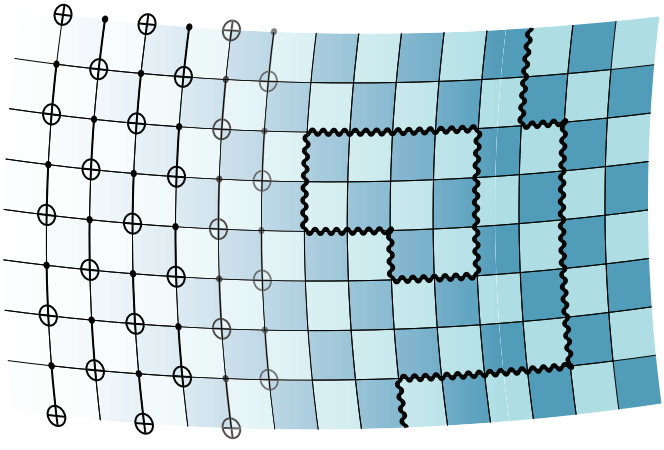
However, because time-symmetric topological phases do not have edge modes, traditional methods cannot be used to probe their long-range entanglement properties, which is necessary for realizing their potential to be used for quantum error correction. Because of the very nature of long-range entanglement, one cannot learn about the material's properties by examining a local region but must instead probe quantum correlations that traverse across the entire volume. Observing such nonlocal properties of a system requires exquisite control of the individual quantum constituents together with entangling interactions and precise measurements. This level of control over the individual components of a complex, many-component quantum system has only recently become possible thanks to state-of-the-art, albeit still rudimentary, quantum computing devices.

Satzinger *et al.* use a nascent quantum processor that consists of a two-dimensional array of 31 coupled superconducting quantum devices. The processor, known as Sycamore, made headlines in 2019 by claiming "quantum supremacy," or the ability to perform certain computing tasks faster than a conventional supercomputer (*3*). By executing a short quantum program, Satzinger *et al.* used Sycamore to stitch together the lowest-energy state of the toric code (*4*)—the canonical example of a topologically ordered quantum phase that can be used for quantum error correction. This together with another short quantum program allowed long-range entanglement to be measured. They also created additional programs to simulate the creation of quasiparticles and to perform a quantum interference experiment that illustrates the expected behavior of the quasiparticles. The authors also demonstrate that quantum information can be encoded into the toric code, protecting it from errors, and that this quantum information can be subsequently read out again. These properties illustrate how the toric code may become the central pillar in an approach to contain errors when scaling up the quantum architecture.

Semeghini *et al.* report on a different experiment that shares the goal of creating and exploring the properties of a topologically ordered phase related to the toric code. Their experiment uses a quantum simulator consisting of 219 atoms arranged on a two-dimensional lattice by using optical tweezers. By controlling the interaction between neighboring atoms, the lattice is coaxed into a topologically ordered phase.

Knitting a quantum topology

Shown is an artist's concept of how topological phases can be constructed by using quantum processors. The circuitry on the left represents a quantum circuit, which can be controlled to form specific patterns of long-range entanglement. The patterns of long-range entanglement—the hallmark of topological quantum phases—are indicated on the right by the wavy lines, of which some form closed loops and some have open strings extending out to the edge.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

Centre for Engineered Quantum Systems, School of Physics, University of Sydney, Sydney NSW, Australia. Email: stephen.bartlett@sydney.edu.au

As with the Sycamore processor, this atomic quantum simulator is programmable. The authors ran programs on the quantum processor and monitored the long-range entanglement properties among the processor's 219 atoms. Specifically, the authors measured how these quantum correlations establish themselves among the spins along a meandering path, which produce data that directly reflect the topological order of the quantum phase. Like Satzinger *et al.*, Semeghini *et al.* demonstrated how quantum information can be encoded into the system, by showing that a bit of encoded information can be read out again, and established a path for creating a quantum memory.

The two experiments represent the first definitive detection of a topologically ordered phase with time-reversal symmetry. Neither experiment was achieved by using new materials, as is usually the case. Instead, the achievement was realized virtually with quantum processors. And although the processors provide a mechanism to create quantum states entangled at long range, their most critical contribution is to provide a way to measure the long-range entanglement structure that characterizes topological order (see the figure). Satzinger *et al.* and Semeghini *et al.* illustrate the potential for quantum computer technology to serve as a tool for exploring quantum many-body systems.

To protect quantum information during computation, it will also be necessary to initialize, manipulate, and measure the quantum information in these codes and to do so by using circuits that are tolerant to errors. For practical applications, the error rates must be reduced even further from the levels achieved by the two experiments. And critically, quantum error correction will require repetitive measurements of the check operators for detecting errors and to update the logical information by decoding these measurements. This will require not only a powerful quantum processor but also a highly integrated classical processer and controller (*5*). However, despite the challenges still in the way of achieving a practically useful quantum processing device, the two experiments mark the first steps toward harnessing topologically ordered quantum phases for error correction in quantum computers. ■

REFERENCES AND NOTES

1. K. J. Satzinger *et al.*, *Science* **374**, 1237 (2021).
2. G. Semeghini *et al.*, *Science* **374**, 1242 (2021).
3. F. Arute *et al.*, *Nature* **574**, 505 (2019).
4. A. Kitaev, *Ann. Phys.* **303**, 2 (2003).
5. S. J. Pauka *et al.*, *Nat. Electron.* **4**, 64 (2021).

10.1126/science.abl8910

INSTRUMENTS

Room-temperature mid-infrared detectors

Two independent groups designed nanoantennas for detecting mid-infrared light

By **Reuven Gordon**

From astronomy to the surveying of greenhouse gases, a wide range of science and engineering applications rely on the detection of mid-infrared (mid-IR) photons. However, because photons from the mid-IR have less than a tenth of the energy of the visible range, the detection is not as easy as taking a photo with a camera. IR detectors are generally very sensitive to thermal noise and must be kept cool, but for applications where a cryogenic setup is impractical, other engineering solutions must be used. On pages 1264 and 1268 of this issue, Chen *et al.* (*1*) and Xomalis *et al.* (*2*), respectively, demonstrate two techniques to "up-convert" mid-IR light into the near-IR band, where it can be detected using standard room-temperature semiconductor detectors.

A common solution to increase the energy of IR photons so that they can be measured efficiently is the use of a process known as up-conversion. This process shifts up the frequency, and therefore the energy, of a photon so that it can be sensed using ordinary detectors. This is often realized by combining the incoming mid-IR photons with near-IR photons from a pump laser to produce a new photon with a higher energy. However, current design schemes to accomplish this require complicated matching of the propagation phase of the mid-IR photons from the source and the near-IR photons from the pump laser.

Chen *et al.* and Xomalis *et al.* each present an antenna design to sidestep the size and momentum mismatch issues that make existing up-conversion schemes inefficient. Both teams achieve high detection efficiency with four orders of magnitude less near-IR laser power. Whereas Chen *et al.* used a mid-IR slot antenna that resonantly confines the mid-IR energy to a narrow slot in a metal film, Xomalis *et al.* used a mid-IR disk antenna that uses a gold film to accomplish the same (see the figure). Both devices boost the interaction between near-IR and mid-IR light by using gold nanospheres in conjunction with mid-IR antennas to confine the light in the gap region containing biphenyl-4-thiol molecules. Both schemes combine two antennas in the same location to interact efficiently with these molecules without the need for complicated phase-matching—one mid-IR antenna and one near-IR one.

As the incoming mid-IR light approaches the antenna from the underside and the pump laser from the topside, they are focused down to an area 1 nm across. This goes well beyond the diffraction limit of the microscope objectives used in the setup. Using IR light with a wavelength of several microns and a beam width of the same size, the antennas used by Chen *et al.* and Xomalis *et al.* reduced the wavelength and beam width of the incoming IR light down to the nanometer range. Both designs achieved a focusing factor of 10,000 beyond the conventional diffraction limit to enhance the interaction with the molecules in the gap. In addition, the focusing factor for the near-IR light was around 1000. The resulting near-IR photon from the up-conversion is only slightly larger in energy than the incident light photons, so it also benefits from this extreme focusing.

The focused IR light is absorbed by a molecule placed inside a nanocavity within the antenna, which causes the molecules to vibrate. The energy from this vibration is then transferred to a photon from a near-IR laser in a process called anti-Stokes Raman scattering. After passing through a series of filters for cleaning up the signal, the photon from the anti-Stokes Raman scattering is detected using a conventional detector.

Raman scattering and IR absorption require vibrational modes with different symmetries. Raman scattering probes vibrations of even symmetry, and IR absorption probes odd. This means that the molecular vibration needs to be chosen to have both even and odd parts for this up-conversion scheme to work. With this in mind, both studies used the biphenyl-4-thiol as an up-converter. This molecule not only has a specific type of bond that

Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada.
Email: rgordon@uvic.ca

allows it to be bound to the gold antenna but also has a prominent vibrational mode that is both Raman-scattering and IR-absorption active.

By multiplying the enhancement factor from the focusing and anti-Stokes Raman scattering, it is estimated that a 10^{10}-fold efficiency enhancement can be achieved, with a potential enhancement factor of up to 10^{13}. The final result is the pronounced improvement in detection sensitivity demonstrated by Chen *et al*. in the detection of a mid-IR signal of less than a microwatt.

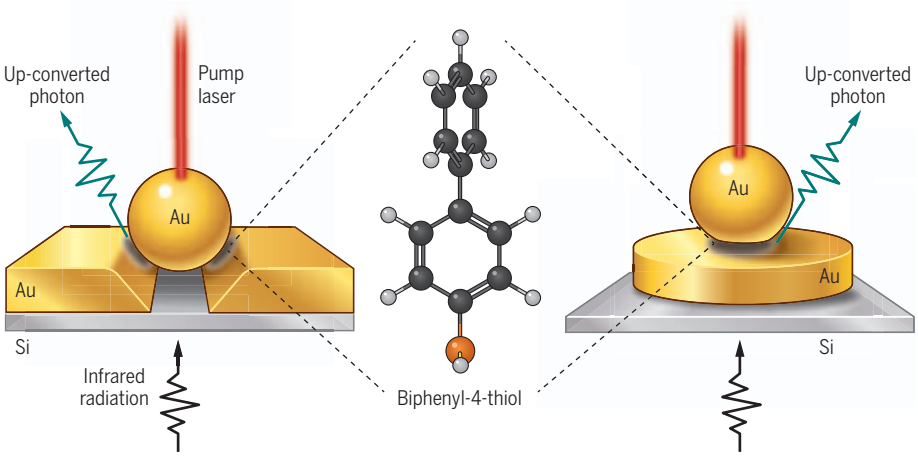
Although both studies showed outstanding sensitivity, far greater sensitivity is possible. For instance, silver may provide a greater enhancement than gold (*3*, *4*). Better antenna design can also bring further improvement. For example, an antenna with higher directivity will increase collection efficiency, such as those used in single-photon quantum light sources (*5*, *6*). With these modifications, it is expected that the same scheme could operate at the single-molecule level in the near term.

Once the single-molecule level is achieved, new quantum physics applications become possible by using the combined Raman scattering and IR absorption platform. Indeed, researchers have been advancing the field of quantum cavity optomechanics for years (*7*, *8*). The ability of a molecule's electronic state to coherently interact with its vibrational states through IR light offers a previously unexplored way to manipulate the quantum state of the molecule. It is also possible to couple between multiple molecules on a single chip using waveguides, which will allow for entanglement or quantum state manipulation, with potential applications in quantum computing.

Chen *et al*. and Xomalis *et al*. demonstrate a sensitive IR detection technique built upon decades of research in surface-enhanced Raman scattering (SERS), surface-enhanced IR absorption spectroscopy (SEIRA), and cavity quantum optomechanics. Substantially greater sensitivity is expected from these platforms by material optimization and antenna design. There is still the challenge of extending the platform to other IR wavelengths by finding molecules with simultaneous Raman and IR active vibrations in different energy ranges of interest. Perhaps more important than their use in room-temperature IR detection, these studies introduce a new quantum playground where unprecedented coherent manipulation and readout are possible. Based on past works on SERS and SEIRA, it is foreseeable that this will soon scale down to the single-molecule level, thereby enabling new quantum physics and information applications. ■

Two antenna designs for detecting infrared light

Both designs by Chen *et al*. and Xomalis *et al*. include a mid-infrared (mid-IR) antenna and a near-IR antenna. Chen *et al*. used a golden nanoparticle inside a 100 nm wide slot on a gold film, whereas Xomalis *et al*. also used a golden nanoparticle, but drop-casted on a microscopic gold disk. Both groups put biphenyl-4-thiol molecules in the tiny gap between the gold nanoparticle and the gold underneath it.



REFERENCES AND NOTES

1. W. Chen *et al*., *Science* **374**, 1264 (2021).
2. A. Xomalis *et al*., *Science* **374**, 1268 (2021).
3. K. Kneipp *et al*., *Phys. Rev. Lett.* **78**, 1667 (1997).
4. S. Nie, S. R. Emory, *Science* **275**, 1102 (1997).
5. A. Ahmed, R. Gordon, *Nano Lett.* **12**, 2625 (2012).
6. A. G. Curto *et al*., *Science* **329**, 930 (2010).
7. F. Benz *et al*., *Science* **354**, 726 (2016).
8. P. Roelli, C. Galland, N. Piro, T. J. Kippenberg, *Nat. Nanotechnol.* **11**, 164 (2016).

ACKNOWLEDGMENTS

R.G. is grateful for editorial comments from A. G. Brolo, R. de Sousa, M. Dobinson, and S. Mathew.

10.1126/science.abm4252

VIEWPOINT: COVID-19

Dissecting the early COVID-19 cases in Wuhan

Elucidating the origin of the pandemic requires understanding of the Wuhan outbreak

By **Michael Worobey**

Some key questions lie at the heart of investigations into the origin of the COVID-19 pandemic, including what is known about the earliest COVID-19 cases in Wuhan, China, and what can be learned from them? Despite assertions to the contrary (*1*), it is now clear that live mammals susceptible to coronaviruses, including raccoon dogs (*Nyctereutes procyonoides*), were sold at Huanan Market and three other live-animal markets in Wuhan before the pandemic (*2*, *3*). Severe acute respiratory syndrome–related coronaviruses (SARSr-CoVs) were found in raccoon dogs during the SARS outbreak, which was facilitated by animal-to-human contact in live-animal markets in China. However, because of the early public health focus on Huanan Market, it remains unclear whether the apparent preponderance of hospitalized COVID-19 cases associated with this market was truly reflective of the initial outbreak. Answering these questions requires resolving several crucial events that took place in December 2019 and early January 2020.

On 30 December 2019, the Wuhan Municipal Health Commission (WHC) issued two emergency notices for internal circulation to local hospitals alerting them to patients with unexplained pneumonia—several of whom worked at Huanan Market—and laying out a treatment and response plan (see fig. S1). The first official public report was WHC's announcement the next day that they had carried out case searches and retrospective investigations related to Huanan Market and found 27 patients. Forty-one of the first known patients formed the basis of an influential study that reported that 66%—i.e., not all early cases—had a link to Huanan Market

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA.
Email: worobey@arizona.edu

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

(*4*). They had been transferred between 29 December and 2 January from other hospitals to Jinyintan Hospital, Wuhan's premier infectious disease center. Notably, individuals were enrolled according to clinical presentation, not epidemiologic information, such as connections to Huanan Market (*4*).

China's Viral Pneumonia of Unknown Etiology (VPUE) mechanism was set up in the wake of SARS to be an early warning reporting system for detecting unknown viral diseases and is overseen by the China Center for Disease Control and Prevention (CCDC) (*5*). PUE cases are supposed to be rapidly reported by clinicians to the national notifiable disease reporting system through an internet-based platform. Evidently, that did not happen in Wuhan in December. The system appears to have been in active use only from 3 January. Although it favored cases having a connection to Huanan Market (*6*–*8*), the VPUE mechanism could not have improperly inflated the proportion of Huanan Market–linked cases in December (*1*). Moreover, reporting began only after the 41 patients were transferred from other hospitals to Jinyintan Hospital. Nevertheless, it is possible that a disproportionate number of cases linked to Huanan Market were transferred to Jinyintan Hospital because of public health officials' early focus there.

There is, however, a way to step back to a period before any such bias could have crept in, by considering what happened in the hospitals that first pieced together that a new viral outbreak was underway. Although not mentioned by name in scientific publications (*9*), media reports reveal that Hubei Provincial Hospital of Integrated Chinese and Western Medicine (HPHICWM) was the first hospital to alert district, municipal, and provincial public health authorities about the mysterious pneumonia cases (see fig. S1). Zhang Jixian, director of respiratory and critical care medicine, noticed on 27 December that an elderly couple had large "ground glass" opacities in computed tomography (CT) images of their lungs, distinct from those she had seen in other cases of viral pneumonia. Zhang insisted that the couple's son, who was not a patient and had no symptoms, undergo a CT scan, and the same unusual lesions were observed. The husband and wife evidently are "cluster 1" in the World Health Organization (WHO)–China report (*1*): They are the earliest known case cluster and the only cluster admitted by 26 December. They had no known connection to Huanan Market.

Another patient with similar CT imaging, a worker at Huanan Market, was admitted on 27 December. Zhang, concerned about a new, probably infectious viral disease, reported the four cases to hospital officials, who alerted the Jianghan District CDC that same day. Over 28 and 29 December, three more patients, all of whom worked at Huanan Market, were admitted and recognized to have the same unknown respiratory disease. A vice president of HPHICWM, Xia Wenguang, brought together 10 experts from the hospital, including Zhang, for an emergency meeting on 29 December, and they concluded that the situation was extraordinary. Upon learning of similar patients, also linked to Huanan Market, at Tongji and Union (Xiehe) Hospitals, Xia alerted the Wuhan and Hubei CDCs on 29 December.

A notably similar situation unfolded at Wuhan Central Hospital. On 18 December, Ai Fen, director of the emergency department, encountered her first unexplained pneumonia patient, a 65-year-old man who had become ill on either 13 or 15 December. Unbeknownst to Ai at the time, the patient was a deliveryman at Huanan Market. A CT scan revealed infection in both lungs, and he did not respond to antibiotics or anti-influenza drugs. On 24 December, a bronchoalveolar lavage specimen collected from him was sent to Vision Medicals, a metagenomics sequencing company. They identified a new SARSr-CoV on 26 December and relayed the finding by telephone to the hospital on 27 December. By 28 December, Wuhan Central Hospital had identified seven cases, of which four turned out to be linked to Huanan Market. Notably, these seven cases, like those at HPHICWM, were ascertained before epidemiologic investigations concerning Huanan Market commenced on 29 December.

At Zhongnan Hospital in the Wuchang District of Wuhan, 15 km away from Huanan Market and on the opposite bank of the Yangtze River, Vice President Yuan Yufeng asked units on 31 December to search for unexplained pneumonia cases, and the Respiratory Medicine Department reported two. The first lived in Wuchang District but worked at Huanan Market (in Jianghan District). The second did not work at Huanan Market but had friends who did and who had visited his home. On 3 January, three more cases were identified—a family cluster unlinked to Huanan Market. Clearly, hospitals in the first weeks of the outbreak were identifying cases both with and without a known connection to Huanan Market. And Wuhan hospitals were not swamped with unexplained pneumonia cases at the end of December—that would come later.

Thus, 10 of these hospitals' 19 earliest COVID-19 cases were linked to Huanan Market (~53%), comparable both to Jinyintan's 66% (of 41 cases) (*4*) and to the WHO-China report's 33% of 168 retrospectively identified cases within Wuhan across December 2019 (*1*). Regarding cases at the Wuhan Central Hospital and HPHICWM, patients with a history of exposure at Huanan Market could not have been "cherry picked" before anyone had identified the market as an epidemiologic risk factor. Hence, there was a genuine preponderance of early COVID-19 cases associated with Huanan Market.

How can this knowledge inform our understanding of the pandemic? If Huanan Market was the source, why were "only" one- to two-thirds of early cases linked to the market? Perhaps a better question is why would one expect all cases ascertained weeks into the outbreak to be confined to one market? Given the high transmissibility of SARS-CoV-2 and the high rate of asymptomatic spread, many symptomatic cases would inevitably soon lack a direct link to the location of the pandemic's origin. And some cases counted as "unlinked" may have been only one or two transmissions away, as exemplified by the second patient identified at Zhongnan Hospital. That so many of the >100 COVID-19 cases from December (*1*) with no epidemiologic link to Huanan Market nonetheless lived in its direct vicinity is notable (see the figure) and provides compelling evidence that community transmission started at the market.

Additionally, the earliest known cases should not necessarily be expected to be the first infected or linked to Huanan Market: They probably postdated the outbreak's index case by a considerable period (*10*). Moreover, only ~7% of SARS-CoV-2 infections lead to hospitalization (*11*); most fly under the radar. Similarly, it is entirely expected that early, ascertained cases from a seafood market would be workers who were not necessarily directly associated with wildlife sales once the outbreak began spreading from human to human. The index case was most likely one of the ~93% who never required hospitalization and indeed could have been any of hundreds of workers who had even brief contact with infected live mammals.

Crucially, however, the now famous "earliest" COVID-19 case (*1*), a 41-year-old male accountant, who lived 30 km south of Huanan Market and had no connection to it—illness onset reported as 8 December—appears to have become ill with COVID-19 considerably later (*12*). When interviewed, he reported that his COVID-19 symptoms started with a fever on 16 December; the 8 December illness was a dental problem related to baby teeth retained into adulthood (*12*). This is corroborated by hospital records and a scientific paper that reports his COVID-19 onset date as 16 December and date of hospitalization as 22 December (see fig. S1). This indicates that he was infected through community transmission after the virus had begun spreading from Huanan Market. He believed that he may have been infected in a hospital (presum-

ably during his dental emergency) or on the subway during his commute; he had also traveled north of Huanan Market shortly before his symptoms began (*12*). His symptom onset came after multiple cases in workers at Huanan Market, making a female seafood vendor there the earliest known case, with illness onset 10 December (see fig. S1). Notably, she reported knowledge of several possible COVID-19 cases in clinics and hospitals that were near Huanan Market from 11 December, and Huanan Market patients were hospitalized at Union Hospital as early as 10 December (see fig. S1).

Although a widely cited report (*7*) credits the VPUE mechanism with uncovering the pandemic, it was HPHICWM that identified both the outbreak and the Huanan Market connection and passed on these fully formed discoveries to district, municipal, and provincial public health officials by 29 December (*9*). National officials reportedly did not learn about the outbreak until CCDC Director George Gao encountered online group chats about the WHC emergency notices on the evening of 30 December. Concerned that so many cases had not been reported to the VPUE system, he quickly notified the National Health Commission (*13*) (see fig. S1).

Therefore, the preponderance of early cases connected to Huanan Market could not have been an artifact of ascertainment bias introduced by case definitions in the VPUE system. Although mechanisms like China's VPUE system are potentially invaluable, they will fail without both widespread buy-in from health care providers and rapid data sharing from local to central authorities. Key problems with the VPUE system were known before the pandemic, including that most clinicians in China had little awareness of the VPUE system and were not reporting cases to it—for example, 0 of 335 PUE cases in one study from 2019 (*5*). China should be commended, however, for having such a system, which is lacking in most countries. The focus now should be on fixing the problems that COVID-19 has exposed and blanketing the globe with a highly functional PUE early warning system.

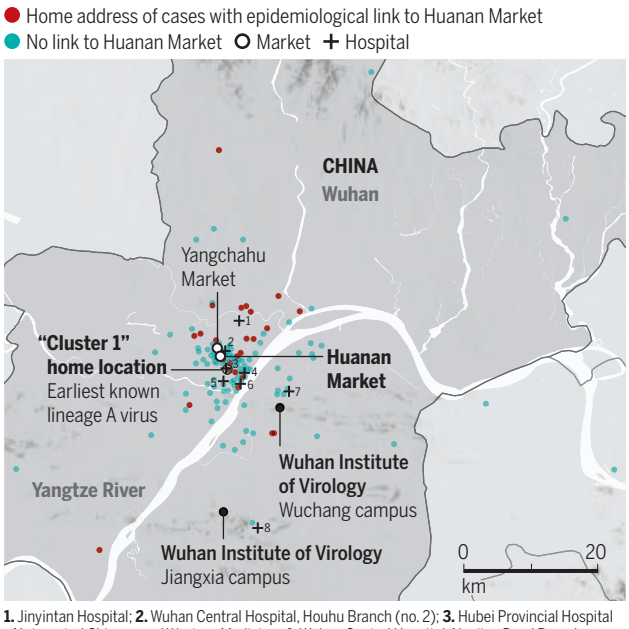
Samples from the earliest COVID-19 patients in Wuhan have been sequenced, and two distinct SARS-CoV-2 lineages, A and B, have been identified. Given that the elderly couple at HPHICWM was the WHO report's cluster 1, it follows that the husband, illness onset 26 December (*1*), must be the source of the earliest lineage A sequence, Wuhan/IME-WH01/2019 (GenBank accession number MT291826) (see fig. S1), which he most likely got from his wife, who became ill 15 December. This raises the possibility that the Yangchahu Market that they visited may have been a site of a separate animal spillover. The recent discovery that there may be no true lineage A or B intermediates in humans (*14*) also raises the possibility of separate spillovers of both lineages. However, the earliest known lineage A genomes have close geographical connections to Huanan Market: one from a patient (age and gender not reported) who stayed in a hotel near Huanan Market in the days before illness onset in December (*15*) and the other from the 62-year-old husband in cluster 1 who visited Yangchahu Market, just a few blocks north of Huanan Market (*1*), and lived just to the south (see the figure). Therefore, if lineage A had a separate animal origin from lineage B, both most likely occurred at Huanan Market, and the association with Yangchahu Market, which does not appear to have sold live mammals, is likely due to community transmission starting in the neighborhoods surrounding Huanan Market.

With SARS, live-animal markets continued to sell infected animals for many months, allowing zoonotic spillover to be established as the origin and revealing multiple independent jumps from animals into humans (*3*). Unfortunately, no live mammal collected at Huanan Market or any other live-animal market in Wuhan has been screened for SARS-CoV-2–related viruses (*1*), and Huanan Market was closed and disinfected on 1 January 2020. Nevertheless, that most early symptomatic cases were linked to Huanan Market—specifically to the western section (*1*) where raccoon dogs were caged (*2*)—provides strong evidence of a live-animal market origin of the pandemic.

This would explain the extraordinary preponderance of early COVID-19 cases at one of the handful of sites in Wuhan—population 11 million—that sell some of the same animals that brought us SARS. Although it may never be possible to recover related viruses from animals if they were not sampled at the time of emergence, conclusive evidence of a Huanan Market origin from infected wildlife may nonetheless be obtainable through analysis of spatial patterns of early cases and from additional genomic data, including SARS-CoV-2–positive samples from Huanan Market, as well as through integration of additional epidemiologic data. Preventing future pandemics depends on this effort. ■

COVID-19 cases in Wuhan in December 2019

The map shows that most of the earliest cases of COVID-19 were in close proximity to Huanan Market, even if they were not directly connected with the market through working there or visiting. This suggests that transmission in the community around the market was occurring in December 2019. The map is based on a subset of data from 174 COVID-19 cases in and around Wuhan (*1*).



1. Jinyintan Hospital; **2.** Wuhan Central Hospital, Houhu Branch (no. 2); **3.** Hubei Provincial Hospital of Integrated Chinese and Western Medicine; **4.** Wuhan Central Hospital, Nanjing Road Branch; **5.** Tongji Hospital; **6.** Union Hospital; **7.** Zhongnan Hospital; **8.** Wuhan Jiangxia First People's Hospital

GRAPHIC: K. FRANKLIN/*SCIENCE*

REFERENCES AND NOTES

1. WHO, WHO-convened global study of origins of SARS-CoV-2: China Part (2021); https://bit.ly/3wjUXze.
2. X. Xiao *et al.*, *Sci. Rep.* **11**, 11898 (2021).
3. E. C. Holmes *et al.*, *Cell* **184**, 4848 (2021).
4. C. Huang *et al.*, *Lancet* **395**, 497 (2020).
5. N. Xiang *et al.*, *BMC Infect. Dis.* **19**, 770 (2019).
6. T. K. Tsang *et al.*, *Lancet Public Health* **5**, e289 (2020).
7. Q. Li *et al.*, *N. Engl. J. Med.* **382**, 1199 (2020).
8. D. L. Yang, "Wuhan officials tried to cover up covid-19—and sent it careening outward," *Washington Post*, 10 March 2020.
9. The 2019-nCoV Outbreak Joint Field Epidemiology Investigation Team, *China CDC Weekly* **2**, 79 (2020).
10. J. Pekar *et al.*, *Science* **372**, 412 (2021).
11. S. Mahajan *et al.*, *Am. J. Med.* **134**, 812 (2021).
12. P. Ju, "Looking for the first infected person in the South China Seafood Market," video, *The Paper*, 25 March 2020; https://bit.ly/2YikwEa.
13. D. L. Yang, "China's early warning system didn't work on covid-19. Here's the story," *Washington Post*, 24 February 2020.
14. J. Pekar *et al.*, *Virological* **754** (2021); https://virological.org/t/evidence-against-the-veracity-of-sars-cov-2-genomes-intermediate-between-lineages-a-and-b/754.
15. R. Lu *et al.*, *Lancet* **395**, 565 (2020).

ACKNOWLEDGMENTS

Thanks to four anonymous reviewers and to A. Crits-Christoph, E. Holmes, D. Robertson, J. Wertheim, J. Pekar, K. Andersen, S. Goldstein, A. Rambaut, H. Mourant, D. Yang, L. Wang, S. Chen, C. Di, and Q. Jiang for assistance and discussions. The author is supported by the David and Lucile Packard Foundation and the Bill and Melinda Gates Foundation.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm4454

Published online 18 November 2021
10.1126/science.abm4454

POLICY FORUM

RESEARCH AND REGULATION

Challenges in confirming drug effectiveness after early approval

Reform requires clarity about whether, when, and how meaningful postapproval trials are possible

By **Holly Fernandez Lynch**[1] *and* **Christopher T. Robertson**[2]

It's easy to understand the urge to make potentially beneficial drugs quickly available to patients in need. It's also easy to go too far. Through its 2021 approval of Aduhelm (aducanumab) for treatment of Alzheimer's disease, the US Food and Drug Administration (FDA) showed a willingness to embrace early approval pathways in ways that risk FDA's reputation and undermine its core role in keeping the market free of worthless or dangerous medical products. Early approval pathways are intended to strike a careful compromise: access to promising therapies today, confirmatory evidence tomorrow. However, this often results in access—often at a hefty cost—accompanied by a persistent lack of evidence of benefit. To improve the balance between access and proof, we must understand why postapproval studies often flounder. Is it due to insufficient incentives for companies to pursue rigorous trials after approval, disincentives for patients to enroll in them, or both?

To guide policy-makers in capturing the promise of early approval and to inform assessment of newly proposed pathways favored by some patient groups, we recommend careful empirical assessment of what is currently inhibiting postapproval studies and what could make them more successful.

EARLY APPROVAL PATHWAYS

Traditional FDA approval requires "substantial evidence" of effectiveness, typically through a demonstration of clinical benefit in how patients feel, function, or survive—or at least improvement on a validated surrogate endpoint. By contrast, the emergency use authorization (EUA) pathway used widely during COVID-19 allows temporary marketing on the basis that a product "may be effective," a weaker standard and one with mixed success (contrast problematic hydroxychloroquine and highly effective COVID-19 vaccines). Fortunately, EUAs terminate with the end of the public health emergency, returning companies to the baseline requirement to prove safety and effectiveness before further market access.

Another early approval pathway, accelerated approval, allows effectiveness to be demonstrated through reliance on unvalidated surrogate endpoints deemed "reasonably likely to predict clinical benefit." Accelerated approval drugs like Aduhelm can be marketed without special restrictions, but in return for accepting uncertainty at the outset, these approvals are paired with a requirement for postmarket study to confirm predicted benefit. The European Medicines Agency (EMA) offers a similar "conditional approval" pathway that relies on a limited clinical data package, followed by "specific obligations" for postmarket study.

ACCESS BEFORE EVIDENCE

Reliance on future confirmatory evidence is perilous for several reasons. First, confirmatory evidence may be long delayed. A recent study found a median time of 50 months from accelerated approval to the deadlines FDA set for sponsors to report postmarket trial results, a median of 30 months more than the duration of the trials themselves once launched (*1*). Earlier studies note substantial delays in starting postmarket trials, as well as concern about early termination (*2*). The Institute for Clinical and Economic Review (ICER) found that of 145 accelerated approvals predating 2016, 13% have not yet produced evidence to support transition to full approval (or withdrawal), despite being on the market for a median of 9.5 years (*3*). For Aduhelm, FDA gave the sponsor 9 years to submit the final report for its confirmatory trial, details of which have not yet been settled. In another controversial example, FDA granted accelerated approval to Exondys 51 (eteplirsen) for Duchenne muscular dystrophy in 2016, but the confirmatory trial is not slated to end until 2026, after a 3-year delay in initiation. Most EMA postmarketing obligations after conditional approval are also delayed.

Second, studies following accelerated approval often lack rigorous design characteristics, such as blinding, randomization, and concurrent controls. In addition, they sometimes continue to rely on surrogate endpoints (*4, 5*), rather than clinical outcomes, with the same being true in Europe. As a result, even confirmatory trials and full approval may not ensure that a drug will meaningfully improve patient health (*3, 6*).

Third, unlike EMA's conditional approvals, which must be renewed annually until specific obligations are fulfilled, FDA's accelerated approvals are not time limited. Moreover, failure to meet postapproval requirements or confirm benefit does not necessarily lead to enforcement action, such as fines or speedy withdrawal (*3*, *6*, *7*). Instead, FDA has discretion about how to proceed, potentially granting extensions or calling for more studies. For example, upon finding 10 "dangling approvals" whose required trials did not confirm benefit, sponsors voluntarily withdrew four indications, but FDA's advisory committee recommended withdrawal for only two of the remaining six (*8*). Especially when alternatives are lacking, FDA may be convinced that something is better than nothing, allowing even questionable drugs to remain on the market (*7*). Even if FDA recommends withdrawal, sponsors can push back through lengthy administrative proceedings.

WHAT'S STANDING IN THE WAY?

These shortcomings in producing and acting upon postapproval evidence—which have been attributed to insufficient resources and legal authority at FDA, as well as "the prevailing political economy" of pressure from Congress, patients, and companies (*7*)—raise questions about the balance sought by early approval pathways (*9*). One of the most critical is whether it is truly feasible to produce adequate, high-quality confirmatory evidence reasonably quickly once a product has been allowed on the market (see the table).

It's the companies

Companies may fail to produce rigorous postmarket evidence because FDA is not forcing them to do so—and Congress has not forced FDA. Because FDA's gatekeeping authority allows it to exclude new drugs from the market until their safety and effectiveness have been demonstrated, in the preapproval period companies have strong incentives to produce that evidence as quickly as possible. Moreover, FDA's author-

[1]University of Pennsylvania, Philadelphia, PA, USA.
[2]Boston University, Boston, MA, USA.
Email: lynchhf@pennmedicine.upenn.edu

ity to determine precisely what evidence will be adequate for approval and to review the details of that evidence helps ensure both quality and confidence beyond what would be possible through market forces alone (*10*).

However, once FDA approves a drug, companies begin to profit, and patient demand builds, company incentives to pursue further rigorous study drop off precipitously—future trials might fail to confirm benefit and companies already have what they want. If followed, FDA's guidance recommending that confirmatory trials be underway at the time of accelerated approval could help address this problem (*11*), but the agency often makes exceptions. FDA also does not necessarily require the parameters of confirmatory trials to be agreed upon as a condition of accelerated approval (*6*). Both types of laxity stem from FDA's reticence to further hold up promising products while postmarket study details are worked out (*6*), but they give companies leverage to "bargain down" after approval. Company incentives are further weakened by knowledge that even with problematic study designs and results, FDA may allow accelerated approvals to stand (*6*, *8*).

On this explanation, early approval pathways do not inherently sacrifice rapid, confirmatory evidence—FDA and Congress merely have to insist upon it. An appropriate policy solution would entail making early approvals time-limited with automatic withdrawal for failure to confirm benefit by a set deadline (*3*). The EMA is on the right track with its requirement for periodic renewal of conditional approvals, although that is watered down by not limiting the number of permissible renewals. A true expiration date, such as that reflected in a new provisional approval authority proposed in a bill dubbed the Promising Pathway Act (PPA, S.1644) currently pending in Congress would help address challenges to FDA's willingness to act on evidence once produced by reducing its discretion. However, for the threat of withdrawal to be meaningful, required confirmatory evidence would also have to be truly confirmatory, generated through trials with appropriate randomization, blinding, and controls, measuring meaningful health outcomes. Unfortunately, even if enacted, the PPA would not meet this standard, instead proposing reliance on "real world evidence" (RWE) and registries.

In recent years, RWE, including observational clinical data produced outside trials, has increasingly been valorized as a potential solution to some of the tension between access and evidence. So far, however, reality has not lived up to the hype, leading ICER to conclude that, at least at present, RWE likely will not suffice "in smaller patient populations with greater uncertainty surrounding the studied endpoints" (*3*). RWE can likely help complement rigorous confirmatory trials, but for diseases with a highly variable course of progression or drugs anticipated to have only modest effect, concurrent, randomized control groups are critical to determining whether observed differences can be causally attributed to the intervention.

A mother moves her son, who has Duchenne muscular dystrophy, out of bed. Accelerated approval was used to bring a Duchenne muscular dystrophy drug, Exoyndys 51, to market in 2016. Confirmatory trials are ongoing, and insurers have balked at coverage.

PHOTO: TRISTAN SPINSKI/THE NEW YORK TIMES/REDUX

It's the patients

Companies also may fail to produce rigorous postapproval trials because not enough patients are willing to participate in them. Once a product receives marketing approval, many patients facing serious disease are likely to want to try it, especially when they lack other options. If their clinicians will prescribe and insurers will pay, patients may have insufficient reason to enroll in trials in which they may be randomized to something else (*2*, *3*). This is why some initially wondered if further randomized trials of Aduhelm might be feasible only outside the US, despite major open questions.

Although typical ethical concerns about withholding proven products from trial participants can be avoided when benefit is truly uncertain, exploitation concerns remain at the prospect of allowing US patients to reap the potential benefits of early access while foisting the burdens of confirmatory trials onto patients elsewhere, especially for high-cost drugs. Patients and contexts abroad may also differ in meaningful ways from those in the US, potentially inhibiting generalizability of international trial data. For example, there may be differences in disease stage at diagnosis, standard care, and environmental factors, not all of which can be fully addressed through enrollment criteria.

The alternative to international trial sites is to gather confirmatory evidence in the US using weaker designs more amenable to recruitment, such as without randomization or use of placebo, as has been common for accelerated approval drugs in the past. It may also be possible to make rigorous trials more attractive to patients, for example, by minimizing the proportion randomized to the control group either at the outset or through response-adaptive randomization or using crossover designs that provide all participants the opportunity to receive the intervention even if initially randomized to the control group. Access to the intervention for free as part of a trial may help entice patients who would otherwise be unable to afford it, as could offering other financial incentives. However, it is not clear that these approaches would be successful, especially for patients facing fast-moving, terminal conditions who may seek any means necessary to access an intervention FDA has designated as potentially beneficial.

If patient willingness is the problem facing early approval pathways, no amount of insistence on companies producing rigorous confirmatory trials will help. In this case, one policy response could be to require confirmation of benefit before allowing a drug to be marketed—thereby rejecting early approval, and its upsides—while simultaneously supporting preapproval access though broader inclusion in clinical trials and improved use of FDA's expanded access pathway. Expanded access ("compassionate use") allows individuals or groups of patients with serious or life-threatening conditions to access unapproved drugs outside trials when they lack alternatives and such access will not interfere with a product's testing and development. As currently conceived, however, expanded access serves far fewer patients than marketing approval, owing to a variety of logistical hurdles, including clinician

Typical interests and influences of key stakeholders before and after approval of a new drug

	FOOD AND DRUG ADMINISTRATION (FDA)	DRUGMAKERS	PATIENTS	PAYERS
Preapproval	• Demand proof of safety and effectiveness (and access to underlying data) to meet statutory approval standards • Consider patient and political pressure for access to products with potential clinical benefit	• Perform minimum necessary studies (size, length, endpoints) to support FDA approval and payer reimbursement • Develop relationships with patients and FDA to support rapid approval to maximize intellectual property protections and profit	• Seek rapid access to potentially life-saving products, including through participation in randomized, placebo-controlled trials, if necessary • In some cases, push for early approval with later confirmation of benefit • May receive drugmaker support for advocacy organizations	• Refuse to pay for "experimental" products
Postapproval	• Face political barriers to enforcement action, including withdrawing approval, especially absent alternative therapies, even if benefit is not confirmed	• Maximize return on investment through high prices and efforts to encourage use • Avoid trials that could raise questions about approved indication; have low incentive for rigor or speed • Pursue studies to support sales, especially for off-label indications	• Pursue approved, reimbursed therapeutic options over participation in randomized trials • May receive funds from drugmakers for advocacy and to cover access costs, if other payers are unavailable or decline	• Cover "medically necessary" care, typically aligned with FDA approval • Can pursue mechanisms to avoid or limit coverage of high-priced products supported by weak evidence, but constrained by political, legal, and market pressures

awareness and burden, as well as economic challenges about who should pay for treatment use of investigational drugs (*12*).

FINANCIAL GATEKEEPING

In contrast to other countries, where drug approval is often followed by national decisions about whether to pay for those drugs, US payment decisions rarely diverge from FDA approvals. For Aduhelm, however, several prominent hospital systems have declined to use the drug, and major payers have hesitated to extend coverage or announced restrictive conditions for payment.

US laws and insurance contracts already allow payers to exercise restraint on covering products that lack evidence of clinical benefit. However, some American payers have a "pass through" mentality, in which they lack incentives to meaningfully police the value of products and services that they reimburse. In the short run, saying "yes" minimizes public relations problems raised by sympathetic patients, even if it inflates government-subsidized insurance premiums and national debt in the long run. Saying "no" is harder.

The solution here may be to reconceive coverage decisions as more-or-less, rather than yes-or-no. The Center for Medicare and Medicaid Services has authority to implement "coverage with evidence development" (CED), whereby it covers items or services only for patients participating in a clinical study or registry, a process currently being considered for Aduhelm. However, to date, CED has been infrequently used and rarely involves prospective randomized controlled trials. More generally, a government mandate for scaled reimbursement or adjusting rebates based on the quality of evidence could create incentives both for drugmakers to develop rigorous evidence and for patients to enroll in trials.

EVIDENCE-BASED POLICY

Continued acceptance of early approval pathways should be based on evidence regarding whether, when, and how meaningful postapproval trials will be possible. As the US Department of Health and Human Services Office of the Inspector General (OIG) undertakes its review of the accelerated approval pathway, we recommend that it prioritize generation of this evidence.

One approach OIG should consider is examining the details of trials that have further evaluated a drug's approved indication after marketing approval, looking to distinguish the features of trials that were more versus less successful. Several reviews have studied postmarket trials, providing valuable insight on broad characteristics such as study type, indication, and population; trial size, duration, and endpoints; use of randomization, blinding, and comparator groups; and trial registration, completion, and results reporting (*4, 5, 13–15*). However, other essential questions have received far less attention but may be relevant to the conditions under which patients are willing to enroll and trials able to complete. How serious and fast-moving is the disease? Are there other approved products for that indication? What were the target recruitment numbers compared to final enrollment? How many sites participated in the trial and how accessible were those sites? Were sites international? Was placebo used? What was the randomization ratio? Were participants given the opportunity to cross over from control to intervention? Were any financial incentives offered? It could also be valuable to examine the perspectives of patients, company executives, and regulators through qualitative methods such as interviews (*7*) and focus groups about how best to encourage rigor, enrollment, and completion of these studies, as well as the barriers and disincentives to doing so. Important patterns may emerge from this type of analysis that could help inform both drugmakers and FDA when deciding how to design postmarket trials and whether to grant early approval based on their likely feasibility.

Accelerated approval is an important regulatory pathway worth trying to save, if the evidence suggests that meaningful improvements in confirmatory trials are possible. While this evidence is gathered, companies, patients, and policy-makers should prioritize efforts to meaningfully improve access to investigational products in the preapproval period, without further pushing the boundaries of accelerated approval. ■

REFERENCES AND NOTES

1. J. D. Wallach, R. Ramachandran, T. Bruckner, J. S. Ross, *JAMA Netw. Open* **4**, e2133601 (2021).
2. J. D. Wallach, J. S. Ross, H. Naci, *Clin. Trials* **15**, 219 (2018).
3. Institute for Clinical and Economic Review, "Strengthening the FDA's Accelerated Approval Pathway" (2021); https://icer.org/assessment/fda-accelerated-approval-pathway/.
4. H. Naci, K. R. Smalley, A. S. Kesselheim, *JAMA* **318**, 626 (2017).
5. B. Gyawali, S. P. Hey, A. S. Kesselheim, *JAMA Intern. Med.* **179**, 906 (2019).
6. B. Gyawali, J. S. Ross, A. S. Kesselheim, *JAMA Intern. Med.* **181**, 1275 (2021).
7. M. Herder, *Milbank Q.* **97**, 820 (2019).
8. K. Powell, M. P. Lythgoe, V. Prasad, *JAMA Oncol.* (2021).
9. H. F. Lynch, A. Bateman-House, *J. Law Med. Ethics* **48**, 365 (2020).
10. A. Kapczynski, *Minn. Law Rev.* **102**, 2357 (2018).
11. US Food and Drug Administration, "Guidance for Industry: Expedited programs for serious conditions–Drugs and biologics" (2014); www.fda.gov/regulatory-information/search-fda-guidance-documents/expedited-programs-serious-conditions-drugs-and-biologics.
12. K. M. Folkers, A. Bateman-House, C. Robertson, *Wake Forest J. Law Policy* **11**, 85 (2020).
13. J. J. Skydel, A. D. Zhang, S. S. Dhruva, J. S. Ross, J. D. Wallach, *Clin. Trials* **18**, 488 (2021).
14. J. D. Wallach, A. T. Luxkaranayagam, S. S. Dhruva, J. E. Miller, J. S. Ross, *BMC Med.* **17**, 117 (2019).
15. J. D. Wallach *et al.*, *BMJ* **361**, k2031 (2018).

ACKNOWLEDGMENTS

The authors are members of the NYU Grossman School of Medicine Division of Medical Ethics Working Group on Compassionate Use and Preapproval Access (CUPA).

10.1126/science.abk3495

Reducing wildlife-vehicle collisions should increase road safety and help countries achieve sustainable transport systems.

Edited by **Jennifer Sills**

Wildlife collisions put a dent in road safety

In September 2020, the United Nations (UN) adopted a resolution to halve road traffic deaths and injuries by 2030 (*1*). A Global Plan for the Decade of Action for Road Safety (*2*) was developed emphasizing the importance of a holistic approach to road safety, including the improvement of the design of roads and vehicles and the enhancement of laws. However, the Global Plan overlooks wildlife-vehicle collisions.

Millions of wildlife-vehicle collisions occur every day around the world (*3*, *4*), causing countless human deaths and injuries as well as high vehicle damage costs (*5*). The collisions also threaten biodiversity (*4*, *6*), given that virtually all species that live in the vicinity of roads are at risk. Hence, reducing wildlife-vehicle collisions goes hand in hand with improving road safety, while also promoting biodiversity conservation.

In line with the UN resolution, we suggest promoting simple but effective mitigation actions. Installation of proper fencing linked to existing road-crossing structures (such as culverts for water drainage) can prevent wildlife crossings on the road (*7*). Given that most wildlife-vehicle collisions occur at night, nocturnal traffic should be substantially decreased or slowed, where possible. Such measures should be integrated into road construction and mitigation legislation. More ambitiously, the car industry should target the integration of artificial intelligence technology, which could help to prevent collisions by automatically detecting animals approaching the road and warning the driver in real time (*8*). Adopting plans and policies to reduce wildlife-vehicle collisions will be a step toward meeting the UN goal of a sustainable transport system.

Fernando Ascensão[1]*, Rafael Barrientos[2], Marcello D'Amico[3]
[1]Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal. [2]Road Ecology Lab, Department of Biodiversity, Ecology and Evolution, Faculty of Biology, Complutense University of Madrid, E-28040 Madrid, Spain. [3]Department of Conservation Biology, Doñana Biological Station, Spanish National Research Council, Seville, Spain.
*Corresponding author. Email: fjascensao@fc.ul.pt

REFERENCES AND NOTES

1. UN General Assembly, "Improving global road safety" (2020); https://undocs.org/en/A/RES/74/299.
2. World Health Organization, "Global Plan for the Decade of Action for Road Safety 2021–2030" (2021); www.who.int/publications/m/item/global-plan-for-the-decade-of-action-for-road-safety-2021-2030.
3. C. Grilo, E. Koroleva, R. Andrášik, M. Bíl, M. González-Suárez, *Front. Ecol. Environ.* **18**, 323 (2020).
4. R. Barrientos, F. Ascensão, M. D'Amico, C. Grilo, H. M. Pereira, *Perspect. Ecol. Conserv.* **19**, 411 (2021).
5. M. R. Conover, *Human–Wildlife Interact.* **13**, 12 (2019).
6. Van der Ree, D. J. Smith, C. Grilo, *Handbook of Road Ecology* (John Wiley & Sons, 2015).
7. D. Lesbarrères, L. Fahrig, *Trends Ecol. Evol.* **27**, 374 (2012).
8. F. Ascensão, C. Branquinho, E. Revilla, *Nat. Electron.* **3**, 295 (2020).

10.1126/science.abm8468

Rethink roads through the Chaco Serrano forest

Despite global initiatives to protect ecosystems (*1*), human development continues to threaten the environment. In central Argentina, the subtropical dry Chaco Serrano forest has lost 94% of its original area (*2*) to the economic growth of the region led by agriculture, urbanization, and mining (*2*, *3*). Now, plans for new road infrastructure (*4*) suggest that the Chaco Serrano forest will continue to shrink, further fragmenting the forest ecosystem. The Córdoba provincial government has granted an environmental license for the construction of two additional roads that will cut through areas of Chaco Serrano forest in the Punilla and Paravachasca valleys. These recently burned forest areas of high conservation priority, according to national laws (*5*, *6*), should be protected without exception.

These roads have been advertised as beneficial for tourism, transportation, and commercial and urban development. However, most speakers at the nonbinding public hearings (including people from local communities, civil organizations, neighborhood assemblies, and Indigenous peoples, as well as small landholders and scientists) rejected these projects (*7*). Land speculation will likely increase deforestation, while the more pressing priorities of the region, such as water access, schools, hospitals, and fire brigades, are completely ignored. Moreover,

PHOTO: TECHNOTR/ISTOCK.COM

the imposition of road development projects undermines communities that aspire to protect nature and strive for community self-determination (*3*, *8*).

Despite humanity's profound dependence on nature (*9*, *10*), government policies and market transactions are still based on maximizing economic gains through capitalist extractivism. Even in cases where national laws rightly protect the environment, damaging policies are approved through loopholes and workarounds. We urge local decision-makers, as well as the foreign governments that finance such projects (*11*), to work toward a sustainable way of life, with real participation of communities and the recognition of the rights of nature (*12*). We must guarantee the conservation of the scarce Chaco Serrano forest, on which the people of Argentina depend.

Guadalupe Peralta[1]*, Francisco M. F. Bertea[2], Luciano Cagnolo[1]
[1]Laboratorio de Interacciones Insecto-Planta, Instituto Multidisciplinario de Biología Vegetal, Consejo Nacional de Investigaciones Científicas y Técnicas and Universidad Nacional de Córdoba (CONICET-UNC), Córdoba, Argentina. [2]Centro de Investigaciones y Estudios sobre Cultura y Sociedad, CONICET-UNC, Córdoba, Argentina.
*Corresponding author.
Email: gdlp.peralta@gmail.com

REFERENCES AND NOTES

1. UN Environment Programme, "New UN Decade on Ecosystem Restoration offers unparalleled opportunity for job creation, food security, and addressing climate change" (2019).
2. M. R. Zak, M. Cabido, J. G. Hodgson, *Biol. Conserv.* **120**, 589 (2004).
3. J. U. Deon, *Iconos Rev. Cienc. Soc.* **70**, 151 (2021) [in Spanish].
4. Legislatura de Córdoba, "Land for road works declared of public utility" (2020); https://legislaturacba.gob.ar/declaran-de-utilidad-publica-terrenos-para-obras-viales/ [in Spanish].
5. Law 26.331, "Minimum Budgets for the Environmental Protection of Native Forests" (2007).
6. Law 27.604, "Fire management" (2020).
7. G. Pedraza, "Autovía Paravachasca: Entre gallos, medianoche, y a un costo insólito," *La Tinta* (2021); https://latinta.com.ar/2021/07/autovia-paravachasca/ [in Spanish].
8. A. Escobar, *Sustain. Sci.* **10**, 451 (2015).
9. D. J. Haraway, *Staying with the Trouble: Making Kin in the Chthuluceno* (Duke University Press, 2016).
10. S. Díaz *et al.*, *Science* **359**, 270 (2018).
11. "Decreto 295/2021," Boletín Oficial de la República Argentina (2021); www.boletinoficial.gob.ar/detalleAviso/primera/243920/20210504 [in Spanish].
12. A. Acosta, E. Martínez, *La Naturaleza con Derechos: De la Filosofía a la Política* (Adya-Yala, Quito, 2011).

10.1126/science.abm7988

Sri Lanka's hasty agrochemical ban

Rice, Sri Lanka's major staple crop, is grown in 40% of the country's arable lands (*1*). More than 90% of Sri Lankan farmers use agrochemicals for rice cultivation (*2*). In May, the Sri Lankan government banned the production and importation of agrochemicals, including artificial fertilizers, pesticides, and weedicides, and ordered an immediate and complete switch to organic agriculture (*3*, *4*) in the name of ecological and human health (*3*). Because the agriculture sector in Sri Lanka was unprepared for this sudden policy change, farmers protested the government's decision in fear of yield and income loss (*5*). Although there are good reasons to reduce the use of agrochemicals (*6*), to protect humans who depend on the affected crops, the changes must be made gradually.

Sri Lankan farmers were not given enough time or support to make the switch to all organic materials between the announcement in May and the start of the rice growing season in October. There is still no standardized supply of organic substitutes available to farmers, who have had to make difficult decisions about how to proceed without the approved materials. Exacerbating the situation, imported organic fertilizer from China was found to have harmful bacteria (*Erwinia* and *Bacillus*, pathogenic to plants) and therefore could not be used (*7*).

Although organic agriculture may benefit biodiversity and reduce contamination of ecosystems (*8*), such strategies also risk lower yield and higher food insecurity (*9*). Sri Lanka is already at high risk of food insecurity due to disruption to cultivation and harvesting during the COVID-19 pandemic lockdowns (*10*) and growing season changes caused by climate change (*11*). Given these challenges, Sri Lankan leaders must take people's safety into account as they create environmental legislation. Agricultural policies should aim to ensure food security and farmers' livelihoods. Gradually reducing the volume of agrochemical usage and redesigning landscapes to promote diversified farming practices could be a cost-effective way to balance yield and farmers' income with biodiversity and ecosystem health (*12*).

Tharaka S. Priyadarshana
Asian School of the Environment, Nanyang Technological University, Singapore City, Singapore. E-mail: tharakas001@ntu.edu.sg

REFERENCES AND NOTES

1. US Department of Agriculture, "Grain and feed annual: Sri Lanka" (2020); https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Grain and Feed Annual_New Delhi_Sri Lanka_03-27-2020.
2. VeritéResearch, "Sri Lanka's first independent island-wide farmers' survey on the government's chemical fertiliser ban" (2021); www.veriteresearch.org/farmerspulse/.
3. Presidential Secretariat, "Importation of chemical fertilizers will be stopped completely" (2021); www.presidentsoffice.gov.lk/index.php/2021/04/22/importation-of-chemical-fertilizers-will-be-stopped-completely/.
4. US Department of Agriculture, "Sri Lanka restricts and bans the import of fertilizers and agrochemicals" (2021); https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Sri Lanka Restricts and Bans the Import of Fertilizers and Agrochemicals_New Delhi_Sri Lanka_05-14-2021.pdf.
5. Z. Farzan, "Protests grow as Sri Lanka's farmers oppose chemical fertilizer ban," *NewsFirst* (2021); www.newsfirst.lk/2021/10/18/protests-grow-as-sri-lankas-farmers-oppose-chemical-fertilizer-ban/.
6. W. Aktar, D. Sengupta, A. Chowdhury, *Interdiscip. Toxicol.* **2**, 1 (2009).
7. S. Rubatheesan, "Maha season begins, but farmers are not in the field," *The Sunday Times* (2021); www.sundaytimes.lk/211017/news/maha-season-begins-but-farmers-are-not-in-the-field-458876.html.
8. V. Seufert, N. Ramankutty, *Sci. Adv.* **3**, e1602638 (2017).
9. E.-M. Meemken, M. Qaim, *Annu. Rev. Resour. Econ.* **10**, 39 (2018).
10. N. Singh *et al.*, *Curr. Dev. Nutr.* **5**, 248 (2021).
11. W. Karunarathne, "How Sri Lankan farmers experience climate change," SLYCAN Trust (2021); www.preventionweb.net/news/how-sri-lankan-farmers-experience-climate-change.
12. T. Tscharntke, I. Grass, T. C. Wanger, C. Westphal, P. Batáry, *Trends Ecol. Evol.* **36**, 919 (2021).

10.1126/science.abm9186

TECHNICAL COMMENT ABSTRACTS

Comment on "A Middle Pleistocene *Homo* from Nesher Ramla, Israel"

Assaf Marom and Yoel Rak

Hershkovitz *et al.* (Report, 25 June 2021, p. 1424) conclude that the Nesher Ramla (NR) fossils represent a distinctive *Homo* paleodeme that played a role as a source population for Neanderthals. However, the highly diagnostic features of the Neanderthal mandible—clearly displayed by the NR fossils—are largely overlooked. Our analyses indicate that the NR fossils represent simply a Neanderthal.

Full text: dx.doi.org/10.1126/science.abl4336

Response to Comment on "A Middle Pleistocene *Homo* from Nesher Ramla, Israel"

Hila May, Rachel Sarig, Ariel Pokhojaev, Cinzia Fornai, María Martinón-Torres, José María Bermúdez de Castro, Gerhard W. Weber, Yossi Zaidner, Israel Hershkovitz

Marom and Rak claim, on the basis of a few mandibular features, that the Nesher Ramla (NR) *Homo* is a Neanderthal. Their comments lack substance and contribute little to the debate surrounding the evolution of Middle Pleistocene *Homo*. Limitations and preconceptions in their study prevented them from achieving resolution beyond a dichotomous interpretation of the NR as either a Neanderthal or a modern human.

Full text: dx.doi.org/10.1126/science.abl5789

ERRATA

Erratum for the Research Article "The genetic architecture of the human cerebral cortex," by K. L. Grasby *et al.*, *Science* 374, eabm7211 (2021). Published online 22 October 2021; 10.1126/science.abm7211

Erratum for the Report "Hierarchical crack buffering triples ductility in eutectic herringbone high-entropy alloys," by P. Shi *et al.*, *Science* 374, eabm7979 (2021). Published online 15 October 2021; 10.1126/science.abm7979

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CARBON CYCLE

Up in the air

Understanding ocean-atmospheric carbon dioxide (CO_2) fluxes in the Southern Ocean is necessary for quantifying the global CO_2 budget, but measurements in the harsh conditions there make collecting good data difficult, so a quantitative picture still is out of reach. Long *et al*. present measurements of atmospheric CO_2 concentrations made by aircraft and show that the annual net flux of carbon into the ocean south of 45°S is large, with stronger summertime uptake and less wintertime outgassing than other recent observations have indicated. —HJS *Science*, abi4355, this issue p. 1275

Aerial surveys above the ocean around Antarctica indicate greater net annual uptake of carbon dioxide than was found by previous measurements.

MECHANICAL BONDS

Pumping macrocycles onto surfaces

Numerous chemical processes, ranging from water purification to catalysis, involve sorption of small molecules onto surfaces. Typically, spontaneous attractive interactions favor the binding event. Feng *et al*. report a mechanisorption process that requires redox manipulations to pump macrocycles from bulk solution onto axles immobilized on a metal-organic framework. The resulting rotaxanes store energy through nonequilibrium charge concentration in their mechanical bonds. Ultimately, the technique could also prove useful for actively partitioning compounds with particular functionality between surface and bulk environments. —JSY

Science, abk1391, this issue p. 1215

EXOPLANETS

A nearby iron-rich sub-Earth planet

The mass and radius of an exoplanet determine its mean density, which provides information about the possible interior structure. Lam *et al*. have identified a planet on a 7.7-hour orbit around a nearby red dwarf star. The authors determined the planet's radius from the transit, then used radial velocity observations to measure the mass. They found a sub-Earth–sized planet with a density almost equivalent to pure iron. Its high surface temperature is close to the vaporization point of iron, suggesting that it is the iron core of a planet that has lost its outer mantle. —KTS

Science, aay3253, this issue p. 1271

PHYSIOLOGY

Controlling cardiomyocyte proliferation

Mammalian cardiomyocytes proliferate during development but lose this capacity in adult life, and this underlies the inability of the heart to regenerate. This limitation is related to cardiomyocyte superspecialization as efficient contractile cells. Tampakakis *et al.* found that sympathetic neurons directly repress cardiomyocyte proliferation in mice through the activation of transcription factors that regulate circadian rhythm. Cardiac sympathetic innervation develops during fetal life as cardiomyocytes increase in size and contractile capacity. This maturation process is further exacerbated postnatally, leading to proliferation-incompetent cardiomyocytes. This study expands our understanding of the barriers that restrict cardiomyocyte proliferation and provides new opportunities for exploring the reactivation of heart regeneration in mammals. —MT

Sci. Adv. 10.1126/sciadv.abh4181 (2021).

HUMAN GENOMICS

Rare variants and blood LDL cholesterol

A current goal in genomics is to identify genetic variation associated with actionable traits of clinical concern. Through exome sequencing of an Old Order

PHOTO: JONATHAN BENT

Amish population, Montasser *et al*. identified a genetic variant that results in an amino acid change in the beta-1,4-galactosyltransferase 1 protein and is correlated with lower levels of cardiovascular disease. Investigation of the mutant protein showed that it affects genes associated with low-density lipoprotein cholesterol (LDL-C), and mice engineered to express the mutant protein exhibited a 38% decrease in blood LDL-C levels. This study suggests that such genomic sequencing and analysis can link genotype to phenotype and identify potentially clinically actionable pathways to treat disease. —LMZ

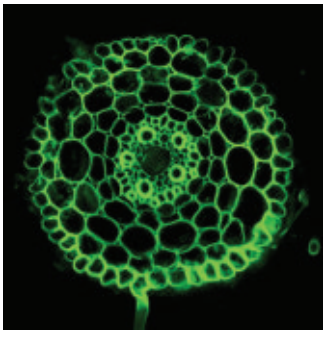
Science, abe0348, this issue p. 1215

PLANT SCIENCE

Dyed roots reveal inner complexity

Plant roots do so much more than just hold a plant up. As a site for air storage during flooding, mycorrhizal symbiosis, or carbohydrate storage, the more complex root can tap more complicated functions. Taking advantage of a dye that stains less the deeper it penetrates the tissue, Ortiz-Ramírez *et al*. applied fluorescence-activated cell sorting to the complex cell layers of the maize root. RNA sequencing applied to the single-cell pools defined a developmental map and showed that the mobile transcription factor SHORT-ROOT travels through multiple cell layers and directs this grass root's anatomical complexity. —PJH

Science, abj2327, this issue p. 1247



A fluorescence microscopy cross section reveals structured layers of cells in the root tissue of the grass *Setaria viridis*.

PHOTOS: (LEFT TO RIGHT) ORTIZ-RAMÍREZ *ET AL*.; ALVAREZ/ISTOCKPHOTO.COM

BIOCHEMISTRY

Revealing the oncogenic effects of RIT1

Various mutant forms of the signaling protein RAS underlie multiple types of cancers, including lung cancers. Mutations or chromosomal amplifications in the RAS-related protein RIT1 have been found in about 15% of lung adenocarcinomas. Lo *et al*. performed comparative multiomics in lung epithelial cells expressing mutant or wild-type forms of RIT1 or RAS. In contrast to RAS, which promoted oncogenic and metastasis-associated changes when mutated, RIT1 promoted these changes whether mutated or not. The findings suggest that overexpression of wild-type RIT1 may be an unappreciated driver in tumor development. —LKF

Sci. Signal. **14**, eabc4520 (2021).

MOLECULAR BIOLOGY

Maturation for survival

Although cells with defects in DNA replication usually die under stress conditions, some cells acquire new mutations and survive. Sun *et al*. identified an error-prone, stress-induced Okazaki fragment maturation pathway that induces tandem duplications and enables the survival of cells that have defects in removing the 5′ RNA-DNA flap during DNA replication. In these cells, stress conditions activate DUN1 signaling and induce conversion of the 5′ flap to a 3′ flap that can form secondary structures and be extended and ligated to the downstream DNA fragment, generating alternative duplication mutations similar to the ones in human cancers. The revealed information is analogous to the mechanism in cancer cell evolution and drug resistance. —DJ

Science, abj1013, this issue p. 1252

IN OTHER JOURNALS

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

NEUROSCIENCE

Echoes of speech in the brain

The brain's response to people talking is shifted by the listener paying attention and understanding the speech. The brain still tracks sounds that the listener is trying to ignore. The ability to track the neural response to speech may help in settings where a person is unable to respond in other ways. Gillis *et al.* tracked neural responses to the linguistic aspects of speech while controlling for the acoustic aspects of the speech. Young adult subjects were asked to pay attention to audiobook stories read by the same speaker as electroencephalography (EEG) data were collected. The results show the brain EEG reflecting linguistic aspects of speech over and above the sounds of speech. —PJH

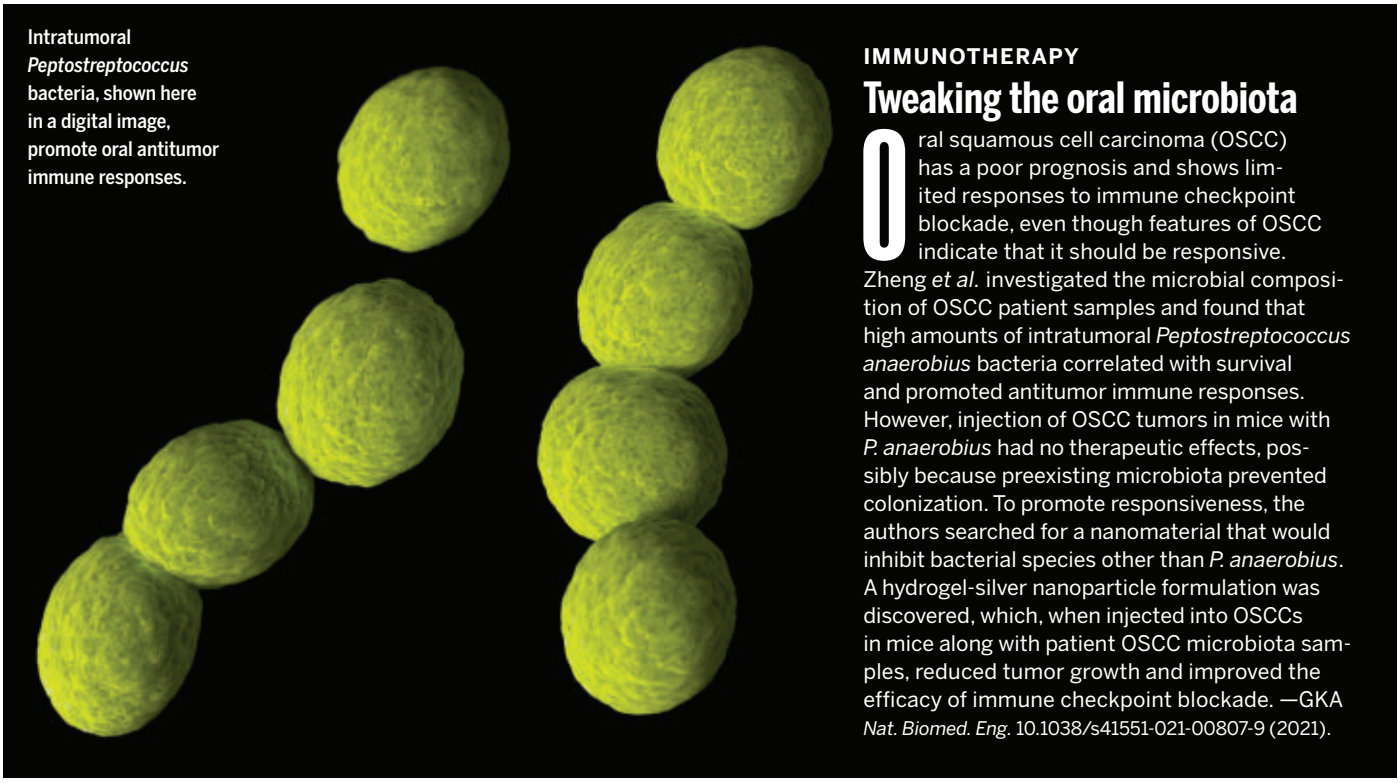
J. Neurosci. 10.1523/jneurosci.0812-21.2021 (2021).

Listening to audiobooks reveals how the brain separates linguistic aspects of speech from the sounds of speech.

SIGNALING

Explaining fast receptor signaling

Stimulation of the small guanosine triphosphatase Ras, which is activated in response to receptor activation at the cell surface, can occur surprisingly rapidly. The time for activation of the guanine nucleotide exchange factor Son of Sevenless (SOS), which activates Ras, has been estimated to take about a minute. In cells, however, activation occurs in only a few seconds. Mathematical modeling and in vitro experiments by Huang *et al.* showed that this reflects the processive activation of multiple SOS molecules, with one SOS molecule activating hundreds of Ras molecules. Thus, the



Intratumoral *Peptostreptococcus* bacteria, shown here in a digital image, promote oral antitumor immune responses.

IMMUNOTHERAPY

Tweaking the oral microbiota

Oral squamous cell carcinoma (OSCC) has a poor prognosis and shows limited responses to immune checkpoint blockade, even though features of OSCC indicate that it should be responsive. Zheng *et al.* investigated the microbial composition of OSCC patient samples and found that high amounts of intratumoral *Peptostreptococcus anaerobius* bacteria correlated with survival and promoted antitumor immune responses. However, injection of OSCC tumors in mice with *P. anaerobius* had no therapeutic effects, possibly because preexisting microbiota prevented colonization. To promote responsiveness, the authors searched for a nanomaterial that would inhibit bacterial species other than *P. anaerobius*. A hydrogel-silver nanoparticle formulation was discovered, which, when injected into OSCCs in mice along with patient OSCC microbiota samples, reduced tumor growth and improved the efficacy of immune checkpoint blockade. —GKA

Nat. Biomed. Eng. 10.1038/s41551-021-00807-9 (2021).

earliest activated SOS molecules initiate signaling more quickly than molecular kinetics would indicate and may be a common feature in other cellular signaling systems. —LBR

Proc. Natl. Acad. Sci. U.S.A. **118**, e2103598118 (2021).

DISEASE GENOMICS

Driver genes differ in type 2 diabetes

Type 2 diabetes (T2D) is characterized by high blood glucose levels resulting from cellular insulin resistance. Untreated, insulin resistance can lead to serious health problems. T2D has been the object of many genetic analyses to determine risk genes or genetic variants across populations. Mansour Aly *et al.* delved into the clinical variables associated with T2D in a Swedish cohort of adult patients. Five clusters of clinical characteristics, disease progression, and outcome became apparent. Genetic and familial analyses showed different heritability among the T2D subtypes. Genome-wide analyses coupled with phenotype, such as body mass index, were used to determine a polygenic risk score. From this, three severe T2D subtypes with distinct genetic backgrounds and pathology were identified, emphasizing how more thorough phenotyping can illuminate how genetic risk contributes to polygenic disease. —LMZ

Nat. Genet. **53**, 1534 (2021).

TEXTILES

A powerful cloth

A textile that can function as an energy harvester would be of great interest for powering things such as wearable electronic devices. Wang *et al.* developed a scalable and flexible magnetoelectrical clothing generator that works by swinging the arms. Magnetic fibers were developed by directly mixing magnetic particles into the polymer. These fibers were used in the arms of a jacket to enable a current to be generated in coils integrated into the jacket sides. The power generated was sufficient to light several hundred light-emitting diodes or charge up a commercial capacitor. —BG

Adv. Funct. Mater. 10.1002/adfm.202107682 (2021).

INTEGRATED PHOTONICS

Local fixes for global performance

Complex networks of optical components patterned into silicon dioxide are used to make beam splitters, phase shifters, interferometers, and logic gates, and are finding application in optical computing, communication, and machine learning. However, small-scale errors caused by fabrication imprecision in the individual components propagate through the system and can compromise global functionality. Bandyopadhyay *et al.* present an approach to mitigate for these hardware errors by homing in on the individual components and correcting them locally. By breaking down the network into individual parts, they characterized the performance and function of each component and then corrected for any errors that might be found. Reassembly resulted in resilient global performance. The approach provides a route to scale up integrated photonics, allowing the fabrication of large-scale and complex circuits in optical chips. —ISO

Optica **8**, 1247 (2021).

PLASTIC WASTE

A cost of contagion

Plastic pollution is a serious environmental problem, and the increased use of single-use plastics caused by the COVID-19 pandemic has made it even worse. Peng *et al.* report that the pandemic has resulted in the generation of more than eight million tons of plastic waste, which is 1 to 2% of the global total. More than 25,000 tons of this waste has entered the ocean. Hospital waste accounts for nearly 75% of the global discharge, and Asia is responsible for more than 70% of the total. The fraction contributed by personal protection equipment and online-shopping packaging is minor. —HJS

Proc. Natl. Acad. Sci. U.S.A. **118**, e2111530118 (2021).

PHOTO: SCIEPRO/SCIENCE SOURCE

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

Understanding the earliest cases in Wuhan

The earliest symptomatic cases of COVID-19 in Wuhan, China, could be informative about the origins of the pandemic and suggest ways of detecting unknown diseases sooner. In a Perspective, Worobey draws from diverse sources to put together a timeline of the earliest cases of COVID-19, the connections of these patients to the Huanan market that sold wild animals, and the events that led to the identification of COVID-19 as a new viral pneumonia. This analysis demonstrates that most of the first cases were associated with the market, indicating that spillover from susceptible animals was the likely source of community transmission. —GKA

Science, abm4454, this issue p. 1202

DEVELOPMENTAL BIOLOGY

The nonconserved primitive streak

In human development, a linear structure called the primitive streak appears 14 days after fertilization. This structure marks the transition of the embryo from having radial to bilateral symmetry. The primitive streak also gives anterior-posterior and dorsal-ventral spatial information to cells undergoing gastrulation and forming the various body cell types. In a Review, Sheng *et al.* present a phylogenetic and ontogenetic overview of the primitive streak. They discuss organismal, cellular, and molecular features of the primitive streak and how it functions in amniote gastrulation. The observation that this structure is not conserved and is not required for development in vitro has implications for embryonic stem cell–based models and considerations about human development research. —BAP

Science, abg1727, this issue p. 1213

CORONAVIRUS

Immune memory after vaccination

Vaccination against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has proven highly effective at preventing severe COVID-19. However, the evolution of viral variants and waning antibody levels over time raise questions regarding the longevity of vaccine-induced immune protection. Goel *et al.* examined B and T lymphocyte responses in individuals who received SARS-CoV-2 messenger RNA vaccines. They performed a 6-month longitudinal study of individuals who never had SARS-CoV-2 infection compared with people who had recovered from SARS-CoV-2. Humoral and cellular immune memory was observed in vaccinated individuals, as were functional immune responses against the Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2) viral variants. Analysis of T cell activity suggested that robust cellular immune memory may prevent hospitalization by limiting the development of severe disease. —PNK

Science, abm0829, this issue p. 1214

METABOLISM

Reversing the chain

The mitochondrial electron transport chain is a major part of cellular metabolism and plays key roles in both cellular respiration and the synthesis of critical metabolites. Typically, electrons flow through the electron transport chain in a specific direction, ending up with oxygen as the terminal electron acceptor. Spinelli *et al.* characterized an alternative path of electron flow through the transport chain, ending with fumarate as the electron acceptor (see the Perspective by Baksh and Finley). This pathway operates under conditions of limited oxygen availability, and the authors have confirmed its activity in vivo in a mouse model, observing that the propensity to use this pathway varied between organs. —YN

Science, abi7495, this issue p. 1227; see also abm8098, p. 1196

TOPOLOGICAL MATTER

Synthesizing topological order

Topologically ordered matter exhibits long-range quantum entanglement. However, measuring this entanglement in real materials is extremely tricky. Now, two groups take a different approach and turn to synthetic systems to engineer the topological order of the so-called toric code type (see the Perspective by Bartlett). Satzinger *et al.* used a quantum processor to study the ground state and excitations of the toric code. Semeghini *et al.* detected signatures of a toric code–type quantum spin liquid in a two-dimensional array of Rydberg atoms held in optical tweezers. —JS

Science, abi8378, abi8794, this issue p. 1237, this issue p. 1242; see also abl8910, p. 1200

ORGANIC CHEMISTRY

Radical substitution

Nucleophilic substitution is a venerable reaction in organic chemistry. Typically, an incoming ion delivers two electrons to a carbon center while a departing ion takes two electrons away with it. The one-electron analog homolytic substitution is used less often, in part because the incoming neutral radicals can self-couple instead of bonding to the intended target. Liu *et al.* report that an iron porphyrin catalyst can direct homolytic substitution between primary and tertiary carbon radicals by selectively activating the primary partners. —JSY

Science, abl4322, this issue p. 1258

NANOPHOTONICS

Optomechanical upconversion

Molecules have rich signatures in their spectra at infrared wavelengths and are typically accessed with dedicated spectroscopic instrumentation. Chen *et al.* and Xomalis *et al.* report optomechanical frequency upconversion from the mid-infrared to the visible domain using molecular vibrations coupled to a plasmonic nanocavity at ambient conditions (see the Perspective by Gordon). Using different nanoantenna designs, one with a nanoparticle-on-resonator and the other with nanoparticle-in-groove, both approaches show the ability to upconvert the mid-infrared vibrations of the molecules in the nanocavity to visible light wavelengths. The effect could be used to simplify infrared spectroscopy, possibly with single-molecule sensitivity. —ISO

Science, abk3106, abk2593, this issue p. 1264, this issue p. 1268; see also abm4252, p. 1201

ORGANIC CHEMISTRY

Desaturating acids

Molecular motifs with neighboring C=C and C=O double bonds are central to the synthesis of fine chemicals, pharmaceuticals, and polymers. Wang *et al.* report a versatile palladium-catalyzed reaction to produce this motif from carboxylic acids by oxidation of the adjacent carbons, with molecular oxygen as a viable terminal oxidant (see the Perspective by Iwabuchi). The method relies on fine-tuned ligand geometry and complements approaches constrained by their reliance on enolate chemistry. Swapping in an alternate ligand promotes extension of the desaturation to coupling with terminal alkynes. —JSY

Science, abl3939, this issue p. 1281; see also abm4457, p. 1199

ALZHEIMER'S DISEASE

Inhibiting AKTivity

Triggering receptor expressed on myeloid cells 2 (TREM2) has been implicated in the pathophysiology of Alzheimer's disease (AD), and the R47H variant of the *TREM2* gene is associated with increased risk of late-onset AD. Sayed *et al.* studied the mechanisms responsible for the detrimental effects of the variant. R47H-*TREM2* results in increased expression of inflammatory molecules and enhanced AKT signaling in microglia from patients. The results were replicated in vivo using a knock-in mouse model. Moreover, the mutation worsened AD-like pathology and cognitive deficits in a model of tauopathy. Inhibiting AKT signaling had therapeutic effects in mice, suggesting that targeting AKT might be effective in patients with AD carrying the R47H mutation. —MM

Sci. Transl. Med. **13**, eabe3947 (2021).

PSORIASIS

Amplification of skin inflammation

The proteins BCL10 and MALT1 form ubiquitously expressed signalosomes that control immune and inflammatory pathways in many tissues. Pathological interplay between the immune system and keratinocytes is known to drive pathogenesis in psoriasis; however, the molecular and cellular functions of BCL10 and MALT1 in the complex pathogenesis of psoriasis are not well defined. Kurgyis *et al.* engineered mouse models to activate, inactivate, or attenuate BCL10/MALT1 signaling specifically in keratinocytes. They found that these proteins function to initiate and amplify keratinocyte responses to inflammatory cytokines and that their constitutive activation is sufficient to drive psoriasis. Furthermore, BCL10/MALT1 signaling is frequently altered in human psoriasis patients, suggesting that targeting the BCL10/MALT1 signalosome using MALT1 inhibitors may be a promising therapeutic approach for treating psoriasis. —HMI

Sci. Immunol. **6**, eabi4425 (2021).

REVIEW SUMMARY

DEVELOPMENTAL BIOLOGY

The primitive streak and cellular principles of building an amniote body through gastrulation

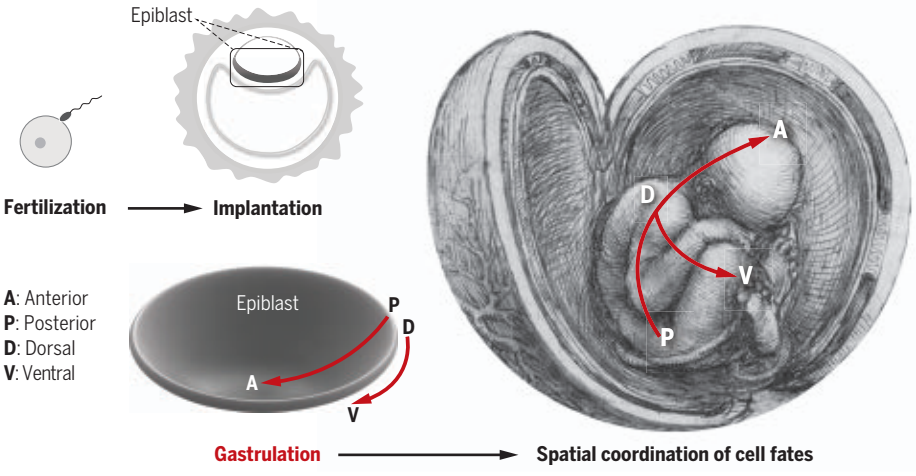
Guojun Sheng*, Alfonso Martinez Arias*, Ann Sutherland*

BACKGROUND: Pluripotent cells are generated by embryonic divisions that occur shortly after fertilization. These cells are transformed into the recognizable outline of an organism through the process of gastrulation, which endows them with lineage and spatial identities in the context of an emerging coordinate system. In many amniote embryos (such as those of reptiles, birds, and mammals), gastrulation has been associated with a transient structure called the primitive streak. Human development also follows this pattern. In humans, the primitive streak forms ~14 days after fertilization. The appearance of the primitive streak breaks the radial symmetry of the epiblast (a sheet of epithelialized pluripotent cells) and has been suggested to symbolize the emergence of human individuality. As such, many countries have established a legal limit of 14 days for the in vitro culture of fertilized human eggs—this is known as the "14-day rule." In recent years, pluripotent stem cells have become a promising in vitro model for studying the cellular and molecular mechanisms associated with early human development. Interpretation of developmental features observed in these in vitro models requires proper understanding of animal gastrulation in general and of the amniote primitive streak in particular.

ADVANCES: In this Review, we offer a phylogenetic and ontogenetic overview of the primitive streak and its role in mediating amniote gastrulation, and we discuss the implications of embryonic stem cell–based models of early mammalian embryogenesis on the function of this iconic structure. We provide evidence that the primitive streak is not a conserved feature in amniote development and that the mammalian and avian primitive streaks have evolved independently through different supracellular mechanisms that led to their morphological emergence. We argue that, in addition to mediating the emergence of germ layers from the pluripotent epiblast, gastrulation is principally a process in which an embryo acquires a coordinate system to organize its primary cell fates and the primordia of organs and tissues relative to each other in space. We highlight that in amniotes this process is regulated by a set of conserved signaling and transcriptional networks through a small collection of cellular behaviors, the tissue-level effects of which are governed by boundary conditions. We suggest that changing boundary conditions, in the form of evolution of extraembryonic lineages such as the trophectoderm and primitive endoderm, have played a key role in the transformation of the blastopore, characteristic of anamniote embryos, into the primitive streak. Variability in the organization of these tissues and the demarcation of embryonic and extraembryonic territories underpins the observed variation in the morphological appearance of the primitive streak in mammals and birds and of primitive streak–related structures in reptiles. Over the past few years, embryonic stem cells have been used as models to recapitulate several aspects of early mammalian embryogenesis. These studies have revealed that the germ layers, and even a rudimentary body plan, can form in the absence of a primitive streak.

OUTLOOK: Our model predicts that the most fundamental feature of a primitive streak–like structure in early amniote development is not its morphological manifestation but rather its capacity to mediate coordinated cell fate specification events in space. Our model also suggests that cell fate specification and tissue-level morphogenesis are regulated independently during gastrulation and then coordinated during embryonic development in vivo. In developmental models in vitro, these two processes can be uncoupled and have been shown to be influenced by different types of biomechanical parameters that mediate their coordination. This modularity leads us to suggest that the in vitro models are useful for studying gastrulation because their use without necessarily having to recapitulate embryonic structures. Future analyses of early amniote development, both in vivo and in vitro, would benefit from putting less emphasis on the primitive streak as a distinct embryological structure and more on its roles as a conduit for symmetry breaking and coordinated germ-layer differentiation.

Research into human development has direct societal and ethical impacts. Current ethical oversight in human embryo research, the 14-day rule, is effective in many countries and reflects an interdisciplinary consensus drawn somewhat arbitrarily to determine the legal rights of a human embryo. Our observations suggest that use of the primitive streak as a key developmental landmark for limiting ex vivo culture of human embryos should be reassessed. An alternative landmark, necessary for exerting ethical oversight in human-related developmental and stem cell biology research, should be selected through a consensual discussion between different stakeholders to ensure scientific and ethical rigor. ■



Schematic view of how a human embryo acquires spatial coordination of its primary cell fates through gastrulation. After fertilization, a human embryo implants into the uterus (between days 6 and 12) and gastrulation begins soon afterward, with the primitive streak emerging at day 14. Gastrulation allocates cell fates and spatial coordinates to epiblast cells, laying down the foundation of a human body.

CREDIT: ("FETUS IN THE WOMB" SKETCH) LEONARDO DA VINCI/ROYAL COLLECTION TRUST/© HER MAJESTY QUEEN ELIZABETH II 2021

The list of author affiliations is available in the full article online.
*Corresponding author. Email: sheng@kumamoto-u.ac.jp (G.S.); alfonso.martineza@upf.edu (A.M.A.); as9n@virginia.edu (A.S.)
Cite this article as G. Sheng *et al.*, *Science* **374**, eabg1727 (2021). DOI: 10.1126/science.abg1727

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abg1727

REVIEW

DEVELOPMENTAL BIOLOGY

The primitive streak and cellular principles of building an amniote body through gastrulation

Guojun Sheng[1]*, Alfonso Martinez Arias[2]*, Ann Sutherland[3]*

The primitive streak, a transient embryonic structure, marks bilateral symmetry in mammalian and avian embryos and helps confer anterior-posterior and dorsal-ventral spatial information to early differentiating cells during gastrulation. Its recapitulation in vitro may facilitate derivation of tissues and organs with in vivo–like complexity. Proper understanding of the primitive streak and what it entails in human development is key to achieving such research objectives. Here we provide an overview of the primitive streak and conclude that this structure is neither conserved nor necessary for gastrulation or early lineage diversification. We offer a model in which the primitive streak is viewed as part of a morphologically diverse yet molecularly conserved process of spatial coordinate acquisition. We predict that recapitulation of the primitive streak is dispensable for development in vitro.

Gastrulation is a stage of animal development that transforms the mass of cells from early postfertilization cell divisions into the recognizable outline of an organism. The term "gastrulation" was coined by the German embryologist Ernst Haeckel when discussing the development of sponges (*1*, *2*). Later, combined with the germ-layer theory, it acquired its common meaning as a phylogenetically conserved stage during which the primary germ layers are defined and organized. Scrutinized at the cellular level, gastrulation was revealed to be associated with species-specific cell movements that, for vertebrates, lead to a conserved pharyngula stage when embryos from different phylogenetic groups resemble each other the most and when organogenesis begins. Over the years, the term gastrulation has come to refer to a set of morphogenetic movements that serve to organize the three germ layers in animal development.

Gastrulation in amniotic vertebrates (reptiles, birds, and mammals; see Fig. 1A for phylogenetic relationships among major animal groups) is thought to involve the primitive streak, a transient structure that forms along the midline of the epiblast (the pregastrulation ectoderm), first observed in a chick embryo by Russian embryologist Christian Pander in 1817 (*3*). The emergence of the primitive streak breaks an early morphologically radial symmetry, outlines the anterior-posterior axis of the embryo, and serves as a channel for the continuous passage of mesenchymal cells toward the interior of the embryo, where they are further assigned an endodermal or mesodermal fate. Bioethicists and government regulatory bodies have associated the primitive streak with human individuality and, as such, have established a legal limit of 14 days—approximately when the primitive streak appears in a human embryo—for the in vitro culture of fertilized human eggs (known as the "14-day rule") (*4*, *5*). Until recently (May 2021), the International Society for Stem Cell Research (ISSCR) (www.isscr.org) has stated its commitment to this 14-day rule, with updated guidelines recommending case-by-case review by national and institutional review boards (*6*). Placing such an emphasis on the primitive streak for the emergence of a human being raises questions about its role in the establishment of the human body plan and whether it represents a stereotyped and generalizable structure. To address these questions, the definition (or lack thereof) of gastrulation must be revisited.

William Ballard, an American embryologist who coined the word "pharyngula" as a phylotypic stage for all vertebrate animals, stated in a discussion on gastrulation that "real problems that require new observation and experimental proof are being glossed over" and that "there has been no progress at all since the 1930's in defining what gastrulation is or when it begins or ends" (*7*). Ballard also quoted Jean Pasteels (*8*) "in that the vertebrate gastrula does not have a definitive form. It is just an abstract collective term for individuals undergoing the movements of gastrulation." We believe that these statements reflect a situation that has persisted to this day. Genetic studies have identified a conserved network of signaling molecules and transcription factors associated with gastrulation and have revealed many common features of cell biology in different species that bridge the gap between unicellular and supracellular behaviors (*9*, *10*). Furthermore, pluripotent stem cells (PSCs) represent a model system for the study of early mammalian development (*11*, *12*), and the interpretation of such experiments raises questions about the nature of gastrulation and, more importantly, whether the behavior of PSCs in vitro can identify fundamental mechanisms underlying the events happening in vivo. Progress in human PSC research has prompted discussions of whether PSC-derived structures should be considered within the same ethical realm as embryos, and the conventional wisdom is converging toward drawing a similar line on the basis of whether these structures assemble a primitive streak and undergo gastrulation. All of these issues suggest that it is time to reassess gastrulation in light of data from more recent experimental techniques and models, from which general rules may emerge, with potential ethical implications.

Here we focus on the primitive streak in amniotic vertebrates and argue that gastrulation is not only a stage but, principally, a process whereby an embryo acquires a coordinate system to organize primary cell fates in space. We highlight that this process is regulated by a set of conserved signaling and transcriptional networks through a small set of cell behaviors, the tissue-level effects of which are governed by boundary conditions. This discussion leads us to suggest that global control by boundary conditions underpins the variability of the gastrulation process in different amniotic clades and to propose a relationship between gastrulation modes observed in amniotic and anamniotic (nonamniotic) vertebrates. Later, the process of organogenesis can also lead to similar conclusions, as supported by data from PSC models.

An organismal view of gastrulation

Gastrulation (literally "formation of a small gut") and its embryological connotation were introduced by Ernst Haeckel in the context of his gastraea theory (*1*), as part of his biogenetic law that lent embryological support to the Darwinian notion of common descent of all living organisms. A gastraea is a hypothetical ancestral organism characteristic of the gastrula stage that all metazoans presumably go through ontogenetically, and gastrulation was the process encompassing this developmental stage as a universal "rite of passage" in Haeckel's recapitulation theory. A key piece of evidence for Haeckel's theory was the presence of an endoderm-like, lineage-restricted cell layer during embryogenesis of sponges (Porifera; a phylum of diploblastic metazoans; Fig. 1A)—hence the names "gastraea" and "gastrulation" [from "gaster" (which means "stomach"), an endoderm derivative]. Recently, this description has been disputed in light of data from morphological and lineage-tracing analyses (*13*, *14*), suggesting that sponges do not undergo

[1]International Research Center for Medical Sciences, Kumamoto University, Kumamoto, Japan. [2]Systems Bioengineering, DCEXS, Universidad Pompeu Fabra, Doctor Aiguader, 88 ICREA, Pag Lluis Companys 23, Barcelona, Spain. [3]Department of Cell Biology, University of Virginia Health System, Charlottesville, VA, USA.
*Corresponding author. Email: sheng@kumamoto-u.ac.jp (G.S.); alfonso.martineza@upf.edu (A.M.A.); as9n@virginia.edu (A.S.)

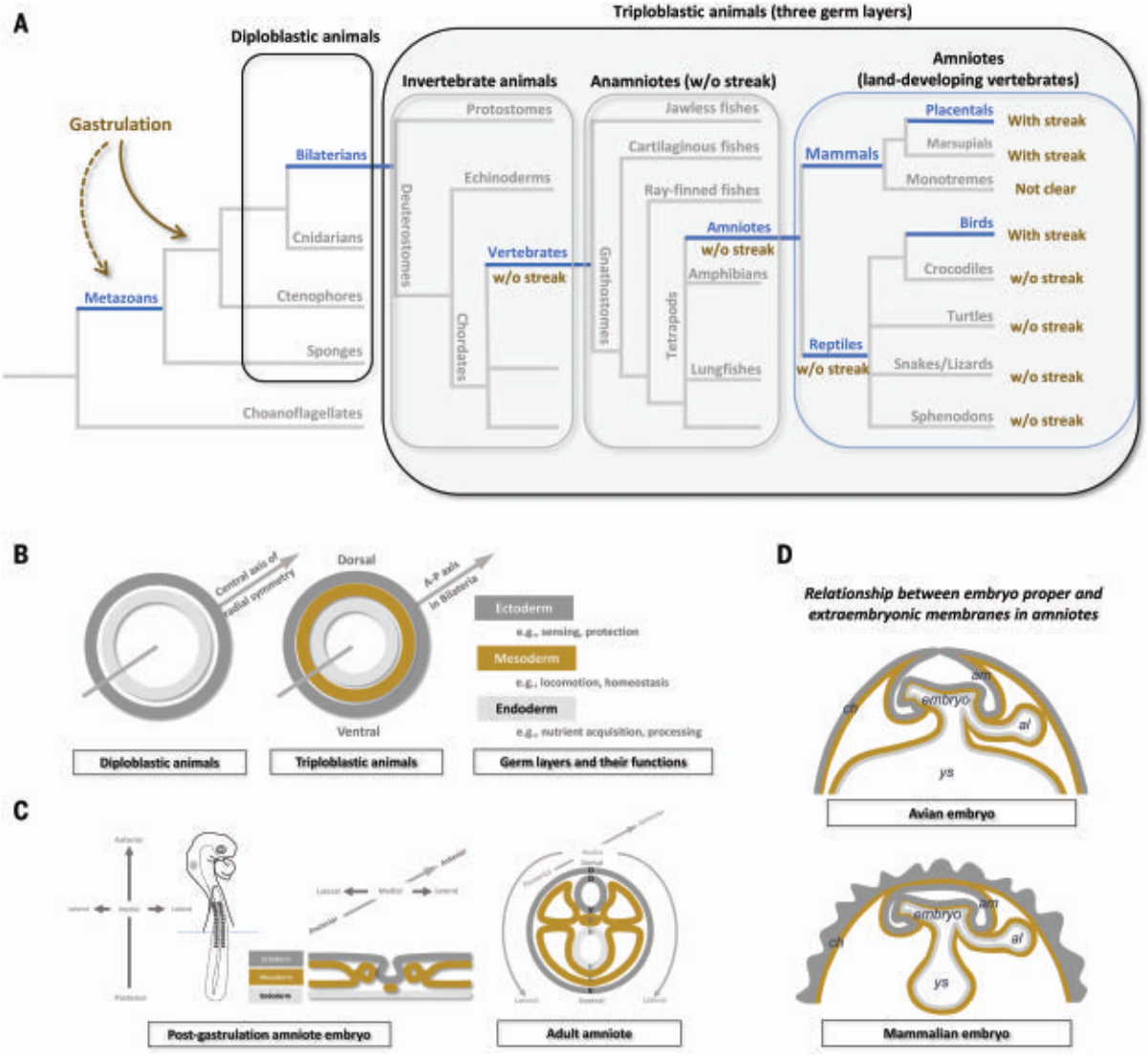


Fig. 1. Animal phylogeny and amniote body plan. (**A**) Phylogenetic relationships among major animal groups. Gastrulation is traditionally viewed as a conserved process of achieving cell lineage diversification in metazoans (animals). The relationship between gastrulation in sponges and that in the rest of the metazoans awaits further clarification (dashed arrow). Diploblastic indicates two germ layers (ectoderm and endoderm); triploblastic indicates three germ layers (ectoderm, mesoderm, and endoderm). Bilaterians are animals with bilateral body symmetry. The presence of a primitive streak–like structure during gastrulation is not conserved among amniotes (land-developing vertebrates, including extant mammals, birds, and reptiles). The term anamniotes collectively denotes vertebrate groups other than the amniotes. w/o, without. (**B**) Germ layers and their biological functions. (Left) Diploblastic animals have two germ layers and one major axis (the central axis of radial symmetry). (Middle) Triploblastic animals have three germ layers and two major axes [the anterior-posterior (A-P) and dorsal-ventral axes]. (Right) Simplified view of cellular functions of each germ layer. This overall functional assignment is conserved in all triploblastic animals. In diploblastic animals, some cellular functions that resemble those of the mesoderm are performed by either ectoderm or endoderm cells. (**C**) Basic 3D organization of a postgastrulation amniote embryo (left), with relationship of three germ layers and spatial coordinates in the anterior-posterior and medial-lateral axes shown in transverse-section view (middle). The medial-lateral axis in early development is transformed into the dorsal (D)–ventral (V) axis in the adult (right). (**D**) Schematic view of the relationship between the embryo and extraembryonic tissues in amniotes. am, amnion; ch, chorion; ys, yolk sac; al, allantois. Dark gray, ectoderm; light gray, endoderm; gold, mesoderm.

gastrulation as commonly seen in triploblastic animals. Nevertheless, comparative genomics and molecular phylogeny studies have shown that most of the genes involved in epithelial organization are present in sponge genomes (*15*, *16*), as are some of the genes involved in germ-layer formation in Bilateria (the protostomes and deuterostomes, including all vertebrate species) (*14*) and in the epithelial-to-mesenchymal transition (EMT) (*17*). In Cnidaria, a phylum of diploblastic animals most closely related to triploblastic Bilateria (Fig. 1, A and B), molecular tool kits for epithelial organization and EMT are present in genomes, and gastrulation generates mesendoderm-like cells that express mesoderm transcriptional regulators and mesoderm-specific lineage markers, despite their lack of a genuine mesoderm germ layer (*18*).

These seemingly contradictory observations highlighted the need for a conceptual framework for understanding the relationship between the primitive streak and amniote gastrulation from the perspective of metazoan phylogeny. The gastraea theory posited that

functional diversification is fundamental to multicellularity and that specification of cells dedicated to nutritional acquisition (i.e., the endoderm) is a conserved process among all metazoans (Fig. 1B). With this came notions of conservation of gastrulation at the genetic and morphogenetic levels, revealed in modern biological terms as molecular regulatory networks and dynamic changes in supracellular organization, respectively. Gastrulation, therefore, can be practically defined by traits that describe cell fates (e.g., neurons, notochord cells, and enterocytes) and shapes (e.g., epithelium, mesenchyme, migration, and EMT), as well as by their associated molecular characteristics (e.g., *Brachyury* for mesoderm and endoderm progenitors and *Nanog* for undifferentiated pluripotent cells). Depending on the choice of experimental models and investigative tools, the particular aspect of gastrulation to emphasize has varied since Haeckel's time. The core, unchanged, and unifying concept of gastrulation, in our view, is for it to serve as a conduit for diversification of cellular lineages and acquisition of a spatial coordinate system for subsequent functional integration in later development.

Central to our current understanding of gastrulation is the primitive streak, which is associated with a spatially oriented, dynamic sequence of individual cells leaving the surface layer in the process of ingression, through the EMT. The primitive streak has been viewed as iconic for gastrulation, although it is present only in specific amniotic vertebrates, such as birds and mammals (*19*) (Figs. 1 and 2). In reptiles (nonavian reptiles, in the context of this article), gastrulation is associated with involution (i.e., rolling of a sheet of cells over a horseshoe-shaped blastopore, an orifice characteristic of amphibian gastrulation) accompanied by ingression in the blastoporal (primitive) plate located posteriorly to the blastopore (*19*–*22*) (Fig. 3). Thus, all functional cell types shared by amniotic vertebrates can be generated regardless of the presence or absence of a primitive streak, leading to the question of whether gastrulation-related morphological features are conserved among all amniotes. Furthermore, relevant to the effort to recapitulate all, or part of, mammalian development in vitro is the question of whether a conserved morphogenetic process similar to the appearance of a primitive streak or a blastopore is required for the functional diversification of cell lineages.

A closer look at amniote embryos undergoing gastrulation, compared with embryos of anamniotes (nonamniote vertebrates), identifies three shared morphological features

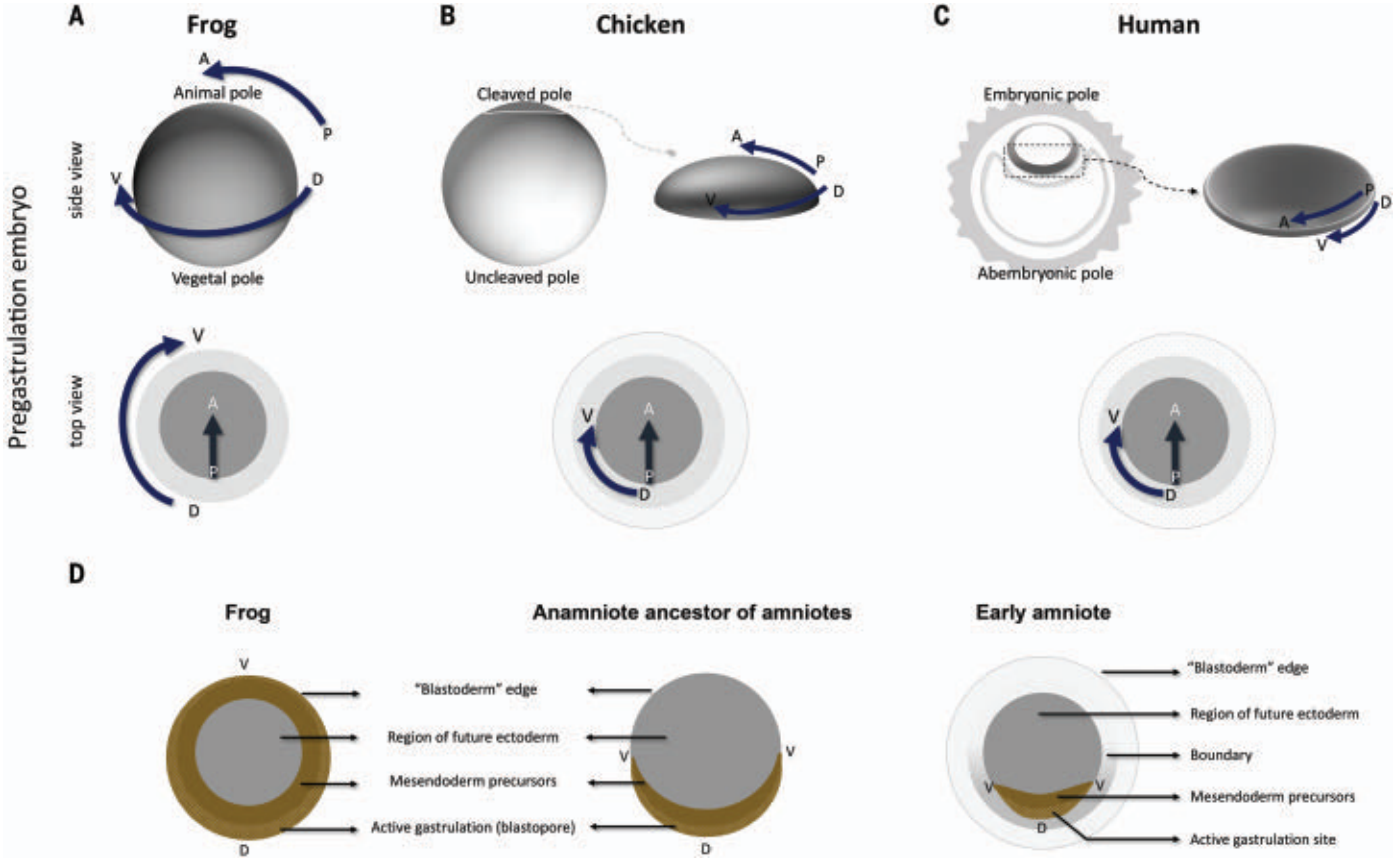


Fig. 2. Phylogenetic comparison of spatial organization in amphibian (frog), avian (chicken), and mammalian (human) embryos before gastrulation and hypothetical transitions in gastrulation morphogenesis during early amniote evolution. Future anterior-posterior and dorsal-ventral axes are indicated in pregastrulation embryos. (**A**) In the frog (*Xenopus*) embryo, the animal pole is where the oocyte nucleus is located before fertilization (it is also lighter because it has less yolk than the vegetal pole). (**B**) In the chicken embryo, the cleaved pole is where the oocyte nucleus is located and where cellularization of the fertilized egg takes place. The uncleaved pole is a part of the oocyte that does not contribute to the cellularized embryo and contains mostly nutritious materials. (**C**) In the human embryo, the embryonic pole is the side where the inner cell mass is located in a blastocyst. The abembryonic pole is the opposite side. The trophectoderm and primitive endoderm are drawn but not labeled. Epiblast cells form the proamniotic cavity, with upper cells giving rise to the amniotic ectoderm and lower cells (area represented by the dashed rectangle) giving rise to the major part of all three germ layers. In the lower panels, dark gray indicates cells that will contribute mainly to the ectoderm lineages; light gray denotes the marginal zone, which will control the location of internalization, contribute to mesendoderm cells, and mediate boundary biomechanical cross-talk—which, in amniotes, will influence the morphogenesis of primitive streak–like structures. (**D**) Hypothetical evolutionary transitions in gastrulation morphogenesis, leading to an amniote-specific mode of mesendoderm internalization.

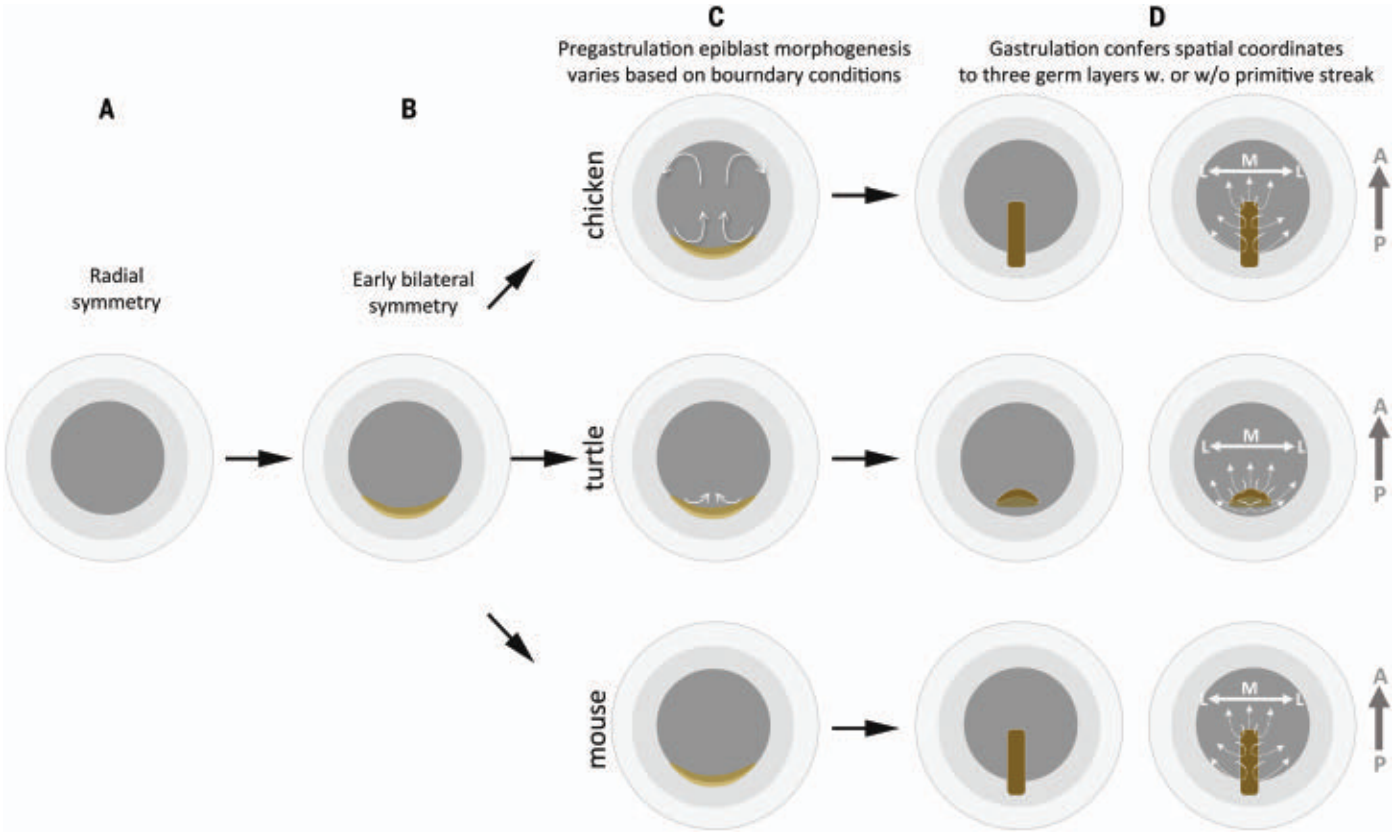


Fig. 3. The primitive streak as a morphogenetic consequence of variable boundary conditions in the pregastrulation epiblast of an amniote embryo. The primitive streak is neither conserved nor necessary for amniote gastrulation. Ontogenetically, in the pregastrulation embryo, radial symmetry (**A**) (similar to the top view in Fig. 2) is transformed into bilateral symmetry when mesoderm and endoderm precursors are induced asymmetrically in the marginal zone (**B**). Geometric organization of active internalization of mesendoderm cells is influenced by planar morphogenesis (white arrows) and cellular proliferation of the epiblast before onset of gastrulation and by boundary conditions reflected as embryo-specific biomechanical constraints on epiblast cells (**C**). In reptilian (e.g., turtle) embryos, such morphogenesis results in the formation of a blastopore as the active center of internalization, whereas in chick and mouse embryos, a primitive streak–like structure forms, with different morphogenetic origins. (**D**) Internalization of mesendoderm cells (small white arrows; the dotted parts indicate cell movement after internalization, contributing to the 3D organization of a postgastrulation embryo). M, medial; L, lateral.

among all amniote clades: (i) Internalization of mesoderm- and endoderm-fated cells takes place in a circumferentially restricted manner (compare an anamniote embryo in Fig. 2A with amniote embryos in Fig. 2, B and C). (ii) Pregastrulation ectoderm (epiblast) cells are organized as a single-cell–layered epithelium. (iii) The epiblast is divided into embryonic and extraembryonic territories, and gastrulation is initiated at their boundary (Fig. 2, B to D). To integrate these three features of amniote gastrulation in a simple model, the primitive streak in birds and therian mammals can be viewed as an independent morphogenetic adaptation of a basal, reptile-like bimodal form of involution and ingression (Figs. 1 to 3) (*19*, *20*). The presence of a neurenteric canal (i.e., an opening of the epiblast that connects the amniotic cavity and the yolk sac) in human and other primate embryos (*23*) and the rudimentary chordal canal associated with head process formation in the mouse embryo (*24*) can be considered to represent a residue of blastoporal involution that continues to play a role in axial mesendoderm internalization. These features and the proposed unifying model, however, highlight only morphogenetic constraints placed on amniote gastrulation, which by themselves may or may not be essential for fulfilling functional roles of gastrulation as a conduit for cell lineage diversification, spatial coordinate establishment, and inter– or intra–germ-layer coordination during tissue and organ formation.

The first two features conserved among amniote vertebrates are also observed in certain groups of anamniotes. For example, unlike *Xenopus*, the main amphibian model, all salamanders (Urodela) (*25*–*29*) and caecilians (Gymnophiona) (*30*, *31*) studied so far exhibit a gastrulation process that is restricted to or extremely biased toward the dorsal marginal zone. Similarly, such dorsally restricted gastrulation is also present in lungfish (the closest relatives of Tetrapoda) (*32*, *33*), dogfish (cartilaginous fish) (*34*), and lampreys (jawless vertebrates) (*35*), which suggests that the amniotes likely inherited this feature of gastrulation from an anamniote ancestor (see Fig. 1A for phylogenetic relationship). The second feature, a unilaminar, epithelialized pregastrulation ectoderm, is also present in certain anamniote groups (e.g., urodeles) (*36*), which suggests a preamniote origin. The third feature, that gastrulation is initiated at the boundary between intraembryonic and extraembryonic territories, is associated with the evolutionary invention of the amnion and chorion in ancestral amniotes (Fig. 1D), where internal fertilization, intrauterine early embryogenesis, and land-based fetal development necessitate the presence of these protective layers. The amnion, composed of ectoderm- and mesoderm-derived cells, provides a protective liquid-filled environment for the developing embryo and is the defining feature of amniotic vertebrates. The chorion, similarly composed of ectoderm and mesoderm cells, forms the external boundary of the embryo (including all intraembryonic and

extraembryonic cell lineages) and is the primary interface for fetal-environmental (including fetal-maternal) exchanges.

Except for birds, in which gastrulation and primitive streak formation takes place after oviposition (egg laying), all extant amniote embryos initiate gastrulation during their intrauterine development. Variations in amniote early development are manifested as differences in oocyte size (e.g., 30 mm in the chick versus 0.1 mm in the human), cleavage pattern (e.g., meroblastic cleavage with delayed cytokinesis and incomplete cytoplasm partitioning in the chick versus holoblastic cleavage with complete partitioning of zygotic cytoplasm in the human), cell cycle duration (e.g., a rapid-to-slow shift in birds versus a slow-to-rapid shift in mammals), and cell number at the onset of gastrulation (e.g., 100-fold difference between the chick epiblast and the mouse epiblast). These variables lead to diverse epiblast "landscapes" during the transition from pluripotency to a rudimentary embryonic architecture with three germ layers. Furthermore, these preconditions are met with the need to demarcate the pregastrulation epiblast into the intraembryonic and extraembryonic territories to facilitate amniogenesis, and as a consequence, internalization of mesoderm and endoderm cells is initiated at the intraembryonic-extraembryonic boundary. In embryos of eutherian animals (placental mammals, including mice and humans), the gastrulation process comes under the additional influences of implantation and placentation, both of which exhibit pronounced diversity and evolutionary adaptability. For example, a horse embryo initiates trophoblast-endometrium contact 1 month after fertilization, when the embryo has already reached organogenesis stages and placentation thereafter remains epitheliochorial (i.e., with superficial, noninvasive feto-maternal contacts). By contrast, a human embryo starts to breach the maternal endometrium soon after blastocyst hatching and completes the invasive implantation process by day 12 after fertilization, well before the initiation of gastrulation or primitive streak formation (*37*, *38*). Functional differentiation of the trophectoderm lineage becomes a crucial early event for eutherian mammals, and the onset and morphogenetic process of gastrulation are affected to a greater or lesser extent, reflecting physical and structural variations in eutherian feto-maternal interactions.

These variable and adaptable features of amniote early development at the organismal level challenge the notion that gastrulation is associated with a specific structure and underscore the importance of looking for components of the process that are more conserved across amniotes, potentially at the molecular and cellular levels. Genetic studies over the past 20 years have indeed revealed a conserved set of transcription factors and signaling molecules associated with gastrulation in both the amniotes and anamniotes. For example, interactions between bone morphogenetic protein (BMP), Nodal, and Wnt signaling lead to *Brachyury* expression at the onset of gastrulation and serve as a gateway for the specification of different organ and tissue primordia. However, this conservation at the molecular level contrasts with the morphogenetic-level variability (*39*) and leads us to search for common motifs at the level of cell behaviors as a way to understand the origin of the variety of gastrulation modes.

A cellular view of gastrulation

Vertebrate gastrulation is generally associated with the transformation of a cellular aggregate into a bilaterally elongated structure with spatial information conferred in all three germ layers (Figs. 1 to 3). This is accomplished in a species-specific manner. In amniotes, the initial step in this process is a mesenchymal-to-epithelial transition (MET) of the early cells to form the pluripotent epiblast epithelium (*40*), followed by ingression and concomitant specification of the endoderm and mesoderm, with an orderly process of EMT placing the three germ layers with respect to the anterior-posterior body axis (Fig. 1C). At the cellular level, the behaviors associated with EMT are quite well conserved between the mammalian and avian primitive streaks (*41*–*45*). The underlying basement membrane is degraded, and epiblast cells apically constrict to become flask-shaped and then detach from neighboring cells to exit the epithelium (*46*–*48*). At the molecular level, this involves down-regulation of E-cadherin expression (*49*) as well as decreased Rho activity and changes in organization of cytoskeletal components [particularly microtubules, (*41*, *50*) in the ingressing cell] without disturbing the overall epithelial nature of the tissue (*41*, *43*). These changes in cell behavior are associated with similar patterns of signaling and gene expression as well—including increases in Wnt, Nodal, and fibroblast growth factor (FGF) and FGF receptor 1 (FGFR1) signaling—that lead to up-regulated expression of *Brachyury*, *Snai1* and *Snai2*, and *FoxA2* transcription factors.

Similarly, at the cellular level, behaviors underlying formation and elongation of the primitive streak are fairly well conserved, including cell shape changes and polarized rearrangement of cells along the mediolateral axis. However, how these cellular level behaviors are deployed and choreographed to generate tissue shape changes are distinctly species specific, likely due to biomechanical constraints imposed by the particular embryonic environment. In birds, the mesendoderm precursors that will form the primitive streak are induced in a sickle-shaped region at the posterior margin of the area pellucida (*51*) and undergo extensive, polarized rearrangements that result in convergence on the midline and anterior extension (*52*–*54*). Epiblast cells in the area anterior to the forming primitive streak rearrange to extend perpendicular to the primitive streak, leading to anterior-posterior contraction and lateral elongation of this region and large-scale counterrotational flows of cells in regions of the epiblast lateral to the forming streak (*52*).

Studies of gastrulation in mammalian embryos reveal distinct tissue-level differences in primitive streak initiation and elongation from the model established in the avian embryo. The mouse embryo, with an elongated cup-shaped epiblast, has no morphological structure equivalent to the sickle-shaped mesendodermal precursor region of the chick, and the mouse primitive streak forms through progressive initiation of EMT rather than convergent extension of a precursor population (*43*). In discoid embryos (such as those of rabbits or pigs), in which the size and geometry of the epiblast are more similar to those in chick embryos, the posterior margin expands posteriorly and becomes less dense cellularly before primitive streak formation, in contrast to the convergent extension of mesendoderm precursors in the avian embryo (*55*, *56*). In addition, the pronounced counterrotational flows characteristic of the avian embryo are not seen in the rabbit embryo (*55*, *57*). Our knowledge of the primitive streak in human embryos, primarily on the basis of anatomical descriptions and comparisons with other primate embryos, reflects similarity to other mammalian embryos, although additional features (such as the neurenteric canal) have been discovered (*58*, *59*). The differences between avian and mammalian primitive streak formation reflect distinct patterns of biomechanical force generation, perhaps as a consequence of distinct biomechanical architecture at the border of intraembryonic and extraembryonic territories that influences tissue-level behaviors in these embryos. In addition to centrifugal tension exerted by expansion of the area opaca (the region of the epiblast that is physically attached to the yolk) in avian embryos (*60*, *61*), polarization and mediolateral intercalation of posterior marginal cells (*52*) initiate graded tangential forces along the margin of the epiblast that lead to cell shape changes in the marginal cells and formation of supracellular myosin cables in groups of 5 to 20 cells (*53*). The forces associated with this tensile ring are anisotropic, providing active tension posteriorly and passive tension anteriorly, and modelling shows that they are sufficient to drive the posterior epiblast cells forward (*53*). Anteriorly, the marginal tension may provide a boundary that directs the lateral movement of cells anterior to the streak to generate the counterrotational

flows of cells toward the posterior of the embryo (Fig. 4B). Furthermore, the marginal tensile ring may well act to regulate gene expression and to define the region of primitive streak formation (*62*).

It is tempting to speculate that such a tensile ring could similarly provide the boundary conditions to generate tissue-wide cell movement in mammalian embryos with discoid epiblasts, including human embryos. However, the posterior expansion and decrease in cellular density observed in some of these embryos (*56*, *63*), as well as the lack of strong posterior convergence movements, seem inconsistent with initiation of a tensile ring at the epiblast margin (Fig. 4D). Additionally, the effects on primitive streak formation of inhibiting myosin activation through Rho kinase inhibition differ substantially between rabbit and avian embryos (*52*, *57*), suggesting that the role of actomyosin contractility during primitive streak formation and gastrulation is not the same in these two embryo types (Fig. 4). Forces tangential to the margin of the epiblast are very unlikely to play a role in mouse embryo gastrulation, given the cuplike geometry of the epiblast. In fact, it may be that the absence of such tangential forces, as well as a less fluidlike epiblast, act as constraints for the behavior of epiblast cells in mouse embryos, leading to formation of the primitive streak without global cell movements. Additionally, although centrifugal forces on the epiblast may be generated through expansion of the fluid-filled proamniotic cavity (*64*), there is no tension generated by the extraembryonic tissues across the epiblast.

Thus, coordinated cell behaviors, acting under the particular constraints of the embryonic environment, produce tissue-level deformations that generate a linear region of cell ingression. Are these morphogenic constraints a necessity for amniote gastrulation or merely a problem to be solved? The variations in the mode of gastrulation among amniote and anamniote vertebrates [as discussed above and in (*65*)] suggest the latter. In fact, recent data show that manipulation of the extent and cellular behaviors of the mesendoderm in the chick embryo can transform the primitive streak into gastrulation modes more similar to those of amphibian or reptile embryos (*62*, *66*). Data from the field of mammalian stem cell and regenerative biology appear to further support this line of argument.

in vitro models of mammalian gastrulation

Over the past few years, PSCs have been used to generate a number of models of early mammalian development. PSCs are clonal derivatives from mammalian blastocysts [embryonic stem cells (ESCs)] or reprogrammed adult cells (induced PSCs). PSCs can be maintained in culture for many generations without losing their ability to produce all cells of an organism, and they can be steered to differentiate into any cell type by controlling the culture conditions. This differentiation is asynchronous, exhibits large heterogeneities in gene expression (*67*, *68*), and reveals the existence of cell-intrinsic programs of gene expression associated with specific fates. In all cases, cells go through patterns of gene expression that mirror events observed in embryos and, early on, cell fate decisions in the early epiblast. Thus, during differentiation into endodermal and mesodermal derivatives in culture, cells go through a sequence of events similar to those of gastrulation—down-regulation of pluripotency genes, MET, activation of Wnt and Nodal signaling, transient expression of *Brachyury*, and engagement into an EMT (*69*)—before expressing specific fate determinants. However, these changes occur without any multicellular coordination or morphogenesis. Despite the sequential events of MET, expression of *Brachyury*, and EMT, no primitive streak–like structure is visible in the culture.

In an attempt to restrain the heterogeneities that arise in adherent culture, human and mouse PSCs have been cultured on two-dimensional (2D) micropatterned structures (*70*). Under these conditions, cells form tight epithelia, and exposure to BMP, Nodal, and Wnt signals results in the emergence of proportionate concentric rings of gene expression identified as the different germ layers and some of their derivatives during gastrulation (*70*–*72*). In these arrangements, also known as 2D gastruloids, radial symmetry can be broken only by spatially controlled asymmetric flow of signal agonists and antagonists (*73*). These experiments have provided insights into the mechanisms by which cells interpret and respond to morphogens in perigastrulation stages. However, likely because of the confinement of cells on the substrate, cellular growth and movement are impaired and patterns of gene expression do not exhibit the structural evolution that they do in the embryo. In the case of mouse PSCs, this experimental setting has been shown to recapitulate relationships between signals and fates in the embryo that are known from genetic studies, thus validating the patterns observed for human ESCs (*72*). On the basis of movements correlated with cells expressing *Brachyury*, it has been suggested that there is a circular

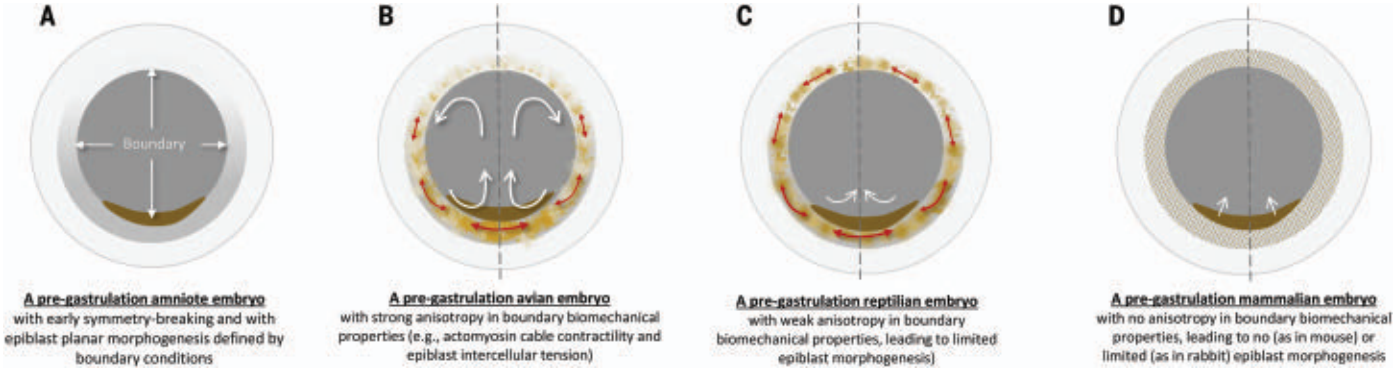


Fig. 4. Biomechanical properties at the embryonic-extraembryonic boundary can influence epiblast morphogenesis before gastrulation. (**A**) An amniote embryo after symmetry breaking but before gastrulation. Epiblast morphogenesis and, consequently, the morphology of a primitive streak–like structure are influenced by embryonic-extraembryonic boundary conditions. (**B**) The avian embryo has strong biomechanical anisotropy at the boundary and undergoes prominent planar rearrangement of epiblast cells, leading to primitive streak formation. (**C**) The reptilian embryo has weak biomechanical anisotropy at the boundary and undergoes limited, regional rearrangement, leading to blastopore formation. (**D**) The mammalian embryo has no or weak biomechanical anisotropy at the boundary and undergoes no (in mice) or limited (in rabbits) epiblast planar rearrangement. A primitive streak–like structure still forms in the mammalian embryo, likely due to directional EMT signal propagation and cell proliferation. In (B) to (D), the dashed line denotes the midline, and the white arrows indicate the direction of global movements of embryonic epiblast cells. In (B) and (C), the red arrows indicate the direction and strength of the intercellular tension force between epiblast cells located at the embryonic-extraembryonic boundary.

primitive streak in these 2D gastruloids (*74*). However, it is not clear that these movements are related to the presence of a primitive streak or whether they are likely to reflect the observation that *Brachyury* promotes cell movement (*69*, *75*). The primitive streak in vivo is a structure with a vectorial component that defines the anterior-posterior axis of the embryo, which is absent in the micropattern experiments. Nevertheless, together with the free adherent culture experiments, this work shows that a primitive streak is not necessary to implement the schedules of gene expression associated with the amniote body plan, as has been suggested by perturbation experiments of avian development in vivo, in which properly patterned mesoderm induction could be uncoupled from primitive streak formation (*76*).

The coordination of signal responses and proportionate patterning of gene expression in experimental systems in vitro highlights the role of cell collectives in cell fate decisions and provides an opportunity to explore the role of biomechanical signals during gastrulation. These arrangements provide a versatile device to address the role that mechanochemical signaling plays in the early stages of development. Thus, BMP and Nodal (*77*) as well as Wnt (*78*) signaling are influenced by the geometry and mechanics of the cellular environment (*77*). In particular, Wnt signaling and *Brachyury* expression, two key regulators of the onset and progression of primitive streak formation, appear to be mechanosensitive. *Brachyury* expression and polarized cell movements can be induced at points of high tension in 2D cultures of human PSCs on soft substrates by Wnt/β-catenin signaling (*78*). This relationship between mechanics, Wnt signaling, and *Brachyury* has been described in other systems (*79*) and suggests that β-catenin might act as a general mechanotransducer during gastrulation.

When PSCs are aggregated and grown in suspension, large-scale patterns emerge that are not predictable from adherent cultures. Thus, under controlled culture conditions PSCs form aggregates that, when exposed to Wnt signaling, undergo symmetry breaking, which is manifest in localized expression of *Brachyury* at one pole of the aggregate, global shape changes, and the expression of mesoderm- and endoderm-specific genes regionally organized with reference to an orthogonal coordinate system (*69*, *80*–*86*). These structures are called "gastruloids" (*80*), and despite the pronounced recapitulation of fate specification and multiaxial organization, they lack a morphologically recognizable primitive streak or any of the aforementioned morphological features associated with amniote gastrulation. Gastruloids emphasize the disconnect between gene expression and morphogenesis revealed in adherent cultures. They also extend the conclusion, derived from micropattern experiments, that multicellular ensembles can pattern gene expression in space and, when allowed to grow, create patterns of structured gene expression that mimic the situation in embryos. Gastruloids also highlight the importance of boundary conditions in early morphogenetic events. In this regard, a recent variation of the gastruloids demonstrated that spatial localization of a Wnt/Nodal signaling center to one end of an aggregate of naïve cells increases the spatial ordering of the fates emerging from the gastruloid (*87*). Mouse and human epiblasts are organized very differently, owing to the arrangement of the extraembryonic tissues, but when these constraints are removed by placing the PCS or epiblast derivatives in culture, both form very similar structures (*85*). This suggests that the primitive streak can be hypothesized as a mechanical response to the boundary conditions surrounding the embryo (Fig. 4). In support of this, when ESCs are placed in culture with trophoblast stem cells (TSCs), they form structures that resemble the early epiblast, in which a primitive streak–like structure can be observed to emerge at the interface between ESCs and TSCs (*88*, *89*). This is even more accentuated when the primitive endoderm is induced in these constructions (*90*). The initiation of gastrulation can also be observed in a human model of amniotic sac formation (*91*), supporting the notion that the primitive streak might be a consequence of mechanical constraints imposed on the epiblast.

Origin and function of the primitive streak: A look beyond

The PSC models of mammalian development provide some insights into the role that gastrulation and its associated structures play in laying down the amniote body plan. In particular, they call into question whether the primitive streak is the hallmark of gastrulation. First, these models reveal that fate-specific schedules of gene expression and morphogenesis, including the primitive streak, are two independent processes that are coordinated during embryonic development in vivo. Gastruloids, which lack a primitive streak**,** still exhibit polarized gene expression and are able to lay down a body plan, but their constituent cells are not spatially constrained at their boundaries, as seen in embryos (Fig. 4). By contrast, when the initial aggregates are 2D and epithelial, they require higher levels of Wnt signaling for patterning, facilitated by BMP from the trophectoderm (*88*), and exhibit EMT-like movements associated with *Brachyury* expression at the interface between the PSCs and TSCs, similar to what occurs in the embryo. This suggests that aggregates of PSCs have an intrinsic ability to break symmetry associated with polarized *Brachyury* expression, but the signaling threshold for these events is raised when the cells are epithelial, and interactions with the trophectoderm bias the onset of symmetry breaking and trigger the localized emergence of the primitive streak. We surmise that having an epithelial substrate for the initiation of the primitive streak in embryos (*40*, *92*) allows for controlled, precise patterning, which is absent in gastruloids. Although PSC models (gastruloids in particular) exhibit a well-organized outline of the primordia for tissues and organs, this outline lacks the precision and detailed patterning characteristic of embryos, which are necessary for the generation of a functional organism. Therefore, the primitive streak may represent a structural feature that allows alignment of the primordia and serves as a morphogenetic conduit for precision in spatial and temporal coordination of early development.

We would like to propose that the avian primitive streak, which has served as the reference structure for amniote gastrulation, is not a conserved structure but rather part of a continuum that spans species-specific structures seen in different amniote clades. It has arisen independently a number of times, as the common ancestor of birds and mammals likely lacked a primitive streak (Fig. 1). One may also view primitive streak–like structures—including avian primitive streaks, reptilian blastopores, and various types of mammalian primitive streaks—in a broader sense, as a manifestation of the midline that indicates the emergence of anterior-posterior polarity from the initial radial symmetry and serves as a reference point for gastrulation movement and precise coordination of three-germ-layer differentiation. In all amniotes, the primitive streak–like structure begins at a precise location that is not only the posterior but also the seed of the midline of the organism. Adopting such a view implies putting less emphasis on the primitive streak as a distinct embryological structure and more on its roles as a conduit for symmetry breaking and coordinated germ-layer differentiation. Furthermore, it supports the view of gastrulation as a process that may begin before the visualization of patterned cell movements. Recapitulation of these functions of the primitive streak in vitro, rather than of its morphological manifestation, will have profound implications in developmental and translational biology.

Our discussion supports the recent reassessment of the primitive streak as it pertains to ethical debates surrounding rules and regulations of in vitro studies of human development (*6*). Currently, as a key element of ethical oversight in human embryo research, the 14-day rule is effective in many countries, including the US, UK, Japan, China, and Spain. This rule reflects an interdisciplinary consensus in developmental and stem cell biology and assisted human reproduction and represents a somewhat arbitrary concept to determine the legal rights of a human embryo (*93*). The scientific

argument was that the primitive streak, formed after ~14 days of human development post-fertilization, represented a definitive sign of human individuality—i.e., from that point on, only one human individual would emerge, endowed with rights and sanctity that must be protected through ethical and legal means. We have argued that the primitive streak is neither conserved nor necessary for germ-layer formation or spatial coordinate acquisition, therefore emphasizing the arbitrariness of the original 14-day decision and supporting the new ISSCR guidelines. However, this line of reasoning does not dispute the necessity of exerting ethical oversight in human-related developmental and stem cell biology research, nor does it preclude defining an alternative developmental landmark as a limit for the culture of human embryos ex vivo. This should be done, as it was by the Warnock committee, through a consensual discussion among different stakeholders to ensure scientific and ethical rigor.

REFERENCES AND NOTES

1. E. Haeckel, *Die Kalkschwämme: Eine Monographie in zwei Bänden Text und einen Atlas mit 60 Tafeln Abbildungen* (Georg Reimer, 1872).
2. E. Haeckel, Die Gastrae Theorie, die phylogenetische Classification des Thierreichs und die Homologie der Keimblatter. *Jenaische Zeitschrift fur Naturwissenschaft* **8**, 1–55 (1874).
3. C. H. Pander, *Beiträge zur Entwickelungsgeschichte des Hühnchens im Eye* (Heinrich Ludwig Brönner Würzburg, 1817).
4. Committee of Inquiry into Human Fertilisation and Embryology, Department of Health and Social Security, M. Warnock, "Report of the Committee of Inquiry into Human Fertilisation and Embryology" (Her Majesty's Stationery Office, Great Britain, 1984).
5. Ethics Advisory Board, Department of Health Education and Welfare, "HEW support of research involving human in vitro fertilization and embryo transfer: Report and conclusions" (US Government Printing Office, 1979).
6. R. Lovell-Badge *et al*., ISSCR Guidelines for Stem Cell Research and Clinical Translation: The 2021 update. *Stem Cell Reports* **16**, 1398–1408 (2021). doi: 10.1016/j.stemcr.2021.05.012; pmid: 34048692
7. W. W. Ballard, Problems of Gastrulation: Real and Verbal. *Bioscience* **26**, 36–39 (1976). doi: 10.2307/1297297
8. J. Pasteels, Un aperçu comparatif de la gastrulation chez les chordés. *Biol. Rev.* **15**, 59–106 (1940). doi: 10.1111/j.1469-185X.1940.tb00942.x
9. L. Solnica-Krezel, D. S. Sepich, Gastrulation: Making and shaping germ layers. *Annu. Rev. Cell Dev. Biol.* **28**, 687–717 (2012). doi: 10.1146/annurev-cellbio-092910-154043; pmid: 22804578
10. C. D. Stern, Ed., *Gastrulation: From Cells to Embryo* (Cold Spring Harbor Laboratory Press, 2004).
11. I. Heemskerk, A. Warmflash, Pluripotent stem cells as a model for embryonic patterning: From signaling dynamics to spatial organization in a dish. *Dev. Dyn.* **245**, 976–990 (2016). doi: 10.1002/dvdy.24432; pmid: 27404482
12. J. Fu, A. Warmflash, M. P. Lutolf, Stem-cell-based embryo models for fundamental research and translation. *Nat. Mater.* **20**, 132–144 (2021). doi: 10.1038/s41563-020-00829-9; pmid: 33199861
13. S. P. Leys, D. Eerkes-Medrano, Gastrulation in Calcareous Sponges: In Search of Haeckel's Gastraea. *Integr. Comp. Biol.* **45**, 342–351 (2005). doi: 10.1093/icb/45.2.342; pmid: 21676779
14. N. Nakanishi, S. Sogabe, B. M. Degnan, Evolutionary origin of gastrulation: Insights from sponge development. *BMC Biol.* **12**, 26 (2014). doi: 10.1186/1741-7007-12-26; pmid: 24678663
15. H. Belahbib *et al*., New genomic data and analyses challenge the traditional vision of animal epithelium evolution. *BMC Genomics* **19**, 393 (2018). doi: 10.1186/s12864-018-4715-9; pmid: 29793430
16. B. Fahey, B. M. Degnan, Origin of animal epithelia: Insights from the sponge genome. *Evol. Dev.* **12**, 601–617 (2010). doi: 10.1111/j.1525-142X.2010.00445.x; pmid: 21040426
17. A. V. Ereskovsky, E. Renard, C. Borchiellini, Cellular and molecular processes leading to embryo formation in sponges: Evidences for high conservation of processes throughout animal evolution. *Dev. Genes Evol.* **223**, 5–22 (2013). doi: 10.1007/s00427-012-0399-3; pmid: 22543423
18. M. Salinas-Saavedra, A. Q. Rock, M. Q. Martindale, Germ layer-specific regulation of cell polarity and adhesion gives insight into the evolution of mesoderm. *eLife* **7**, e36740 (2018). doi: 10.7554/eLife.36740; pmid: 30063005
19. F. Bertocchini, C. Alev, Y. Nakaya, G. Sheng, A little winning streak: The reptilian-eye view of gastrulation in birds. *Dev. Growth Differ.* **55**, 52–59 (2013). doi: 10.1111/dgd.12014; pmid: 23157408
20. M. J. Stower, F. Bertocchini, The evolution of amniote gastrulation: The blastopore-primitive streak transition. *WIREs Dev. Biol.* **6**, e262 (2017). doi: 10.1002/wdev.262; pmid: 28177589
21. M. J. Stower *et al*., Bi-modal strategy of gastrulation in reptiles. *Dev. Dyn.* **244**, 1144–1157 (2015). doi: 10.1002/dvdy.24300; pmid: 26088476
22. M. Coolen *et al*., Molecular characterization of the gastrula in the turtle Emys orbicularis: An evolutionary perspective on gastrulation. *PLOS ONE* **3**, e2676 (2008). doi: 10.1371/journal.pone.0002676; pmid: 18628985
23. A. Rulle *et al*., On the Enigma of the Human Neurenteric Canal. *Cells Tissues Organs* **205**, 256–278 (2018). doi: 10.1159/000493276; pmid: 30481762
24. A. Jurand, Some aspects of the development of the notochord in mouse embryos. *J. Embryol. Exp. Morphol.* **32**, 1–33 (1974). pmid: 4141719
25. D. R. Shook, R. Keller, Epithelial type, ingression, blastopore architecture and the evolution of chordate mesoderm morphogenesis. *J. Exp. Zool. B* **310**, 85–110 (2008). doi: 10.1002/jez.b.21198; pmid: 18041055
26. G. Swiers, Y. H. Chen, A. D. Johnson, M. Loose, A conserved mechanism for vertebrate mesoderm specification in urodele amphibians and mammals. *Dev. Biol.* **343**, 138–152 (2010). doi: 10.1016/j.ydbio.2010.04.002; pmid: 20394741
27. C. A. Hurney *et al*., Normal table of embryonic development in the four-toed salamander, Hemidactylium scutatum. *Mech. Dev.* **136**, 99–110 (2015). doi: 10.1016/j.mod.2014.12.007; pmid: 25617760
28. T. Kaneda, J. Y. Motoki, Gastrulation and pre-gastrulation morphogenesis, inductions, and gene expression: Similarities and dissimilarities between urodelean and anuran embryos. *Dev. Biol.* **369**, 1–18 (2012). doi: 10.1016/j.ydbio.2012.05.019; pmid: 22634398
29. R. P. Elinson, E. M. del Pino, Developmental diversity of amphibians. *WIREs Dev. Biol.* **1**, 345–369 (2012). doi: 10.1002/wdev.23; pmid: 22662314
30. P. Sarasin, F. Sarasin, *Ergebnisse Naturwissenschaftlicher Forschungen auf Ceylon in der Jahren 1884-86* (C.W. Kreidel's Verlag, 1887–1890).
31. R. Keller, D. Shook, in *Gastrulation: From Cells to Embryo*, C.D. Stern, Ed. (Cold Spring Harbor Press, 2004), pp. 171–203.
32. J. S. Budgett, On the Breeding-habits of some West-African Fishes, with an Account of the External Features in Development of Protopterus annectens, and a Description of the Larva of Polypterus lapradei. *Trans. Zool. Soc. Lond.* **16**, 115–136 (1901). doi: 10.1111/j.1096-3642.1901.tb00028.x
33. J. G. Kerr, The external features in the development of *Lepidosiren paradoxa*, fitz. *Proc. R. Soc. London* **65**, 160–161 (1900). doi: 10.1098/rspl.1899.0017
34. T. Sauka-Spengler, B. Baratte, M. Lepage, S. Mazan, Characterization of Brachyury genes in the dogfish S. canicula and the lamprey L. fluviatilis. Insights into gastrulation in a chondrichthyan. *Dev. Biol.* **263**, 296–307 (2003). doi: 10.1016/j.ydbio.2003.07.009; pmid: 14597203
35. M. Takeuchi, M. Takahashi, M. Okabe, S. Aizawa, Germ layer patterning in bichir and lamprey; an insight into its evolution in vertebrates. *Dev. Biol.* **332**, 90–102 (2009). doi: 10.1016/j.ydbio.2009.05.543; pmid: 19433081
36. D. R. Shook, C. Majer, R. Keller, Urodeles remove mesoderm from the superficial layer by subduction through a bilateral primitive streak. *Dev. Biol.* **248**, 220–239 (2002). doi: 10.1006/dbio.2002.0718; pmid: 12167400
37. P. Chavatte-Palmer, M. Guillomot, Comparative implantation and placentation. *Gynecol. Obstet. Invest.* **64**, 166–174 (2007). doi: 10.1159/000101742; pmid: 17934314
38. A. M. Carter, A. C. Enders, The evolution of epitheliochorial placentation. *Annu. Rev. Anim. Biosci.* **1**, 443–467 (2013). doi: 10.1146/annurev-animal-031412-103653; pmid: 25387027
39. D. R. Shook, R. Keller, Morphogenic machines evolve more rapidly than the signals that pattern them: Lessons from amphibians. *J. Exp. Zool. B* **310**, 111–135 (2008). doi: 10.1002/jez.b.21204; pmid: 18041048
40. S. Hamidi *et al*., Mesenchymal-epithelial transition regulates initiation of pluripotency exit before gastrulation. *Development* **147**, dev184960 (2020). doi: 10.1242/dev.184960; pmid: 32014865
41. Y. Nakaya, E. W. Sukowati, Y. Wu, G. Sheng, RhoA and microtubule dynamics control cell-basement membrane interaction in EMT during gastrulation. *Nat. Cell Biol.* **10**, 765–775 (2008). doi: 10.1038/ncb1739; pmid: 18552836
42. E. D. Hay, An overview of epithelio-mesenchymal transformation. *Acta Anat.* **154**, 8–20 (1995). doi: 10.1159/000147748; pmid: 8714286
43. M. Williams, C. Burdsal, A. Periasamy, M. Lewandoski, A. Sutherland, Mouse primitive streak forms in situ by initiation of epithelial to mesenchymal transition without migration of a cell population. *Dev. Dyn.* **241**, 270–283 (2012). doi: 10.1002/dvdy.23711; pmid: 22170865
44. W. Nahaboo, B. Saykali, N. Mathiah, I. Migeotte, Visualizing Mouse Embryo Gastrulation Epithelial-Mesenchymal Transition Through Single Cell Labeling Followed by Ex Vivo Whole Embryo Live Imaging. *Methods Mol. Biol.* **2179**, 135–144 (2021). doi: 10.1007/978-1-0716-0779-4_12; pmid: 32939718
45. G. Sheng, Defining epithelial-mesenchymal transitions in animal development. *Development* **148**, dev198036 (2021). doi: 10.1242/dev.198036; pmid: 33913481
46. P. P. Tam, E. A. Williams, W. Y. Chan, Gastrulation in the mouse embryo: Ultrastructural and molecular aspects of germ layer morphogenesis. *Microsc. Res. Tech.* **26**, 301–328 (1993). doi: 10.1002/jemt.1070260405; pmid: 8305722
47. C. Viebahn, B. Mayer, M. H. de Angelis, Signs of the principle body axes prior to primitive streak formation in the rabbit embryo. *Anat. Embryol.* **192**, 159–169 (1995). doi: 10.1007/BF00186004; pmid: 7486012
48. Y. Nakaya, G. Sheng, An amicable separation: Chick's way of doing EMT. *Cell Adhes. Migr.* **3**, 160–163 (2009). doi: 10.4161/cam.3.2.7373; pmid: 19262172
49. E. A. Carver, R. Jiang, Y. Lan, K. F. Oram, T. Gridley, The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition. *Mol. Cell. Biol.* **21**, 8184–8188 (2001). doi: 10.1128/MCB.21.23.8184-8188.2001; pmid: 11689706
50. Y. Nakaya, E. W. Sukowati, G. Sheng, Epiblast integrity requires CLASP and Dystroglycan-mediated microtubule anchoring to the basal cortex. *J. Cell Biol.* **202**, 637–651 (2013). doi: 10.1083/jcb.201302075; pmid: 23940118
51. A. Lawson, G. C. Schoenwolf, Cell populations and morphogenetic movements underlying formation of the avian primitive streak and organizer. *Genesis* **29**, 188–195 (2001). doi: 10.1002/gene.1023; pmid: 11309852
52. E. Rozbicki *et al*., Myosin-II-mediated cell shape changes and cell intercalation contribute to primitive streak formation. *Nat. Cell Biol.* **17**, 397–408 (2015). doi: 10.1038/ncb3138; pmid: 25812521
53. M. Saadaoui, D. Rocancourt, J. Roussel, F. Corson, J. Gros, A tensile ring drives tissue flows to shape the gastrulating amniote embryo. *Science* **367**, 453–458 (2020). doi: 10.1126/science.aaw1965; pmid: 31974255
54. O. Voiculescu, F. Bertocchini, L. Wolpert, R. E. Keller, C. D. Stern, The amniote primitive streak is defined by epithelial cell intercalation before gastrulation. *Nature* **449**, 1049–1052 (2007). doi: 10.1038/nature06211; pmid: 17928866
55. V. Halacheva *et al*., Planar cell movements and oriented cell division during early primitive streak formation in the mammalian embryo. *Dev. Dyn.* **240**, 1905–1916 (2011). doi: 10.1002/dvdy.22687; pmid: 21761476
56. R. Hassoun, P. Schwartz, K. Feistel, M. Blum, C. Viebahn, Axial differentiation and early gastrulation stages of the pig embryo. *Differentiation* **78**, 301–311 (2009). doi: 10.1016/j.diff.2009.07.006; pmid: 19683851
57. V. Stankova, N. Tsikolia, C. Viebahn, Rho kinase activity controls directional cell movements during primitive streak formation in the rabbit embryo. *Development* **142**, 92–98 (2015). doi: 10.1242/dev.111583; pmid: 25516971
58. S. Ghimire, V. Mantziou, N. Moris, A. Martinez Arias, Human gastrulation: The embryo and its models. *Dev. Biol.* **474**, 100–108 (2021). doi: 10.1016/j.ydbio.2021.01.006; pmid: 33484705

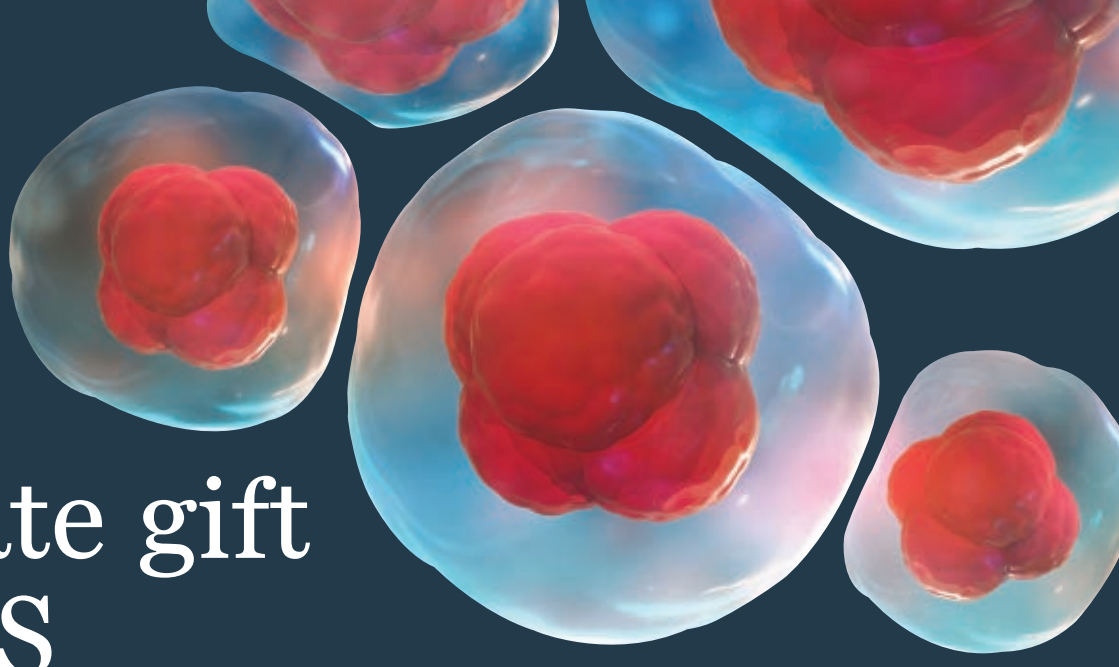
59. J. Rossant, P. P. L. Tam, Exploring early human embryo development. *Science* **360**, 1075–1076 (2018). doi: 10.1126/science.aas9302; pmid: 29880675
60. R. Bellairs, D. R. Bromham, C. C. Wylie, The influence of the area opaca on the development of the young chick embryo. *J. Embryol. Exp. Morphol.* **17**, 195–212 (1967). pmid: 6040550
61. D. A. New, The adhesive properties and expansion of the chick blastoderm. *J. Embryol. Exp. Morphol.* **7**, 146–164 (1959). pmid: 14426769
62. P. Caldarelli, A. Chamolly, O. Alegria-Prévot, J. Gros, F. Corson, Self-organized tissue mechanics underlie embryonic regulation. bioRxiv 2021.10.08.463661 [Preprint] (2021); doi: 10.1101/2021.10.08.463661
63. C. Viebahn, C. Stortz, S. A. Mitchell, M. Blum, Low proliferative and high migratory activity in the area of Brachyury expressing mesoderm progenitor cells in the gastrulating rabbit embryo. *Development* **129**, 2355–2365 (2002). doi: 10.1242/dev.129.10.2355; pmid: 11973268
64. Y. Imuta *et al*., Mechanical control of notochord morphogenesis by extra-embryonic tissues in mouse embryos. *Mech. Dev.* **132**, 44–58 (2014). doi: 10.1016/j.mod.2014.01.004; pmid: 24509350
65. D. Arendt, K. Nübler-Jung, Rearranging gastrulation in the name of yolk: Evolution of gastrulation in yolk-rich amniote eggs. *Mech. Dev.* **81**, 3–22 (1999). doi: 10.1016/S0925-4773(98)00226-3; pmid: 10330481
66. M. Chuai, G. Serrano Nájera, M. Serra, L. Mahadevan, C. J. Weijer, Reconstruction of distinct vertebrate gastrulation modes via modulation of key cell behaviours in the chick embryo. bioRxiv 2021.10.03.462938 [Preprint] (2021); doi: 10.1101/2021.10.03.462938
67. Z. S. Singer *et al*., Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55**, 319–331 (2014). doi: 10.1016/j.molcel.2014.06.029; pmid: 25038413
68. D. Gökbuget, R. Blelloch, Epigenetic control of transcriptional regulation in pluripotency and early differentiation. *Development* **146**, dev164772 (2019). doi: 10.1242/dev.164772; pmid: 31554624
69. D. A. Turner, P. Rué, J. P. Mackenzie, E. Davies, A. Martinez Arias, Brachyury cooperates with Wnt/β-catenin signalling to elicit primitive-streak-like behaviour in differentiating mouse embryonic stem cells. *BMC Biol.* **12**, 63 (2014). doi: 10.1186/s12915-014-0063-7; pmid: 25115237
70. A. Warmflash, B. Sorre, F. Etoc, E. D. Siggia, A. H. Brivanlou, A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* **11**, 847–854 (2014). doi: 10.1038/nmeth.3016; pmid: 24973948
71. S. Chhabra, L. Liu, R. Goh, X. Kong, A. Warmflash, Dissecting the dynamics of signaling events in the BMP, WNT, and NODAL cascade during self-organized fate patterning in human gastruloids. *PLOS Biol.* **17**, e3000498 (2019). doi: 10.1371/journal.pbio.3000498; pmid: 31613879
72. S. M. Morgani, J. J. Metzger, J. Nichols, E. D. Siggia, A. K. Hadjantonakis, Micropattern differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate patterning. *eLife* **7**, e32839 (2018). doi: 10.7554/eLife.32839; pmid: 29412136
73. A. Manfrin *et al*., Engineered signaling centers for the spatially controlled patterning of human pluripotent stem cells. *Nat. Methods* **16**, 640–648 (2019). doi: 10.1038/s41592-019-0455-2; pmid: 31249412
74. I. Martyn, E. D. Siggia, A. H. Brivanlou, Mapping cell migrations and fates in a gastruloid model to the human primitive streak. *Development* **146**, dev.179564 (2019). doi: 10.1242/dev.179564; pmid: 31427289
75. V. Wilson, L. Manson, W. C. Skarnes, R. S. Beddington, The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* **121**, 877–886 (1995). doi: 10.1242/dev.121.3.877; pmid: 7720590
76. C. Alev, Y. Wu, Y. Nakaya, G. Sheng, Decoupling of amniote gastrulation and streak formation reveals a morphogenetic unity in vertebrate mesoderm induction. *Development* **140**, 2691–2696 (2013). doi: 10.1242/dev.094318; pmid: 23698348
77. X. Xue *et al*., Mechanics-guided embryonic patterning of neuroectoderm tissue from human pluripotent stem cells. *Nat. Mater.* **17**, 633–641 (2018). doi: 10.1038/s41563-018-0082-9; pmid: 29784997
78. J. M. Muncie *et al*., Mechanical Tension Promotes Formation of Gastrulation-like Nodes and Patterns Mesoderm Specification in Human Embryonic Stem Cells. *Dev. Cell* **55**, 679–694.e11 (2020). doi: 10.1016/j.devcel.2020.10.015; pmid: 33207224
79. T. Brunet *et al*., Evolutionary conservation of early mesoderm specification by mechanotransduction in Bilateria. *Nat. Commun.* **4**, 2821 (2013). doi: 10.1038/ncomms3821; pmid: 24281726
80. S. C. van den Brink *et al*., Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* **141**, 4231–4242 (2014). doi: 10.1242/dev.113001; pmid: 25371360
81. L. Beccari *et al*., Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018). doi: 10.1038/s41586-018-0578-0; pmid: 30283134
82. A. Hashmi, S. Tlili, P. Perrin, A. Martinez-Arias, P.-F. Lenne, Cell-state transitions and collective cell movement generate an endoderm-like region in gastruloids. bioRxiv 2020.05.21.105551 [Preprint] (2020); doi: 10.1101/2020.05.21.105551
83. J. V. Veenvliet *et al*., Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science* **370**, eaba4937 (2020). doi: 10.1126/science.aba4937; pmid: 33303587
84. D. A. Turner, J. Trott, P. Hayward, P. Rué, A. Martinez Arias, An interplay between extracellular signalling and the dynamics of the exit from pluripotency drives cell fate decisions in mouse ES cells. *Biol. Open* **3**, 614–626 (2014). doi: 10.1242/bio.20148409; pmid: 24950969
85. N. Moris *et al*., An in vitro model of early anteroposterior organization during human development. *Nature* **582**, 410–415 (2020). doi: 10.1038/s41586-020-2383-9; pmid: 32528178
86. S. Vianello, M. P. Lutolf, In vitro endoderm emergence and self-organisation in the absence of extraembryonic tissues and embryonic architecture. bioRxiv 2020.06.07.138883 [Preprint] (2020); doi: 10.1101/2020.06.07.138883
87. P. F. Xu *et al*., Construction of a mammalian embryo model from stem cells organized by a morphogen signalling centre. *Nat. Commun.* **12**, 3277 (2021). doi: 10.1038/s41467-021-23653-4; pmid: 34078907
88. M. U. Girgin *et al*., Bioengineered embryoids mimic post-implantation development in vitro. *Nat. Commun.* **12**, 5140 (2021); doi: 10.1038/s41467-021-25237-8; pmid: 34446708
89. B. Sozen *et al*., Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat. Cell Biol.* **20**, 979–989 (2018). doi: 10.1038/s41556-018-0147-7; pmid: 30038254
90. G. Amadei *et al*., Inducible Stem-Cell-Derived Embryos Capture Mouse Morphogenetic Events In Vitro. *Dev. Cell* **56**, 366–382.e9 (2021). doi: 10.1016/j.devcel.2020.12.004; pmid: 33378662
91. Y. Shao *et al*., A pluripotent stem cell-based model for post-implantation human amniotic sac development. *Nat. Commun.* **8**, 208 (2017). doi: 10.1038/s41467-017-00236-w; pmid: 28785084
92. G. Sheng, Epiblast morphogenesis before gastrulation. *Dev. Biol.* **401**, 17–24 (2015). doi: 10.1016/j.ydbio.2014.10.003; pmid: 25446532
93. S. Franklin, Developmental Landmarks and the Warnock Report: A Sociological Account of Biological Translation. *Comp. Stud. Soc. Hist.* **61**, 743–773 (2019). doi: 10.1017/S0010417519000252

ACKNOWLEDGMENTS

We thank B. Steventon, N. Moris, and C. Stern for discussions. **Funding:** G.S. is supported by JST e-ASIA Joint Research Project JPMJSC19E5, JSPS Kakenhi 18H02452 and 21H02490, and Takeda Science Foundation grants. A.M.A. is supported by an ERC Advanced Investigator Grant (MiniEmbryoBluePrint–834580). A.S. is supported by a grant from the National Institute for Child Health and Human Development (1R01HD087093). **Competing interests:** G.S. and A.S. declare no competing interests. A.M.E. is an inventor on two patent applications (PCT/GB2019/052668 and PCT/GB2019/052670) filed by the University of Cambridge that cover the generation of mouse and human gastruloids.

10.1126/science.abg1727

1848 SOCIETY
AAAS



An estate gift to AAAS

Going all the way back to 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

1848 SOCIETY
AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern

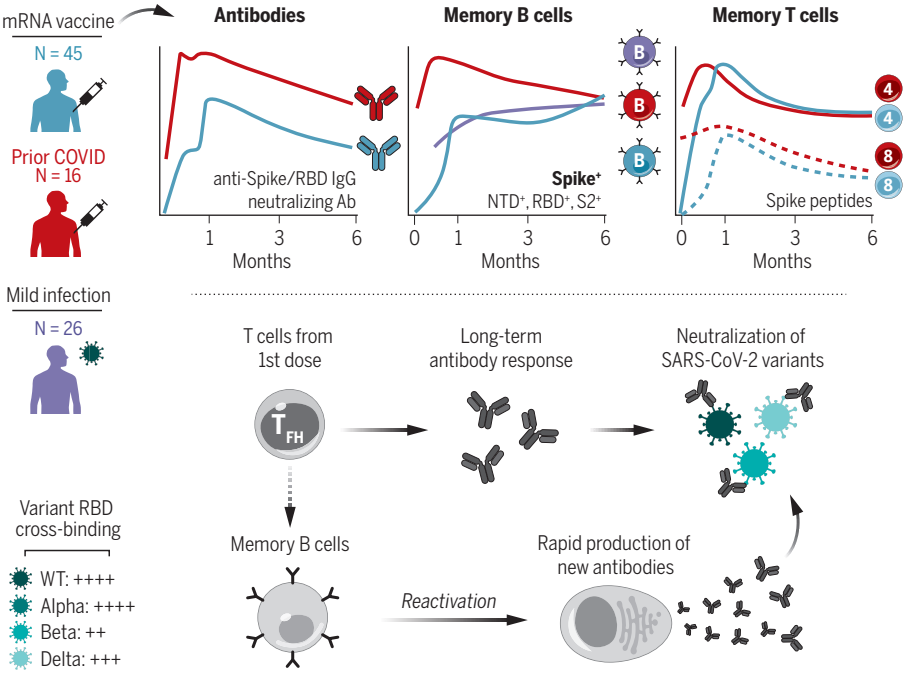
Rishi R. Goel†, Mark M. Painter†, Sokratis A. Apostolidis†, Divij Mathew†, Wenzhao Meng, Aaron M. Rosenfeld, Kendall A. Lundgreen, Arnold Reynaldi, David S. Khoury, Ajinkya Pattekar, Sigrid Gouma, Leticia Kuri-Cervantes, Philip Hicks, Sarah Dysinger, Amanda Hicks, Harsh Sharma, Sarah Herring, Scott Korte, Amy E. Baxter, Derek A. Oldridge, Josephine R. Giles, Madison E. Weirick, Christopher M. McAllister, Moses Awofolaju, Nicole Tanenbaum, Elizabeth M. Drapeau, Jeanette Dougherty, Sherea Long, Kurt D'Andrea, Jacob T. Hamilton, Maura McLaughlin, Justine C. Williams, Sharon Adamski, Oliva Kuthuru, The UPenn COVID Processing Unit, Ian Frank, Michael R. Betts, Laura A. Vella, Alba Grifoni, Daniela Weiskopf, Alessandro Sette, Scott E. Hensley, Miles P. Davenport, Paul Bates, Eline T. Luning Prak, Allison R. Greenplate, E. John Wherry*

INTRODUCTION: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) mRNA vaccines are highly effective at preventing infection and especially severe disease. However, the emergence of variants of concern (VOCs) and increasing infections in vaccinated individuals have raised questions about the durability of immunity after vaccination.

RATIONALE: To study immune memory, we longitudinally profiled antigen-specific antibody, memory B cell, and memory T cell responses in 61 individuals receiving mRNA vaccines from baseline to 6 months postvaccination. A subgroup of 16 individuals had recovered from prior SARS-CoV-2 infection, providing insight into boosting preexisting immunity with mRNA vaccines.

RESULTS: mRNA vaccination induced robust anti-spike, anti–receptor binding domain (RBD), and neutralizing antibodies that remained above prevaccine baseline levels in most individuals at 6 months postvaccination, although antibodies did decline over time. mRNA vaccination also generated spike- and RBD-specific memory B cells, including memory B cells that cross-bound Alpha, Beta, and Delta RBDs, that were capable of rapidly producing functional antibodies after stimulation. Notably, the frequency of SARS-CoV-2–specific memory B cells continued to increase from 3 to 6 months postvaccination. mRNA vaccines also generated a higher frequency of variant cross-binding memory B cells than mild SARS-CoV-2 infection alone, with >50% of RBD-specific memory B cells cross-binding all three VOCs at 6 months. These variant-binding memory B cells were more hypermutated than wild-type–only binding cells. SARS-CoV-2–specific memory $CD4^+$ and $CD8^+$ T cell responses contracted from peak levels after the second vaccine dose, with relative stabilization of SARS-CoV-2–specific memory $CD4^+$ T cells from 3 to 6 months. T follicular helper cell responses after the first vaccine dose correlated with antibodies at 6 months, highlighting a key role for early $CD4^+$ T cell responses. Finally, recall responses to mRNA vaccination in individuals with preexisting immunity led to an increase in circulating antibody titers that correlated with preexisting memory B cell frequency. However, there was no substantial increase in the long-term frequency of memory B and T cells. There was also no significant difference in the decay rates of antibodies in SARS-CoV-2–naïve versus –recovered subjects after vaccination, which suggests that the main benefit of recall responses to mRNA vaccination may be a robust but transient increase in circulating antibodies.

CONCLUSION: These findings demonstrate multicomponent immune memory after SARS-CoV-2 mRNA vaccination, with memory B and T cell responses remaining durable even as antibodies decline. Immune memory was resilient to VOCs and generated an efficient recall response upon antigen reexposure. These durable memory cells may be responsible for continued protection against severe disease in vaccinated individuals, despite a gradual reduction in antibodies. Our data may also inform expectations for the immunological outcomes of booster vaccination. ■



Immune memory after mRNA vaccination. SARS-CoV-2–specific antibody, memory B, and memory T cell responses were measured at six time points after vaccination, highlighting a coordinated evolution of durable immunological memory. B cell memory was also resilient to VOCs and capable of producing new antibodies upon reactivation. IgG, immunoglobulin G; Ab, antibody; NTD, N-terminal domain; T_{FH}, T follicular helper cell; WT, wild-type.

The list of author affiliations is available in the full article online.
***Corresponding author. Email: wherry@pennmedicine.upenn.edu**
†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 R. R. Goel *et al.*, *Science* 374, eabm0829 (2021). DOI: 10.1126/science.abm0829

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm0829

RESEARCH ARTICLE

CORONAVIRUS

mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern

Rishi R. Goel[1,2]†, Mark M. Painter[1,2]†, Sokratis A. Apostolidis[1,2,3]†, Divij Mathew[1,2]†, Wenzhao Meng[1,4], Aaron M. Rosenfeld[1,4], Kendall A. Lundgreen[5], Arnold Reynaldi[6], David S. Khoury[6], Ajinkya Pattekar[2], Sigrid Gouma[5], Leticia Kuri-Cervantes[1,5], Philip Hicks[5], Sarah Dysinger[5], Amanda Hicks[2], Harsh Sharma[2], Sarah Herring[2], Scott Korte[2], Amy E. Baxter[1], Derek A. Oldridge[1,4], Josephine R. Giles[1,7,8], Madison E. Weirick[5], Christopher M. McAllister[5], Moses Awofolaju[5], Nicole Tanenbaum[5], Elizabeth M. Drapeau[5], Jeanette Dougherty[1], Sherea Long[1], Kurt D'Andrea[1], Jacob T. Hamilton[2,5], Maura McLaughlin[1], Justine C. Williams[2], Sharon Adamski[2], Oliva Kuthuru[1], The UPenn COVID Processing Unit‡, Ian Frank[9], Michael R. Betts[1,5], Laura A. Vella[10], Alba Grifoni[11], Daniela Weiskopf[11], Alessandro Sette[11,12], Scott E. Hensley[5], Miles P. Davenport[6], Paul Bates[5], Eline T. Luning Prak[1,4], Allison R. Greenplate[1,2], E. John Wherry[1,2,7,8]*

The durability of immune memory after severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) messenger RNA (mRNA) vaccination remains unclear. In this study, we longitudinally profiled vaccine responses in SARS-CoV-2–naïve and –recovered individuals for 6 months after vaccination. Antibodies declined from peak levels but remained detectable in most subjects at 6 months. By contrast, mRNA vaccines generated functional memory B cells that increased from 3 to 6 months postvaccination, with the majority of these cells cross-binding the Alpha, Beta, and Delta variants. mRNA vaccination further induced antigen-specific $CD4^+$ and $CD8^+$ T cells, and early $CD4^+$ T cell responses correlated with long-term humoral immunity. Recall responses to vaccination in individuals with preexisting immunity primarily increased antibody levels without substantially altering antibody decay rates. Together, these findings demonstrate robust cellular immune memory to SARS-CoV-2 and its variants for at least 6 months after mRNA vaccination.

The COVID-19 pandemic has resulted in substantial morbidity and mortality worldwide. Community-level immunity, acquired through infection or vaccination, is necessary to control the pandemic as the virus continues to circulate (*1*). mRNA vaccines encoding a stabilized version of the full-length severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein have been widely administered, and clinical trial data have demonstrated up to 95% efficacy in preventing symptomatic COVID-19 (*2*, *3*). These mRNA vaccines induce potent humoral immune responses, with neutralizing antibody titers proposed as the major correlate of protection (*4*–*6*). Current evidence suggests that circulating antibodies persist for at least 6 months postvaccination (*7*), although there is some decay from the peak levels achieved after the second dose. This decline from peak antibody levels may be associated with an increase in infections over time compared with the initial months postvaccination (*8*, *9*). Yet, vaccine-induced immunity remains effective at preventing severe disease, hospitalization, and death, even at later time points when antibody levels may have declined (*10*–*12*).

Previous research has largely focused on responses early in the course of vaccination, with transcriptional analysis identifying potential links between myeloid cell responses and neutralizing antibodies (*13*). In addition to the production of antibodies, an effective immune response requires the generation of long-lived memory B and T cells. mRNA vaccines induce robust germinal center responses in humans (*14*, *15*), which result in memory B cells that are specific for both the full-length SARS-CoV-2 spike protein and the spike receptor-binding domain (RBD) (*16*–*18*). mRNA vaccination has also been shown to generate spike-specific memory $CD4^+$ and $CD8^+$ T cell responses (*19*–*22*). Although antibodies are often correlates of vaccine efficacy, memory B cells and memory T cells are important components of the recall response to viral antigens and are a likely mechanism of protection, especially in the setting of exposures in previously vaccinated individuals, where antibodies alone do not provide sterilizing immunity (*23*). In such cases, memory B and T cells can be rapidly reactivated, resulting in the enhanced control of initial viral replication and limiting viral dissemination in the host (*24*, *25*). By responding and restricting viral infection within the first hours to days after exposure, cellular immunity can thereby reduce or even prevent symptoms of disease (i.e., preventing hospitalization and death) and potentially reduce the ability to spread virus to others (*26*, *27*).

Immunological studies of SARS-CoV-2 infection show that memory B and T cell responses appear to persist for at least 8 months after symptom onset (*28*, *29*). However, the durability of these populations of memory B and T cells after vaccination remains poorly understood. The emergence of several SARS-CoV-2 variants, including B.1.1.7 (Alpha), B.1.351 (Beta), and B.1.617.2 (Delta), has also raised concerns about increased transmission and potential evasion from vaccine-induced immunity (*30*–*33*). As such, it is necessary to develop a more-complete understanding of the trajectory and durability of immunological memory after mRNA vaccination, as well as how immune responses are affected by current variants of concern (VOCs). Moreover, the United States and other well-resourced countries have recently announced plans for a third vaccine booster dose, yet information on how preexisting serological and cellular immunity to SARS-CoV-2 are boosted by mRNA vaccination remains limited. Specifically, it is unclear how different components of the immune response may benefit from boosting and whether boosting has any effect on the durability of these components. Here, we investigated these key questions by measuring SARS-CoV-2–specific antibody, memory B cell, and memory T cell responses through 6 months postvaccination in a group of healthy subjects generating primary immune responses to two doses of mRNA

[1]Institute for Immunology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [2]Immune Health, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [3]Division of Rheumatology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [4]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [5]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [6]Kirby Institute, University of New South Wales, Sydney, NSW, Australia. [7]Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [8]Parker Institute for Cancer Immunotherapy, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [9]Division of Infectious Disease, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [10]Division of Infectious Disease, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [11]Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology (LJI), La Jolla, CA, USA. [12]Department of Medicine, Division of Infectious Diseases and Global Public Health, University of California San Diego (UCSD), La Jolla, CA, USA.

*Corresponding author. Email: wherry@pennmedicine.upenn.edu

†These authors contributed equally to this work.

‡The UPenn COVID Processing Unit includes individuals from diverse laboratories at the University of Pennsylvania who volunteered their time and effort to enable study of COVID-19 patients during the pandemic. Members and affiliations are listed at the end of this paper.

vaccine compared with a group of SARS-CoV-2–recovered vaccinees generating recall responses from preexisting immunity. These analyses provide insights into mRNA vaccine–induced immunological memory and may be relevant for future vaccine strategies, including recommendations for additional booster vaccine doses.

Results and discussion

Cohort design

We collected 348 longitudinal samples from 61 individuals receiving either the Pfizer BNT162b2 (N = 54) or Moderna mRNA-1273 (N = 7) SARS-CoV-2 vaccines at six time points (Fig. 1A), ranging from prevaccination baseline to 6 months postvaccination. This study design allowed us to monitor the induction and maintenance of antigen-specific immune responses to the vaccine. Specifically, sampling at 1, 3, and 6 months postvaccination enabled the analysis of immune trajectories from peak responses after the second vaccine dose through the establishment and maintenance of immunological memory. This cohort was divided into two groups on the basis of prior SARS-CoV-2 infection (N = 45 SARS-CoV-2–naïve individuals; N = 16 SARS-CoV-2–recovered individuals). Age and sex were balanced in both groups. Paired serum and peripheral blood mononuclear cell (PBMC) samples were collected from all individuals, allowing for a detailed analysis of both serologic and cellular immune memory to SARS-CoV-2 antigens. Notably, the subjects with a prior infection allowed us to study the dynamics of reactivating preexisting immunity with mRNA vaccines. Though preexisting immunity generated by infection may differ from that generated by vaccination, responses observed in this group may provide insights into the boosting of vaccine-induced immunity using additional doses of vaccine.

Antibody responses to SARS-CoV-2 mRNA vaccines

We first measured anti-spike and anti-RBD binding antibody responses in plasma samples by enzyme-linked immunosorbent assay (ELISA). As reported previously by our group and others, mRNA vaccines induced robust circulating antibody responses to the SARS-CoV-2 spike protein and spike RBD with distinct patterns of early response in SARS-CoV-2–naïve and –recovered individuals (Fig. 1B) (*16*, *34*–*36*).

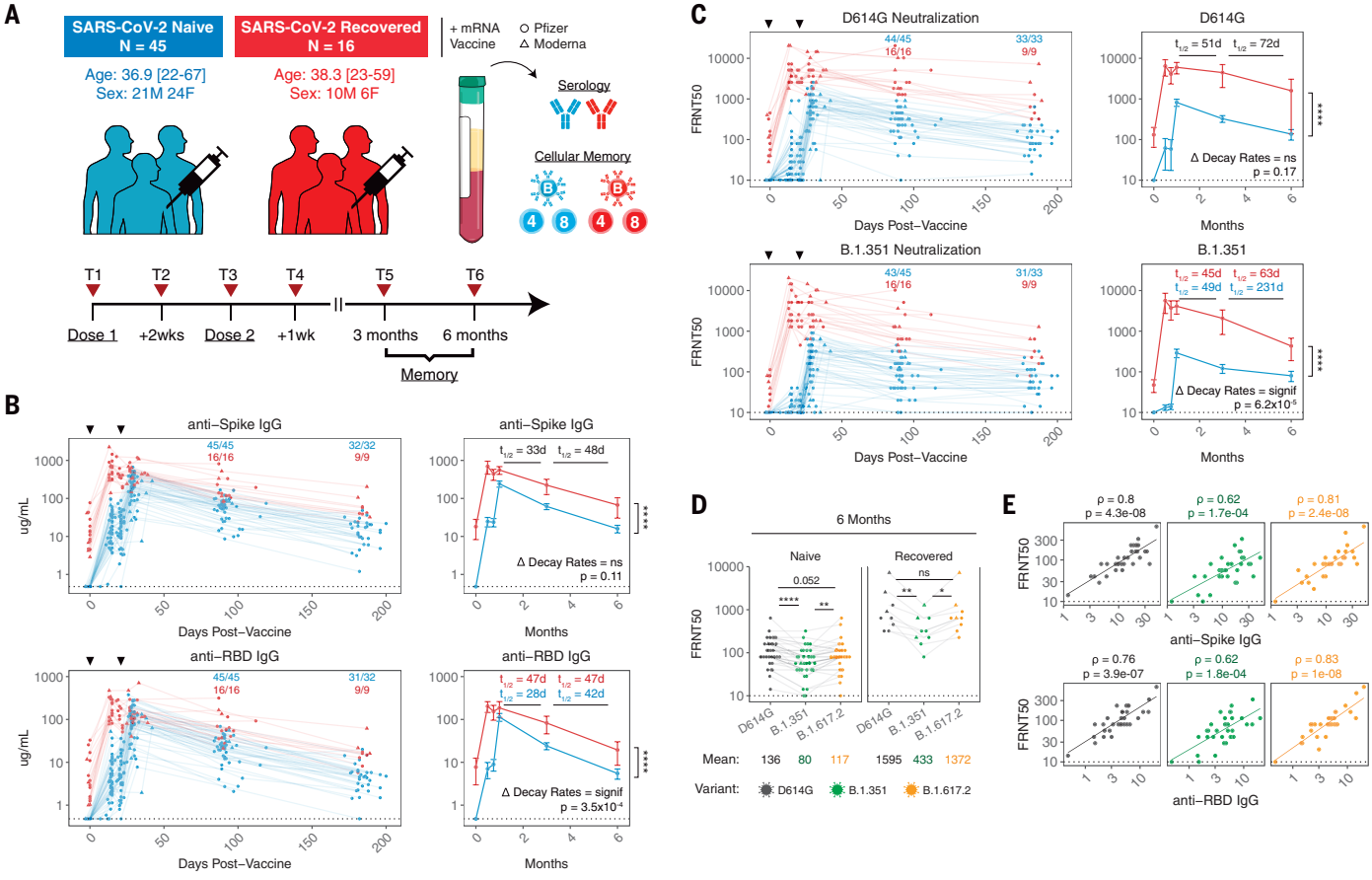


Fig. 1. SARS-CoV-2 mRNA vaccines induce robust antibody responses. (**A**) University of Pennsylvania COVID-19 vaccine study design and cohort summary statistics. (**B**) Anti-spike and anti-RBD IgG concentrations over time in plasma samples from vaccinated individuals. (**C**) Pseudovirus neutralization titers against WT D614G or B.1.351 variant spike protein over time in plasma samples from vaccinated individuals. Data are represented as focus reduction neutralization titer 50% ($FRNT_{50}$) values. (**D**) Comparison of D614G, B.1.351, and B.1.617.2 $FRNT_{50}$ values at 6 months postvaccination. (**E**) Correlation between anti-spike or anti-RBD IgG and neutralizing titers (D614G = black, B.1.351 = green, and B.1.617.2 = orange; statistics were calculated using nonparametric Spearman rank correlation). Dotted lines indicate the limit of detection for the assay. For (B) and (C), black triangles indicate time of vaccine doses, fractions above plots indicate the number of individuals above their individual baseline at memory time points, and summary plots show mean values with the 95% confidence interval. Decay rates were calculated using a piecewise linear mixed-effects model with censoring. Changes in decay rate over time (linear versus two-phase decay) were determined on the basis of a likelihood ratio test. Δ Decay Rates indicates whether decay rates were different in SARS-CoV-2–naïve and –recovered groups. Statistics were calculated using unpaired [(B) and (C)] or paired (D) nonparametric Wilcoxon test with Benjamini-Hochberg (BH) correction. Blue and red values indicate comparisons within naïve or recovered groups. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; ns, not significant.

Peak levels of anti-spike and anti-RBD immunoglobulin G (IgG) were observed 1 week after the second vaccine dose and subsequently declined over the course of the next 2 months with a half-life of ~28 to 33 days (Fig. 1B), consistent with the dynamics of a typical immune response. This decrease in antibody levels slowed from 3 to 6 months postvaccination (decay rates were significantly different before and after day 89 by likelihood ratio rest; $P = 0.004$ for anti-spike IgG; $P = 0.01$ for anti-RBD IgG) (Fig. 1B). Notably, the calculated decay rates for anti-spike IgG were not significantly different between SARS-CoV-2–naïve and –recovered vaccinees. Even after the decrease from peak antibody responses, all individuals had detectable anti-spike IgG at 6 months.

To examine the functional quality of circulating antibodies, we used a neutralization assay with pseudotyped virus expressing either the wild-type (WT) spike with the prevailing D614G mutation or the B.1.351 variant spike (sequences are provided in the Materials and methods). We focused on B.1.351 neutralization because this variant has consistently shown the highest immune evasion among the current VOCs. In line with our binding antibody data, neutralizing titers for D614G and B.1.351 declined from peak levels after the second dose to 6 months for both SARS-CoV-2–naïve and –recovered vaccinees (Fig. 1C). However, neutralizing titers displayed different decay kinetics, with slightly longer half-lives than binding antibody responses. Modeled two-phase decay rates for D614G neutralization were not significantly different between SARS-CoV-2–naïve and –recovered vaccinees with a half-life of 72 days between 3 to 6 months postvaccination (Fig. 1C). By contrast, a relative stabilization of neutralizing titers against the B.1.351 variant was observed between 3 and 6 months postvaccination in individuals without a prior SARS-CoV-2 infection with a half-life of 231 days, compared with 63 days in SARS-CoV-2–recovered subjects (Fig. 1C). We next compared neutralizing titers to D614G, B.1.351, and B.1.617.2 at 6 months postvaccination. Neutralizing antibody titers to B.1.617.2 were similar to D614G (Fig. 1D). By contrast, neutralizing titers to B.1.351 were significantly lower than D614G. Despite this reduced neutralizing ability, 31 of 33 SARS-CoV-2–naïve and 9 of 9 SARS-CoV-2–recovered individuals still had neutralizing antibodies against B.1.351 above the limit of detection at 6 months postvaccination (Fig. 1, C and D). Finally, cross-sectional analysis of 6-month antibody responses also demonstrated that binding antibodies remained highly correlated with neutralizing titers (Fig. 1E), indicating that spike- and RBD-specific antibody responses retain their functional characteristics and neutralizing capacity over time.

Memory B cell responses to SARS-CoV-2 mRNA vaccines

In addition to antibodies, we measured the frequencies of SARS-CoV-2 spike- and RBD-specific memory B cells in peripheral blood using a flow cytometric assay. Antigen specificity was determined on the basis of binding to fluorescent SARS-CoV-2 spike and RBD probes (Fig. 2, A and B). Influenza hemagglutinin (HA) from the 2019 flu vaccine season was also included as a historical antigen control. Full gating strategies are provided in fig. S1A.

SARS-CoV-2–specific memory B cells were detectable in all previously uninfected individuals after two vaccine doses (the currently recommended primary vaccination series) and remained stable as a percentage of total B cells from 1 to 3 months postvaccination (Fig. 2C). All SARS-CoV-2–recovered individuals in our study had a robust population of antigen-specific memory B cells at prevaccination baseline, and these preexisting memory B cells were significantly boosted by the first vaccine dose with little change after the second vaccine dose (Fig. 2C). No changes were observed in influenza HA^+ memory B cells after SARS-CoV-2 vaccination for either group (Fig. 2C).

Longitudinal analysis revealed a continued increase in the frequency of $spike^+$ and $spike^+$ RBD^+ memory B cells from 3 to 6 months postvaccination in SARS-CoV-2–naïve individuals, whereas the frequency of these antigen-specific memory B cells in SARS-CoV-2–recovered subjects continued to decline from peak levels (Fig. 2C). One possible explanation for the observed increase in frequency of vaccine-induced memory B cells over time in SARS-CoV-2–naïve vaccinees is prolonged germinal center activity, resulting in continued export of memory B cells. Antigen-specific germinal center B cells have been documented in axillary lymph nodes at 15 weeks after mRNA vaccination in SARS-CoV-2–naïve subjects (*14*), though germinal center dynamics in vaccinees with prior immunity to SARS-CoV-2 remain to be defined. SARS-CoV-2–recovered individuals had consistently higher frequencies of antigen-specific memory B cells up to 3 months postvaccination (Fig. 2C). However, because of distinct trajectories, both SARS-CoV-2–naïve and SARS-CoV-2–recovered individuals had similar frequencies of $spike^+$ and $spike^+$ RBD^+ memory B cells at 6 months postvaccination (Fig. 2C), perhaps reflecting some upper limit to the frequencies of antigen-specific memory B cells that can be maintained long-term.

We next investigated the phenotype of mRNA vaccine–induced memory B cells. Analysis of immunoglobulin isotypes in SARS-CoV-2–naïve vaccinees revealed a steady increase in IgG^+ memory B cells over time (Fig. 2, D and E, and fig. S2, A to C), indicating ongoing class-switching. By contrast, IgM^+ cells were most abundant at preimmune baseline and early postvaccination time points. IgM^+ and IgA^+ memory B cells represented a minor fraction of the overall response in the blood at later time points (Fig. 2F and fig. S2C). In SARS-CoV-2–recovered vaccinees, the majority of $spike^+$ and $spike^+$ RBD^+ memory B cells were IgG^+ at baseline, and the fraction of IgG^+ cells continued to increase after vaccination (Fig. 2, D and E, and fig. S2, A to C). Moreover, we assessed the activation status of antigen-specific memory B cells by CD71 expression (*37*). The percent of $spike^+$ memory B cells expressing CD71 increased over the course of the primary two-dose vaccine regimen in SARS-CoV-2–naïve individuals, peaking at 1 week after the second vaccine dose (Fig. 2G). The percent of $CD71^+$ antigen–specific memory B cells then steadily declined by the 6-month time point, indicating a transition toward a population of mature resting memory B cells. A similar decrease in CD71 expression was observed from 1 to 6 months postvaccination in SARS-CoV-2–recovered individuals (Fig. 2G).

Given the robust generation of spike- and RBD-binding memory B cells, we next tested whether vaccine-induced memory B cells could produce functional antibodies upon reactivation. This reactivation-induced antibody production from memory B cells may be especially relevant in the setting of antigen reencounter, either through exposure to live virus or an additional vaccine dose (*38*). To this end, we established an in vitro culture system to differentiate memory B cells into antibody-secreting cells (*39*). PBMC samples from vaccinated individuals at the 6-month time point were cultured with a combination of R848, a Toll-like receptor 7 (TLR7) and TLR8 agonist, and interleukin-2 (IL-2), and culture supernatants were collected to measure antibody levels and function (Fig. 2H). Anti-spike IgG was detected in supernatants as early as 4 days after stimulation (Fig. 2I), indicating that memory B cells can act as a rapid source of secondary antibody production. All 6-month samples tested generated significant levels of anti-spike IgG in this assay compared with unstimulated controls (Fig. 2J). This in vitro anti-spike IgG production also correlated with the frequency of $spike^+$ memory B cells detected by flow cytometry (Fig. 2K). We further tested the function of memory B cell–derived antibodies from culture supernatants using an ELISA-based RBD–angiotensin-converting enzyme 2 (ACE2) binding inhibition assay. RBD-ACE2 binding inhibition activity was observed and correlated with the frequency of RBD-specific memory B cells in peripheral blood (Fig. 2L). Moreover, pseudovirus neutralization assays demonstrated that antibodies produced by memory B cells upon restimulation were capable of neutralizing the B.1.351 and B.1.617.2 VOCs (Fig. 2M), and

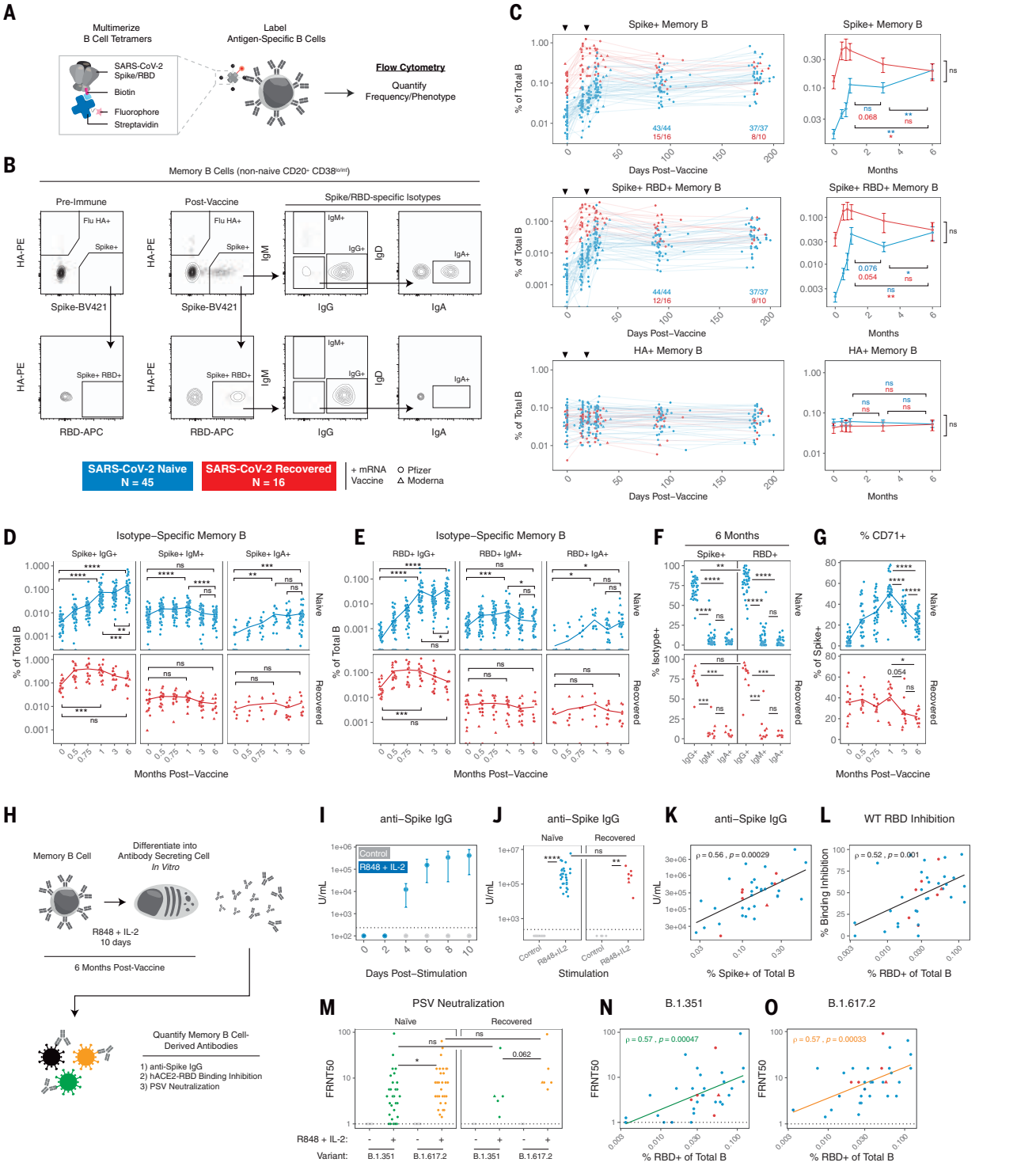


Fig. 2. SARS-CoV-2 mRNA vaccines generate durable and functional memory B cell responses. (**A** and **B**) Experimental design (A) and gating strategy (B) for quantifying the frequency and phenotype of SARS-CoV-2–specific memory B cells by flow cytometry. Antigen specificity was determined on the basis of binding to fluorophore-labeled spike, RBD, and influenza HA tetramers. (**C**) Frequencies of SARS-CoV-2 spike$^+$, spike$^+$ RBD$^+$, and influenza HA$^+$ memory B cells over time in PBMC samples from vaccinated individuals. Data are represented as a percentage of total B cells, black triangles indicate time of vaccine doses, fractions below plots indicate the number of individuals above their individual baseline at memory time points, and summary plots show mean values with the 95% confidence interval. (**D** and **E**) Frequency of isotype-specific spike$^+$ (D) and spike$^+$ RBD$^+$ (E) memory B cells over time. IgA was assessed on a subset of subjects. (**F**) Percent IgG$^+$, IgM$^+$, or IgA$^+$ of SARS-CoV-2–specific memory B cells at 6 months postvaccination. (**G**) Percent CD71$^+$ of total spike$^+$ memory B cells over time. (**H**) Experimental design for in vitro differentiation of memory B cells into antibody-secreting cells. (**I**) Anti-spike IgG levels in culture supernatants over time from PBMCs stimulated with PBS control or R848 + IL-2 (n = 4). (**J**) Anti-spike IgG levels in culture supernatants after 10 days of stimulation (**K**) Correlation of spike$^+$ memory B cell frequencies by flow

cytometry with anti-spike IgG levels from in vitro stimulation. (**L**) Correlation of RBD$^+$ memory B cell frequencies by flow cytometry with hACE2-RBD binding inhibition from in vitro stimulation. (**M**) Pseudovirus (PSV) neutralizing titers against B.1.351 and B.1.617.2 variants in culture supernatants after 10 days of stimulation. (**N** and **O**) Correlation of RBD$^+$ memory B cell frequencies by flow cytometry with PSV neutralizing titers of memory B cell–derived antibodies against B.1.351 (N) and B.1.617.2 (O). For (D), (E), and (G), lines connect mean values at different time points. For (K), (L), (N), and (O), correlations were calculated using nonparametric Spearman rank correlation. Dotted lines indicate the limit of detection of the assay. Statistics were calculated using unpaired nonparametric Wilcoxon test with BH correction for multiple comparisons. Blue and red values indicate comparisons within naïve or recovered groups. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; ns, not significant.

neutralization titers correlated with both anti-spike IgG and RBD-ACE2 binding inhibition (fig. S3, A to D). The neutralization potential of memory B cell–derived antibodies was greater for B.1.617.2 than B.1.351 but was not significantly different between SARS-CoV-2–naïve and –recovered vaccinees. Finally, VOC neutralizing titers in culture supernatants correlated with the frequency of RBD-specific memory B cells by flow cytometry (Fig. 2, N and O), further supporting the functional relevance of quantifying antigen-specific memory B cells in the blood. Taken together, these data demonstrate that mRNA vaccines induced a population of memory B cells that was durable for at least 6 months after vaccination, and these cells were capable of rapidly producing functional antibodies against SARS-CoV-2, including neutralizing antibodies against VOCs, upon stimulation.

Memory B cell responses to major VOCs

We next developed an expanded antigen probe panel to better quantify memory B cell specificities to different regions of the spike protein and test how RBD binding by memory B cells may be affected by the mutations found in emerging VOCs. Specifically, we designed B cell tetramers for eight SARS-CoV-2 antigens, including full-length spike, N-terminal domain (NTD), multiple variant RBDs (WT, B.1.1.7, B.1.351, and B.1.617.2), and the S2 domain (Fig. 3, A and B). Spike-specific memory B cells were defined on the basis of a multiple-discrimination approach, with binding to full-length spike plus one or more additional probes. This strategy also allowed us to identify memory B cells that cross-bind all variant RBDs (all variant$^+$). SARS-CoV-2 nucleocapsid was used as a vaccine-irrelevant antigen (but one for which SARS-CoV-2–infected subjects had detectable preexisting immunity; fig. S4, A and B). Full gating strategies are provided in fig. S1B. We also leveraged a separate cohort of health care workers (HCWs) (table S1) who had mild COVID-19 and were sampled longitudinally after a positive serology test to compare vaccine-induced responses with infection alone (*40*).

mRNA vaccination induced robust memory B cell responses to all SARS-CoV-2 spike antigens in previously uninfected individuals, and the frequency of these memory B cells increased from 3 to 6 months postvaccination (Fig. 3C). In individuals with immunity from prior COVID-19, vaccination resulted in a significant expansion of memory B cells targeting all spike antigens. These responses subsequently contracted from peak levels, remaining slightly above prevaccination frequencies at 6 months postvaccination (Fig. 3C). In the mild-infection HCW cohort, a gradual increase in the frequency of spike$^+$ NTD$^+$ and spike$^+$ all variant$^+$ memory B cells was observed from 2 weeks to 6 months after seropositive test (Fig. 3C). Cross-sectional analysis at 6 months postvaccination or seropositivity revealed similar antigen-specific memory B cell frequencies between all groups (fig. S4B), suggesting that both vaccination and infection can induce durable memory B cell populations.

Because our panel included probes covering much of the spike protein, including NTD, RBD, and S2, we also examined immunodominance patterns and how B cell immunodominance to spike changed over time. In previously uninfected individuals, ~30% of spike-binding memory B cells cobound S2 at prevaccine baseline (Fig. 3D). Previous work has shown that the S2 domain of SARS-CoV-2 spike is more conserved with other coronaviruses, and it is likely that S2-binding memory B cells detected at baseline reflect cross-reactivity to these commonly circulating coronaviruses (*41*, *42*). mRNA vaccination induced robust populations of S2-specific memory B cells in SARS-CoV-2–naïve vaccinees, with S2-binding B cells accounting for 40 to 80% of the total spike-specific memory B cell population at 6 months (Fig. 3D). Although the overall frequency of NTD$^+$ and RBD$^+$ memory B cells increased over time, they were comparatively less immunodominant than S2 as a percentage of total spike$^+$ memory B cells (Fig. 3, C and D). mRNA vaccination induced a gradual increase in NTD specificity over time in SARS-CoV-2–naïve individuals, whereas RBD specificity as a percent of spike$^+$ memory B cells had a more prominent peak 1 week after the second vaccine dose and then stabilized from 3 to 6 months postvaccination (Fig. 3D). When SARS-CoV-2–recovered subjects were immunized with mRNA vaccine, a similar immunodominance pattern was observed, with S2 specificity representing most of the total anti-spike response (Fig. 3D). Vaccination transiently increased NTD and RBD specificity in this group; however, this effect returned to baseline by 6 months postvaccination. In the context of infection only, we found that NTD, RBD, and S2 immunodominance remained relatively stable from early convalescence through late memory, with a slight increase in NTD specificity over time (Fig. 3D).

We next examined memory B cell binding to B.1.1.7 (Alpha), B.1.351 (Beta), and B.1.617.2 (Delta) variant RBDs relative to WT RBD (Fig. 3, E and F, and fig. S4, C and D). All RBD probes were used at the same concentration to facilitate direct comparisons, and specific point mutations are shown in Fig. 3, A and B. Variant-binding memory B cells were detectable in all SARS-CoV-2–naïve individuals after two vaccine doses and were stable as a percentage of WT RBD$^+$ cells from 1 to 6 months postvaccination (Fig. 3F). In SARS-CoV-2–recovered individuals, vaccination resulted in a significant increase in memory B cell cross-binding to the B.1.617.2 variant (Fig. 3F). In convalescent individuals who recovered from a mild infection, there was a gradual increase in cross-binding to variants over time (Fig. 3F). Class-switching to an IgG-dominated response was also observed in all groups, with vaccination producing a higher percentage of IgG$^+$ cells compared with infection alone (fig. S4, E and F). Notably, the variants and corresponding mutations tested in our panel had different magnitudes of effect (Fig. 3, E and F, and fig. S4, C and D). B.1.1.7 RBD with a single N501Y mutation had relatively little change in binding compared with WT RBD. Consistent with the in vitro pseudovirus neutralization data above, B.1.351 RBD resulted in a more-substantial loss of cross-binding, whereas B.1.617.2 RBD had an intermediate effect on binding.

Cross-sectional analysis of variant-binding at the 6-month time point also revealed two major findings. First, all vaccinated individuals in our study maintained variant-specific memory B cells for at least 6 months, with an average of >50% of WT RBD$^+$ memory B cells also cross-binding all three major VOCs (Fig. 3, G and H). Second, mRNA vaccination in SARS-CoV-2–naïve individuals induced a stronger response to B.1.351 than infection alone (Fig. 3H). One possible explanation for this difference is the immunogen itself. Vaccinated individuals mount a primary response to the mRNA-encoded prefusion stabilized spike

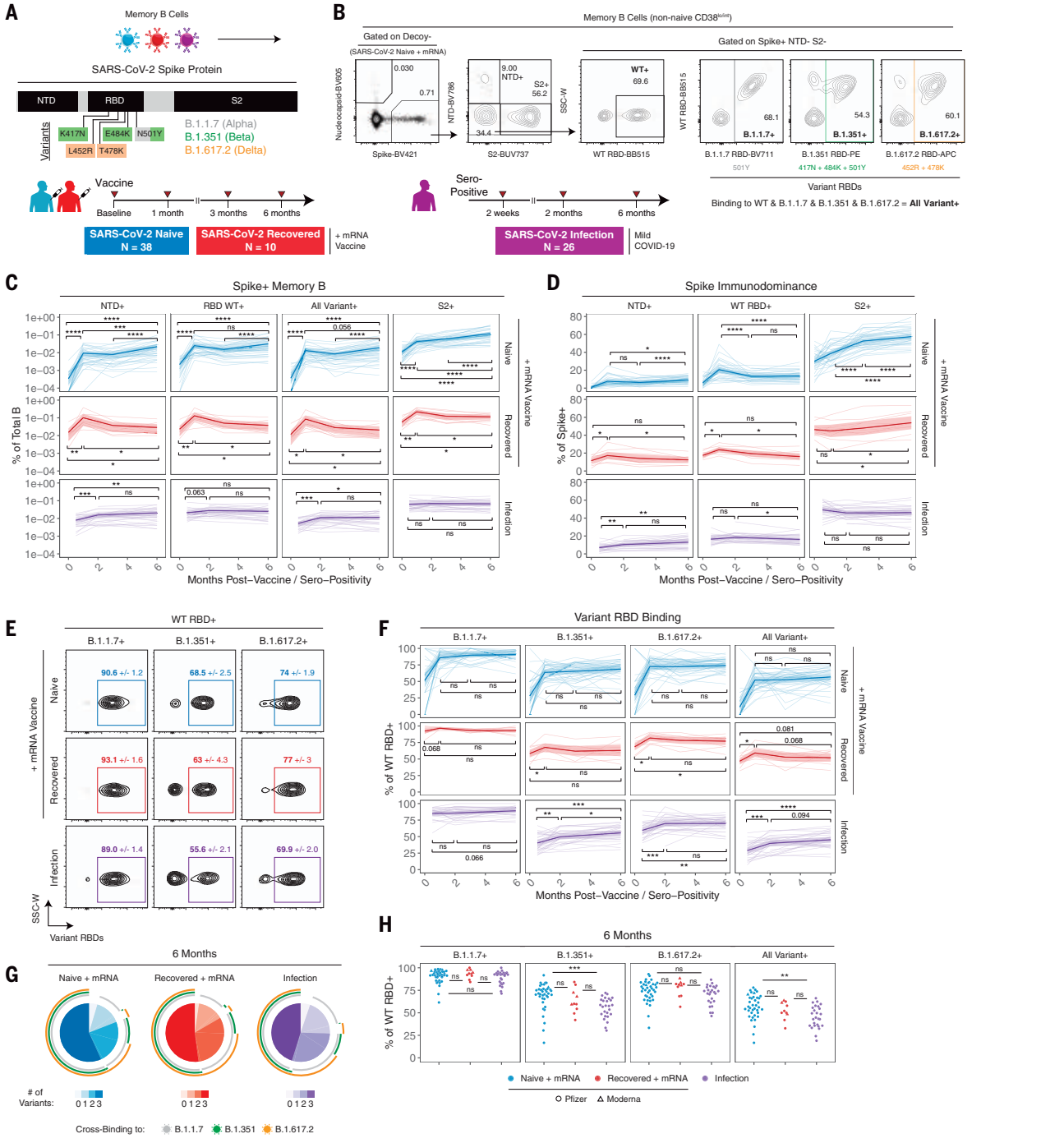


Fig. 3. Memory B cells induced by mRNA vaccination or infection are cross-reactive to SARS-CoV-2 VOCs and increase in frequency over time. (**A** and **B**) Experimental design (A) and gating strategy (B) for quantifying the frequency and phenotype of spike subunit and variant-specific memory B cells by flow cytometry. Specific mutations in B.1.1.7, B.1.351, or B.1.617.2 variant RBDs are indicated. (**C**) Frequencies of spike$^+$ NTD$^+$, spike$^+$ WT RBD$^+$, spike$^+$ all variant$^+$ (all variant RBD binding), and spike$^+$ S2$^+$ memory B cells over time in PBMC samples from vaccinated or convalescent individuals. Data are represented as a percentage of total B cells. (**D**) Percent NTD$^+$, RBD$^+$, or S2$^+$ of total spike$^+$ memory B cells over time. (**E**) Representative plots of variant RBD cross-binding gated on spike$^+$ WT RBD$^+$ cells in vaccinated or convalescent individuals. Mean and standard error values at the 6-month time point are indicated. (**F**) Percent B.1.1.7$^+$, B.1.351$^+$, B.1.617.2$^+$, or all variant$^+$ of WT RBD$^+$ memory B cells over time. (**G**) Boolean analysis of variant cross-binding memory B cell populations in vaccinated, infected then vaccinated, or infected-only individuals at 6 months after vaccination or seropositivity. Pie charts indicate the fraction of WT RBD$^+$ memory B cells that cross-bind zero, one, two, or three variant RBDs. Colored arcs indicate cross-binding to specific variants. (**H**) Cross-sectional analysis of variant binding as a percentage of WT RBD$^+$ memory B cells at 6 months after vaccination or seropositivity. For (C), (D), and (F), thick lines indicate mean values, and thin lines represent individual subjects. Statistics were calculated using paired [(C), (D), and (F)] or unpaired (H) nonparametric Wilcoxon test with BH correction for multiple comparisons. Blue, red, and purple values indicate comparisons within naïve, recovered, or infection-only groups, respectively. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; ns, not significant.

trimer, potentially allowing for increased recruitment and/or selection of specific clones that can bind conserved regions of RBD (*43*, *44*). By contrast, convalescent individuals were primed against native, nonstabilized spike protein. Taken together, our data indicate robust B cell memory to multiple components of the spike protein as well as currently described VOCs that continues to evolve and increase in frequency over time.

Clonal evolution of variant-specific memory B cells

We next asked what differences may underly variant-binding versus nonbinding properties of memory B cells. Here, we focused on the Beta B.1.351 variant RBD containing the K417N, E484K, and N501Y mutations because this variant resulted in the greatest loss of binding relative to WT RBD (Fig. 3, E, G, and H). We designed a sorting panel to identify three populations of memory B cells with different antigen-binding specificities: (i) memory B cells that bind full-length spike but not RBD, (ii) memory B cells that bind full-length spike and WT RBD but not B.1.351 variant RBD, and (iii) memory B cells that bind full-length spike and cross-bind both WT and B.1.351 variant RBD (Fig. 4A and fig. S5A). Naïve B cells were also sorted as a control. These populations were isolated from eight SARS-CoV-2–naïve

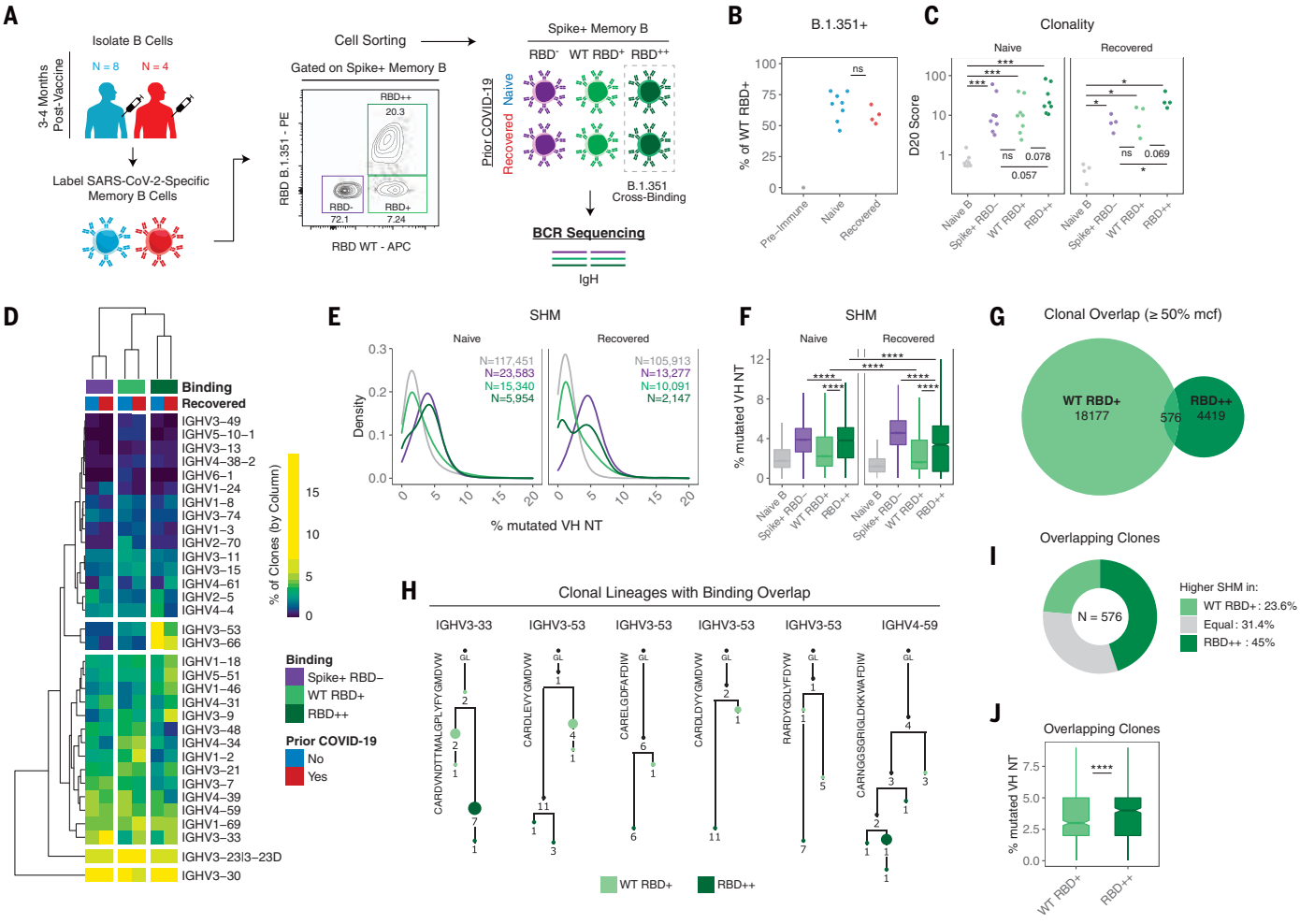


Fig. 4. Variant-binding memory B cell clones use distinct VH genes and evolve through somatic hypermutation. (**A**) Experimental design for sorting and sequencing SARS-CoV-2–specific memory B cells. (**B**) Frequency of RBD^{++} (B.1.351 variant cross-binding) memory B cells as a percentage of total RBD^{+} cells. (**C**) Percentage of sequence copies occupied by the top 20 ranked clones (D20) across naïve B cells and different antigen-binding memory B cell populations. (**D**) Heatmap and hierarchical clustering of VH gene usage frequencies in memory B cell clones across different antigen-binding populations. Data are represented as the percentage of clones with the indicated VH gene per column. (**E** and **F**) Somatic hypermutation (SHM) density plots (E) and boxplots of individual clones across naïve B cells and different antigen-binding memory B cell populations (F). Data are represented as the percent of mutated VH nucleotides. Number of clones sampled for each population is indicated. For (C) to (F), data were filtered on clones with productive rearrangements and ≥2 copies. (**G**) Venn diagram of clonal lineages that are shared between WT RBD and RBD cross-binding (RBD^{++}) populations. Data were filtered on the basis of larger clones with ≥50% mean copy number frequency (mcf) in each sequencing library. (**H**) Example lineage trees of clones with overlapping binding to WT and B.1.351 variant RBD. VH genes and CDR3 sequences are indicated. Numbers refer to mutations compared with the preceding vertical node. Colors indicate binding specificity, black dots indicate inferred nodes, and size is proportional to sequence copy number. GL, germline sequence. (**I**) Classification of SHM within overlapping clones. Each clone was defined as having higher (or equal) SHM in WT RBD binders or RBD^{++} cross-binders on the basis of average levels of SHM for all WT RBD versus RBD^{++} sequence variant copies within each lineage. (**J**) SHM levels within overlapping clones. Data are represented as the percentage of mutated VH nucleotides for WT RBD and RBD^{++} sequence copies. Statistics were calculated using unpaired nonparametric Wilcoxon test, with BH correction for multiple comparisons in (C) and (F). Notches on boxplots in (F) and (J) indicate a 95% confidence interval of the median. *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001; ns, not significant.

and four SARS-CoV-2–recovered individuals at 3 to 4 months postvaccination (Fig. 4A and fig. S5A). Consistent with our previous data, between 50 and 80% of WT RBD$^+$ cells cobound B.1.351 variant RBD (Fig. 4B), which indicates that a majority of RBD epitopes in the response are shared by the WT and mutant RBDs.

To gain insight into the clonal composition of the different spike and/or RBD-binding B cell populations, IgH rearrangements were amplified from the sorted populations (N = 48 total), and related sequences were grouped into clones (table S2). We analyzed the contribution of the top copy number clones to the overall repertoire as measured by the diversity 20 (D20) index (i.e., the percent of the overall response composed of the top 20 clones). The D20 index ranged from <1% for naïve B cells (which is expected for a diverse, nonclonally expanded population) to >20% for some of the antigen-binding populations (Fig. 4C). Clones that cross-bound both WT and B.1.351 RBD trended toward higher D20 scores, suggesting greater clonal expansion and/or lower diversity compared with the other antigen-binding populations (Fig. 4C). The clonality of antigen-binding memory B cell populations was not significantly different after vaccination on the basis of prior immunity, although there was heterogeneity in clonal expansion across individuals.

We further analyzed immunoglobulin heavy-chain variable region (IGHV) gene usage across the different antigen-binding memory B cell populations. Hierarchical clustering revealed that VH gene profiles were overall similar in vaccinated individuals regardless of prior SARS-CoV-2 infection status (Fig. 4D and fig. S5B), indicating that both vaccination and infection followed by vaccination can recruit similar clones into the response. Rather, IGHV gene usage largely clustered on the basis of the antigen specificity, with increased usage of VH3-53 and VH3-66 in RBD cross-binding clones (Fig. 4D and fig. S5B). Notably, both of these IGHV genes are known to be enriched in RBD-binding B cells (*45*, *46*). These differences in IGHV gene usage between WT only and variant cross-binding phenotype suggested that these cells may derive, at least partially, from different B cell clones that were independently recruited into the vaccine response.

Analysis of VH gene sequences also revealed clear differences in somatic hypermutation (SHM) between the different antigen-binding populations. As expected, SARS-CoV-2–specific memory B cell clones had significantly more VH nucleotide mutations compared with naïve B cell clones (Fig. 4, E and F, and fig. S5C). Spike$^+$, RBD nonbinding memory B cells (which include NTD- and S2-binding populations) had high SHM (Fig. 4, E and F, and fig. S5C), consistent with germinal center–dependent responses as well as possible recall responses of preexisting S2 cross-reactive clones. Notably, significantly higher levels of SHM were observed in variant RBD cross-binding clones compared with WT RBD only clones (Fig. 4, E and F, and fig. S5C). Additionally, boosting of infection-acquired immunity by mRNA vaccination in SARS-CoV-2–recovered donors did not produce higher SHM in RBD-binding memory B cell clones compared with vaccination alone (Fig. 4F).

To determine whether variant cross-binding clones could evolve from WT RBD-binding clones, we next investigated whether there was any clonal overlap between these populations. For clonal overlap analysis, we focused on larger clones (defined as having copy numbers at or above 50% of the mean copy number frequency within each sequencing library) (*47*) because larger clones are more readily sampled at both the clonal and subclonal levels. Among such larger clones, 2.5% had sequence variants that were isolated from both WT RBD and cross-binding populations (Fig. 4G and fig. S5D). Lineage analysis revealed that WT and cross-binding sequence variants localized on separate branches (representative lineages shown in Fig. 4H), indicating that the shift in antigen-reactivity was not a result of contamination of the sorted populations (in which case sequence variants localize to the same nodes). Next, to determine whether cross-binding activity arose from WT binding or vice versa, we used SHM as a molecular clock and counted the fraction of overlapping clonal lineages in which variant binding had higher, lower, or equivalent levels of SHM to WT RBD-binding variants. Consistent with the overall SHM data, this analysis of overlapping clones revealed higher levels of SHM in the variant binding sequences compared with WT only binding sequences (Fig. 4, I and J), suggesting a clonal evolution from WT only binding to variant RBD cobinding for at least some clones.

Taken together, these data indicate that mRNA vaccine–induced memory B cells that bind variant RBDs have higher SHM compared with clones that only bind WT RBD. Moreover, the clonal relationships between WT-only and cross-binding RBD-specific memory B cells suggest that variant binding capacity can evolve from clones that initially bound to WT RBD. Ongoing evolution and selection of these clones could therefore facilitate cross-protection against different VOCs. These findings are consistent with earlier work that has suggested that SHM and affinity maturation are important for the acquisition of broader neutralization activity of RBD-binding antibodies that are formed in response to SARS-CoV-2 infection (*48*, *49*). It is presently unclear how additional antigen exposure through booster vaccination, environmental virus exposure, or overt infection may affect additional affinity maturation toward improved variant binding.

Memory CD4$^+$ and CD8$^+$ T cell responses to SARS-CoV-2 mRNA vaccines

In addition to antibodies and memory B cells, memory T cells can contribute to protection upon reexposure to virus. Memory T cell responses have also been shown to be less affected by VOCs than humoral immune responses (*21*, *50*). To determine whether mRNA vaccination induced durable antigen-specific memory T cell responses, we performed a flow cytometric analysis using an activation-induced marker (AIM) assay. PBMCs were stimulated with peptide megapools containing optimized spike epitopes (*51*, *52*). Antigen-specific responses were quantified as the frequency of AIM$^+$ non-naïve T cells in stimulated samples with background subtraction from paired unstimulated controls (Fig. 5, A and B) (*19*). Full gating strategies are provided in fig. S6. Antigen-specific CD4$^+$ T cells were defined on the basis of coexpression of CD40L and CD200. Antigen-specific CD8$^+$ T cells were defined on the basis of expression of four of five total activation markers, as described previously (*19*).

Consistent with recent studies, SARS-CoV-2 mRNA vaccination efficiently primed antigen-specific CD4$^+$ T cells and CD8$^+$ T cells (Fig. 5, C and D) (*20*–*22*). All individuals in our cohort, regardless of prior infection with SARS-CoV-2, had detectable CD4$^+$ T cell responses above their individual baseline 1 week after the second vaccine dose (Fig. 5C). Most (36 of 41) SARS-CoV-2–naïve individuals also generated detectable CD8$^+$ T cell responses after the second dose (Fig. 5D). By contrast, vaccination did little to further boost prevaccination antigen-specific CD8$^+$ T cell frequencies in SARS-CoV-2–recovered individuals (Fig. 5D). A marked contraction phase was observed from peak responses to 3 months postvaccination, with a half-life of 47 days for CD4$^+$ T cells and 27 days for CD8$^+$ T cells (Fig. 5, C and D). These kinetics are consistent with a typical T cell response after the effector phase (*53*). After this initial contraction, antigen-specific memory CD4$^+$ T cell frequencies stabilized from 3 to 6 months postvaccination with a half-life of 187 days, whereas CD8$^+$ T cells continued to decline. Overall, 28 of 31 SARS-CoV-2–naïve individuals had vaccine-induced antigen-specific CD4$^+$ T cell responses at 6 months postvaccination above prevaccination baseline levels, and 13 of 31 had detectable CD8$^+$ T cell responses above baseline (Fig. 5, C and D). In SARS-CoV-2–recovered subjects, mRNA vaccination had only a modest effect on T cell responses and did not elevate the magnitude of long-term antigen-specific CD4$^+$ or CD8$^+$ T cell memory above baseline levels (Fig. 5, C and D). Taken together, these data indicate that mRNA vaccination generates durable SARS-CoV-2–specific CD4$^+$ T cell memory in individuals who were not previously infected with SARS-CoV-2 and only transiently boosts

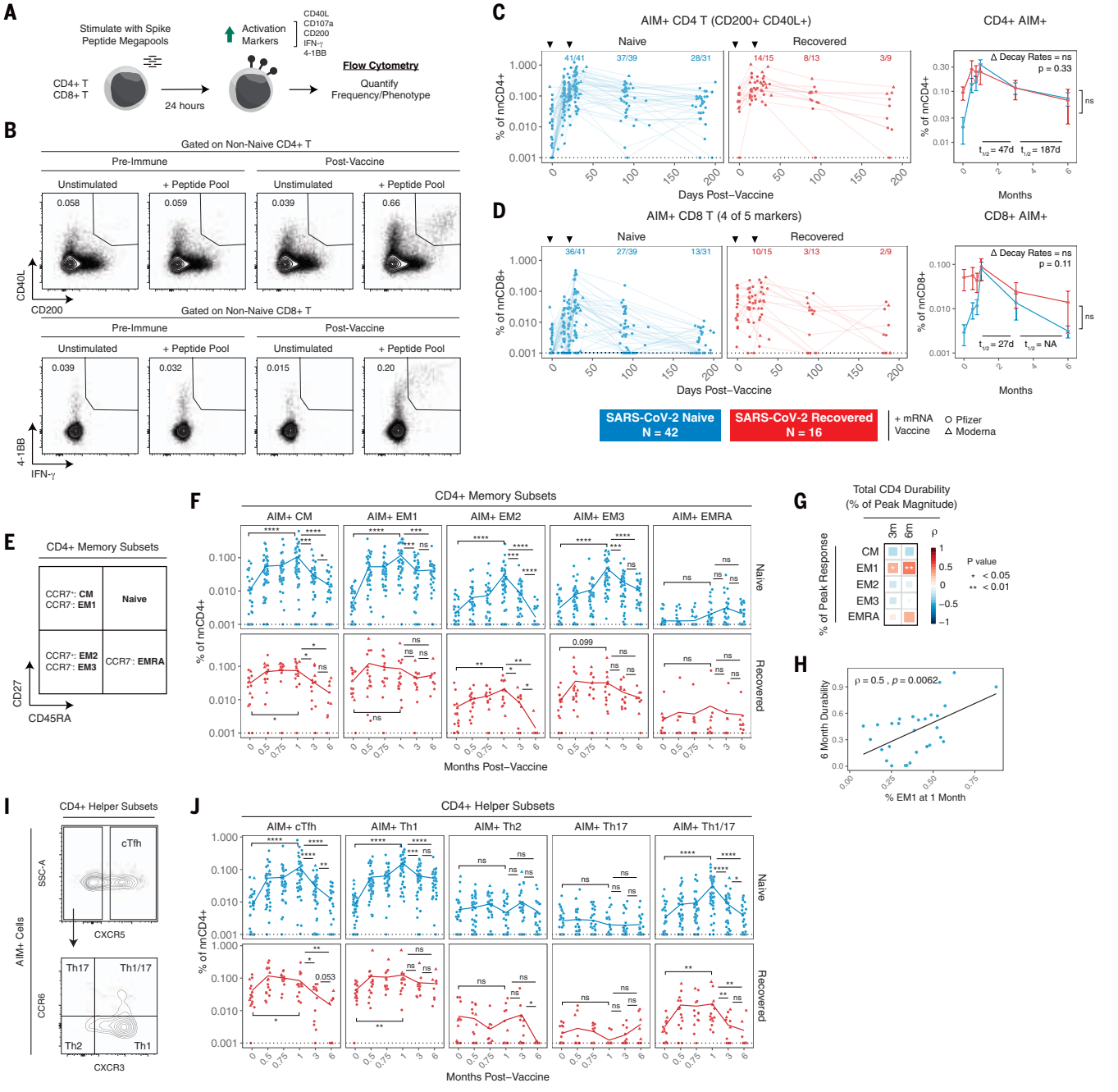


Fig. 5. SARS-CoV-2 mRNA vaccines generate durable memory T cell responses. (**A** and **B**) Experimental design (A) and gating strategy (B) for quantifying the frequency of SARS-CoV-2–specific $CD4^+$ and $CD8^+$ T cells by AIM assay. For $CD4^+$ T cells, antigen specificity was defined on the basis of coexpression of CD40L and CD200. For $CD8^+$ T cells, antigen specificity was defined on the basis of expression of at least four of five activation markers, as indicated in (A). (**C** and **D**) Frequencies of AIM^+ $CD4^+$ T cells (C) and AIM^+ $CD8^+$ T cells (D) over time in PBMC samples from vaccinated individuals. Data were background subtracted using a paired unstimulated control for each time point and are represented as a percentage of non-naïve $CD4^+$ or $CD8^+$ T cells. Black triangles indicate time of vaccine doses, fractions above plots indicate the number of individuals above their individual baseline at memory time points, and summary plots show mean values with the 95% confidence interval. Decay rates were calculated using a piecewise linear mixed-effects model with censoring. Δ Decay Rates indicates whether decay rates were different in SARS-CoV-2–naïve and –recovered groups. (**E**) AIM^+ $CD4^+$ T cell memory subsets were identified on the basis of surface expression of CD45RA, CD27, and CCR7. (**F**) Frequencies of AIM^+ $CD4^+$ T cell memory subsets over time. (**G**) Correlation matrix of memory subset skewing at peak (1-month) response with total AIM^+ $CD4^+$ T cell durability at 3 and 6 months. Durability was measured as the percentage of peak response maintained at memory time points for each individual. (**H**) Correlation between percent of EM1 cells at peak response and 6-month durability. (**I**) AIM^+ $CD4^+$ T helper subsets were defined on the basis of chemokine receptor expression. (**J**) Frequencies of AIM^+ $CD4^+$ T helper subsets over time. For (F) and (J), lines connect mean values at different time points. Dotted lines indicate the limit of detection for the assay. Statistics were calculated using unpaired nonparametric Wilcoxon test with BH correction for multiple comparisons. Correlations were calculated using nonparametric Spearman rank correlation. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns, not significant.

these responses in SARS-CoV-2–recovered individuals.

Antigen-specific T cells can further be classified into different memory subsets using cell surface markers (Fig. 5E). Peak $CD4^+$ T cell responses after SARS-CoV-2 mRNA vaccination were composed of predominantly central memory [(CM); $CD45RA^-$ $CD27^+$ $CCR7^+$] and effector memory 1 [(EM1); $CD45RA^-$ $CD27^+$ $CCR7^-$] cells in both SARS-CoV-2–naïve and –recovered individuals (Fig. 5F) (*19*). During contraction from peak responses, antigen-specific $CCR7^+$ CM cells were largely lost from circulation, whereas antigen-specific CCR7- EM1 cells stabilized in frequency from 3 to 6 months postvaccination. Moreover, the percentage of the peak $CD4^+$ response that was EM1 cells, but not other memory subsets, was significantly associated with the durability of the overall $CD4^+$ T cell response at 3 and 6 months postvaccination (Fig. 5, G and H), which suggests that EM1s are long-lived memory $CD4^+$ T cells and that early skewing toward an EM1 phenotype contributes to durable $CD4^+$ T cell memory. Although our AIM assay allows for the detection of low-frequency memory $CD8^+$ T cell responses for overall quantification, reliable subsetting of antigen-specific $CD8^+$ T cells at memory time points was not feasible because of the low number of events.

mRNA vaccination also preferentially induced antigen-specific $CD4^+$ circulating T follicular helper cells (cT_{FH} cells) and T helper cells (T_H1 cells) in both SARS-CoV-2–naïve and –recovered individuals, whereas T_H2, T_H17, and $T_H1/17$ cells were detected at lower levels in the AIM assay (Fig. 5I). Although the overall frequency of antigen-specific $CD4^+$ T cells stabilized from 3 to 6 months postvaccination, cT_{FH} and T_H1 cells had distinct trajectories. Specifically, cT_{FH} cells declined more rapidly than T_H1 cells both during the initial contraction phase and from 3 to 6 months postvaccination (Fig. 5J), perhaps reflecting redistribution of T_{FH} cells into lymphoid tissues. By contrast, spike-specific T_H1 cells did not decline in the blood from 3 to 6 months postvaccination. Although cT_{FH} cells may be important in the early stages of vaccine response, these data indicate that the durable component of the memory $CD4^+$ T cell response at 6 months postvaccination is largely composed of T_H1 cells, and the boosting of preexisting immunity with mRNA vaccine does not change the magnitude or subset composition of the $CD4^+$ memory T cell response.

Integrated analysis of immune components and vaccine-induced memory to SARS-CoV-2

A goal of this study was to assess the development of multiple components of antigen-specific immune memory over time in the same individuals after SARS-CoV-2 mRNA vaccination. This dataset allowed us to integrate longitudinal antibody, memory B cell, and memory T cell responses to construct an immunological landscape of SARS-CoV-2 mRNA vaccination. To this end, we applied uniform manifold approximation and projection (UMAP) to visualize the trajectory of vaccine-induced adaptive immunity over time. This analysis revealed a continued evolution of the overall immune response in SARS-CoV-2–naïve subjects after mRNA vaccination with different time points occupying largely non-overlapping UMAP space (Fig. 6A). Projection of individual immune components onto the UMAP space revealed that primary vaccination was largely defined by rapid induction of $CD4^+$ T cell immunity (Fig. 6B). The second vaccine dose induced peak antibody, $CD4^+$ T cell, and $CD8^+$ T cell responses. Antibodies and $CD4^+$ T cells then remained durable through later memory time points, coinciding with a trajectory shift toward peak memory B cell responses. Notably, all 6-month samples clustered away from preimmune baseline samples (Fig. 6A), highlighting the durable multicomponent immune memory induced by mRNA vaccination. At 6 months, we observed some heterogeneity in the immune landscape. This heterogeneity may be partially driven by a significant negative correlation between age and anti-spike IgG (fig. S7, A and B). Sex did not appear to have any association with the overall antigen-specific response to mRNA vaccination (fig. S7C). SARS-CoV-2–recovered individuals occupied a wide range of UMAP space at baseline, highlighting the variability of infection-induced virus-specific immunity (Fig. 6A). Time since infection did not appear to fully explain the observed variability for SARS-CoV-2–recovered individuals at prevaccine baseline (fig. S7D). Vaccination uniformly shifted SARS-CoV-2–recovered individuals at 3 months postvaccination to a region defined by high levels of all antigen-specific immune parameters analyzed (Fig. 6A). This region was largely unoccupied by SARS-CoV-2–naïve vaccinees, underscoring the potency of reactivating preexisting immune responses. These distinctively high responses were transient, however, as SARS-CoV-2–recovered individuals at 6 months postvaccination shifted toward the UMAP space occupied by memory time points in SARS-CoV-2–naïve individuals at 3 and 6 months postvaccination.

A second question is how different antigen-specific mRNA vaccine–induced immune components interact with each other over time. Antibody responses after the first or second vaccine dose did not correlate with the magnitude of B cell memory at 6 months (Fig. 6C). However, at 3 and 6 months postvaccination, antibodies were significantly associated with contemporaneous memory B cell responses, an effect most prominent for B.1.351 neutralizing titers (Fig. 6C). Given the role of T_{FH} cells in generating efficient humoral immunity, we next investigated the relationship between antigen-specific T cells and humoral responses. $CD4^+$ T cell responses, especially cT_{FH} responses, as early as 2 weeks after the first dose of mRNA vaccine were positively correlated with antibody responses up to and including 6 months postvaccination (Fig. 6D and fig. S7E). This observation suggested that rapid mobilization of $CD4^+$ T cell responses by the first mRNA vaccine dose had a lasting effect on humoral immunity. Like memory B cells, the magnitude of $CD4^+$ T cell responses at 6 months was also correlated with antibodies at 6 months (Fig. 6D), suggesting that antibody levels may provide a useful (though incomplete) proxy for the magnitude of memory B and $CD4^+$ T cell responses at 6 months postvaccination. Taken together, these data identify key temporal relationships between different branches of the human immune response that are associated with long-term immune memory after mRNA vaccination.

Next, we investigated whether the magnitude of peak responses after the second vaccine dose in SARS-CoV-2–naïve subjects was predictive of memory responses at 3 and 6 months. Peak antibody levels were significantly correlated with later antibody levels (fig. S7F). Memory B cell frequencies 1 week after the second dose were also correlated significantly with 3- and 6-month frequencies (fig. S7F). Like antibodies and memory B cells, peak T cell responses after the second dose were predictive of later time points (fig. S7F). Overall, these data suggest that the magnitude and trajectory of individual components of the immune response are patterned soon after the second vaccine dose in SARS-CoV-2–naïve individuals.

This dataset also presented an opportunity to investigate the effect of mRNA vaccination in subjects with preexisting immunity, in this case from a prior SARS-CoV-2 infection. To investigate the dynamics of these recall responses, we examined the change in individual SARS-CoV-2–specific immune responses from prevaccine baseline levels. Vaccination modestly increased preexisting memory B cell and $CD4^+$ T cell frequencies at 1 month, with a more robust increase in antibody levels (Fig. 6E). To investigate the contribution of preexisting immune memory to these recall antibody responses, we correlated the magnitude of prevaccine memory responses with the change in antibody levels after vaccination. The frequency of SARS-CoV-2–specific memory B cells was the only feature of preexisting immunity that correlated significantly with antibody responses after vaccination (Fig. 6F), consistent with a major role for memory B cells

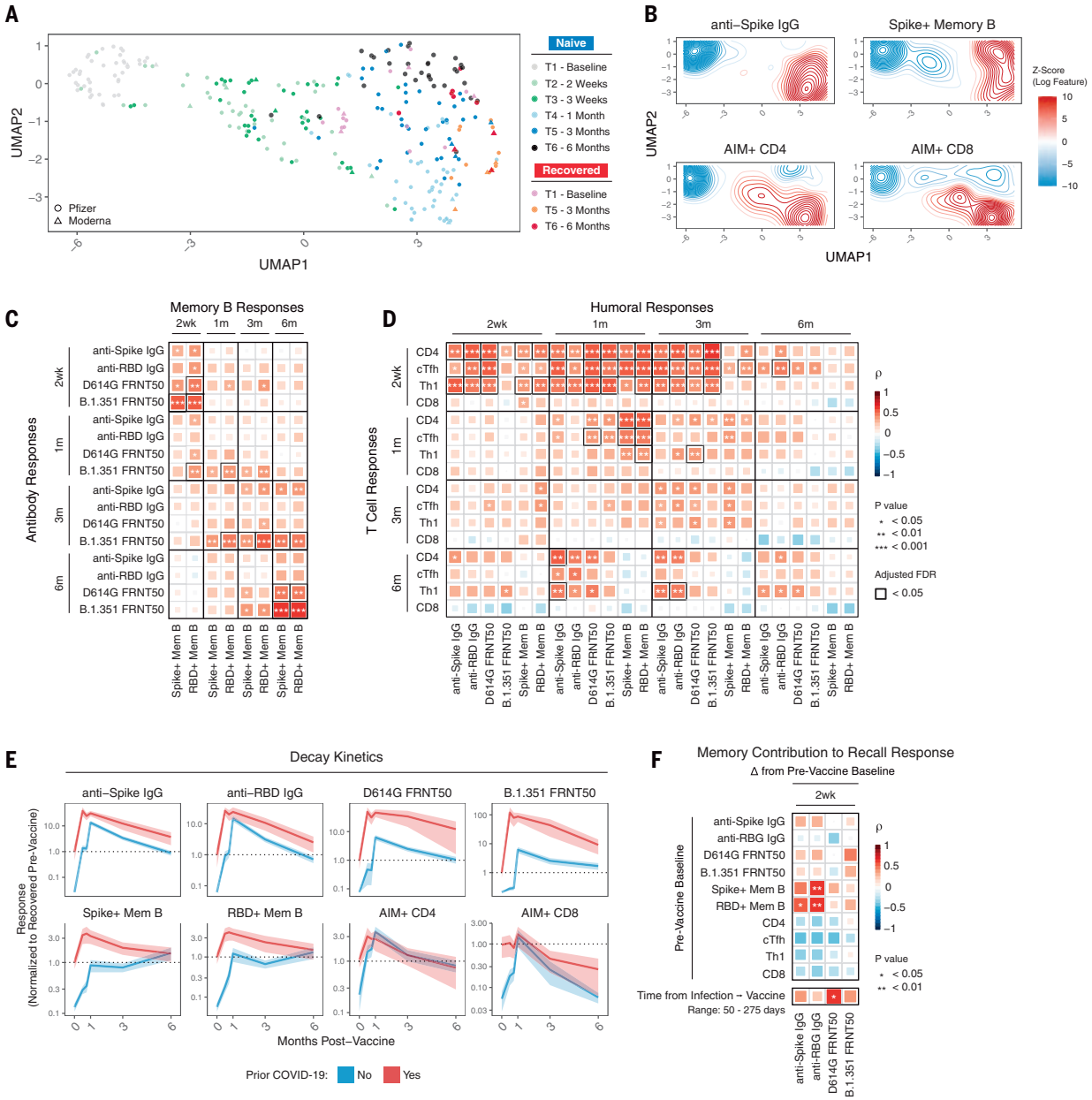


Fig. 6. Immune trajectories and relationships in response to SARS-CoV-2 mRNA vaccination. (**A**) UMAP of 12 antigen-specific parameters of antibody, memory B, and memory T cell responses to mRNA vaccination in SARS-CoV-2–naïve and –recovered subjects. Data points represent individual participants and are colored by time point relative to primary vaccine. (**B**) Kernel density plots of anti-spike IgG, spike^{+} memory B, AIM CD4^{+}, and AIM^{+} CD8^{+} T cells. Red contours represent areas of UMAP space that are enriched for specific immune components. (**C**) Correlation matrix of antibody and memory B cell responses over time in SARS-CoV-2–naïve subjects. (**D**) Correlation matrix of T cell and humoral responses over time in SARS-CoV-2–naïve subjects. (**E**) Decay kinetics of antibody, memory B cell, and memory T cell parameters over time in SARS-CoV-2–naïve and –recovered vaccinees. Data are normalized to prevaccine levels in SARS-CoV-2–recovered individuals to evaluate the effect of boosting preexisting immunity. Lines connect mean values at different time points, ribbons represent the 95% confidence interval of the mean, and dotted lines indicate mean values at baseline. (**F**) Correlation matrix of baseline memory components and time since infection with antibody recall responses after vaccination in SARS-CoV-2–recovered individuals. Recall responses were calculated as the difference between postvaccination levels and prevaccine baseline. All statistics were calculated using nonparametric Spearman rank correlation.

in recall responses. Because we observed that memory B cell frequencies continue to increase in the months postvaccination, we investigated whether time since infection affected the magnitude of the antibody recall response. A longer interval between infection and vaccination was correlated with a significantly greater neutralizing antibody recall response to D614G, with similar trends for B.1.351 neutralization and for binding antibodies to spike and RBD (Fig. 6F). Thus, these data suggest that there may be some benefit to a longer interval between initial priming and subsequent restimulation or boost of immune responses to SARS-CoV-2.

Finally, we evaluated the decay kinetics of SARS-CoV-2–specific recall responses. Boosting of spike- and RBD-specific memory B cell and memory $CD4^+$ T cell responses was transient and returned to prevaccination baseline by 3 to 6 months (Fig. 6E). $CD8^+$ T cell responses were not boosted in SARS-CoV-2–immune subjects and decayed from peak at a comparable rate to that in SARS-CoV-2–naïve vaccinees (Fig. 6E). The increase in anti-spike and anti-RBD binding antibodies was also transient and returned to near baseline by 6 months postvaccination (Fig. 6E). D614G and B.1.351 neutralizing antibody remained substantially above prevaccine baseline levels (~10-fold increase at 6 months), but these antibody levels were also declining over time. Notably, the decay rate of antibodies was similar between SARS-CoV-2–naïve and SARS-CoV-2–recovered vaccinees (Fig. 6E). Lastly, we estimated the benefit of mRNA vaccine–mediated boosting of preexisting immunity in this setting by calculating, on the basis of antibody half-lives, the time it would take for recall responses to return to prevaccine antibody levels. We estimated from these calculations that recall responses to mRNA vaccination will maintain antibodies above prevaccination levels in this cohort of mostly young individuals who recovered from mild COVID-19 for ~7 to 16 months. Additionally, recall responses in this cohort remained above peak responses in SARS-CoV-2–naïve vaccinees, where clinical efficacy is well established, for 2 to 3 months for spike-binding antibodies and 6 to 10 months for neutralizing titers (table S3). Overall, these data suggest that boosting of infection-induced immunity with mRNA vaccination does not substantially enhance already durable memory B cell or memory T cell responses. Rather, the benefit of vaccination in the context of preexisting immunity may be limited to a significant but transient increase in antibodies, with some of this benefit to antibody levels remaining at 6 months.

Concluding remarks

These studies provide insight into the evolution of immunological memory after SARS-CoV-2 mRNA vaccination. Specifically, the continued increase in SARS-CoV-2–specific memory B cells between 3 and 6 months after mRNA vaccination, even as antibody levels declined in the same individuals, suggests that prolonged germinal center reactions (*14*) continue to generate circulating memory B cells for at least several months after vaccination. A majority of these memory B cells were able to cross-bind VOCs, including B.1.1.7 (Alpha), B.1.351 (Beta), and B.1.617.2 (Delta), and clonal relationships indicated that at least some of these cross-binding memory B cells evolved through somatic hypermutation from clones that initially lacked variant binding. This evolution of variant binding may have implications for booster strategies aimed at targeting antibody responses to future variants. As demonstrated here, these memory B cells are capable of mounting rapid recall responses, providing a new source of antibodies upon infection or booster vaccination. Furthermore, there may be differences in immunity generated by mRNA vaccination versus infection, as memory B cells 6 months postvaccination were qualitatively superior at binding VOCs compared with memory B cells 6 months after recovering from mild COVID-19. Variant binding developed rapidly after two-dose mRNA vaccination but evolved more slowly after infection, consistent with conclusions drawn from other approaches (*17*). In addition to durable B cell memory, SARS-CoV-2–specific memory $CD4^+$ T cells were relatively stable from 3 to 6 months after mRNA vaccination, and the vast majority of vaccinees maintained robust $CD4^+$ T cell responses at 6 months. Early $CD4^+$ T cell responses correlated with 3- and 6-month humoral responses, highlighting a role for T cell immunity in shaping the overall response to vaccination. Together, these data identify durable cellular immunity for at least 6 months after mRNA vaccination, with persistence of high-quality memory B cells and strong $CD4^+$ T cell memory in most individuals.

These data may also provide context for understanding potential discrepancies in vaccine efficacy at preventing infection versus severe disease, hospitalization, and death (*10*, *11*). Declining antibody titers over time likely reduce the potential that vaccination will completely prevent infection or provide near-sterilizing immunity. However, the durability of cellular immunity, here demonstrated for at least 6 months, may contribute to rapid recall responses that can limit initial viral replication and dissemination in the host, thereby preventing severe disease. Finally, by examining individuals with preexisting immunity after infection, we were able to gain insights into the possible effects of booster vaccination. In this setting, boosting of preexisting immunity from prior infection with mRNA vaccination mainly resulted in a transient benefit to antibody titers with little-to-no long-term increase in cellular immune memory. Antibody decay rates were similar in SARS-CoV-2–naïve and –recovered vaccinees, which suggests that additional vaccine doses will temporarily prolong antibody-mediated protection without fundamentally altering the underlying landscape of SARS-CoV-2 immune memory. It will be important to examine whether similar dynamics exist after other types of immune boosting, including a third dose of mRNA vaccine in previously vaccinated individuals or SARS-CoV-2 infections that occur after vaccination. Nevertheless, these data provide evidence for durable immune memory at 6 months after mRNA vaccination and are relevant for interpreting epidemiological data on rates of infections in vaccinated populations and the implementation of booster vaccine strategies.

Despite the overall strengths of this study, including the large sample size and integrated measurement of multiple components of the antigen-specific adaptive immune response, there are several limitations. First, the overall number of subjects, although substantial for studies with high depth of immune profiling, was still limited compared with epidemiological or phase 3 clinical trials. In particular, only 9 to 10 individuals with preexisting immunity from SARS-CoV-2 infection were fully sampled through 6 months postvaccination. Second, it is possible that the time points in this study do not perfectly capture the full kinetics of the response for each individual immune component. For example, it is possible that antibody levels could stabilize at time points beyond 6 months rather than continuing to decay at the observed rates. Additionally, the comparison of variant-specific immune memory induced by vaccination versus infection is limited to mild COVID-19 cases and does not include more-severe disease. Time points for sampling of infection only, although broadly consistent with the vaccination studies, were also not perfectly aligned with the date of actual infection because samples were longitudinally collected after a positive serology test rather than an acutely positive polymerase chain reaction (PCR) test in most cases. Regarding $CD8^+$ T cell responses, our AIM assay was effective at capturing peak responses after vaccination; however, this assay may not be sensitive enough to detect very low-frequency $CD8^+$ T cells at memory time points. Other approaches, such as major histocompatibility complex (MHC) tetramers, will be necessary in the future to further interrogate memory $CD8^+$ T cell responses after vaccination. Finally, our cohort is skewed toward young, healthy individuals. As such, the results described may not fully represent the durability of vaccine-induced immunity in older individuals or in populations with chronic diseases and/or compromised immune systems, and future studies will be required to better quantify the immune response over time in these populations.

Materials and methods

Clinical recruitment and sample collection

Sixty-one individuals (45 SARS-CoV-2–naïve; 16 SARS-CoV-2–recovered) were consented and enrolled in the longitudinal vaccine study with approval from the University of Pennsylvania Institutional Review Board (IRB no. 844642). All participants were otherwise healthy and, based on self-reported health screening, did not have any history of chronic health conditions.

Subjects were stratified on the basis of self-reporting and laboratory evidence of a prior SARS-CoV-2 infection. All subjects received either Pfizer (BNT162b2) or Moderna (mRNA-1273) mRNA vaccines. Samples were collected at six time points: baseline, ~2 weeks after primary immunization, day of secondary immunization, ~1 week after secondary immunization, ~3 months after primary immunization, and ~6 months after primary immunization. Eighty to 100 mL of peripheral blood samples and clinical questionnaire data were collected at each study visit. A separate cohort of 26 SARS-CoV-2–convalescent individuals was used to compare vaccine-induced immune responses with immune responses upon SARS-CoV-2 infection. This cohort was a subset from a sero-monitoring study previously described (*40*) that was approved by the University of Pennsylvania Institutional Review Board (IRB no. 842847). Recent or active SARS-CoV-2 infections were identified on the basis of SARS-CoV-2 RBD antibody levels and/or SARS-COV-2 PCR testing. Longitudinal samples were collected from seropositive participants up to ~200 days after seroconversion to study long-term immune responses. Full cohort and demographic information is provided in table S1. Additional healthy donor samples were collected with approval from the University of Pennsylvania Institutional Review Board (IRB no. 845061)

Peripheral blood sample processing

Venous blood was collected into sodium heparin and EDTA tubes by standard phlebotomy. Blood tubes were centrifuged at 3000 rpm for 15 min to separate plasma. Heparin and EDTA plasma were stored at −80°C for downstream antibody analysis. Remaining whole blood was diluted 1:1 with R1 [RPMI + 1% fetal bovine serum (FBS) + 2 mM L-glutamine + 100 U penicillin/streptomycin] and layered onto SEPMATE tubes (STEMCELL Technologies) containing lymphoprep gradient (STEMCELL Technologies). SEPMATE tubes were centrifuged at 1200 g for 10 min and the PBMC fraction was collected into new tubes. PBMCs were then washed with R1 and treated with ACK lysis buffer (Thermo Fisher) for 5 min. Samples were washed again with R1, filtered with a 70 μm filter, and counted using a Countess automated cell counter (Thermo Fisher). Aliquots containing 5 to 10 × 10^6 PBMCs were cryopreserved in fresh 90% FBS 10% dimethyl sulfoxide (DMSO).

Detection of SARS-CoV-2 spike- and RBD-specific antibodies

Plasma samples were tested for SARS-CoV-2–specific antibody by ELISA as previously described (*16*, *54*). Plasmids encoding the recombinant full-length spike protein and the RBD were provided by F. Krammer (Mt. Sinai) and purified by nickel-nitrilotriacetic acid resin (Qiagen). ELISA plates (Immulon 4 HBX; Thermo Fisher Scientific) were coated with phosphate-buffered saline (PBS) or 2 ug/mL recombinant protein and stored overnight at 4°C. The next day, plates were washed with PBS containing 0.1% Tween-20 (PBS-T) and blocked for 1 hour with PBS-T supplemented with 3% nonfat milk powder. Samples were heat-inactivated for 1 hour at 56°C and diluted in PBS-T supplemented with 1% nonfat milk powder. After washing the plates with PBS-T, 50 μL diluted sample was added to each well. Plates were incubated for 2 hours and washed with PBS-T. Next, 50 μL of 1:5000 diluted goat anti-human IgG-HRP (Jackson ImmunoResearch Laboratories) or 1:1000 diluted goat anti-human IgM-HRP (Southern-Biotech) was added to each well and plates were incubated for 1 hour. Plates were washed with PBS-T before 50 μL SureBlue 3,3′,5,5′-tetramethylbenzidine substrate (KPL) was added to each well. After 5 min incubation, 25 μL of 250 mM hydrochloric acid was added to each well to stop the reaction. Plates were read with the SpectraMax 190 microplate reader (Molecular Devices) at an optical density (OD) of 450 nm. Monoclonal antibody CR3022 was included on each plate to convert OD values into relative antibody concentrations. Plasmids to express CR3022 were provided by I. Wilson (Scripps).

Detection of SARS-CoV-2 neutralizing antibodies

HEK 293T cells were seeded for 24 hours at 5 × 10^6 cells per 10-cm dish and were transfected using calcium phosphate with 35 μg of pCG1 SARS-CoV-2 S D614G delta18, pCG1 SARS-CoV-2 S B.1.351 delta18, or pCG1 SARS-CoV-2 S B.1.617.2 delta18 expression plasmid encoding a codon optimized SARS-CoV-2 S gene with an 18-residue truncation in the cytoplasmic tail (provided by S. Pohlmann). Mutations in pseudovirus constructs are indicated: D614G (WT) = D614G; B.1.351 = L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V; B.1.617.2 = T19R, G142D, del156-157, R158G, L452R, T478K, D614G, P681R, D950N. Twelve hours after transfection, cells were fed with fresh media containing 1 mM sodium butyrate to increase expression of the transfected DNA. Twenty-four hours after transfection, the SARS-CoV-2 spike-expressing cells were infected for 2 hours with VSV-G pseudotyped VSVΔG-RFP at a multiplicity of infection (MOI) of ~1. Virus-containing media was removed, and the cells were re-fed with media without serum. Media containing the VSVΔG-RFP SARS-CoV-2 pseudotypes was harvested 28 to 30 hours after infection, clarified by centrifugation twice at 6000 g, then aliquoted and stored at −80°C until used for antibody neutralization analysis. All sera were heat-inactivated for 30 min at 55°C before use in the neutralization assay. Vero E6 cells stably expressing TMPRSS2 were seeded in 100 μl at 2.5 × 10^4 cells per well in a 96-well collagen coated plate. The next day, twofold serially diluted serum samples were mixed with VSVΔG-RFP SARS-CoV-2 pseudotype virus (100 to 300 focus forming units per well) and incubated for 1 hour at 37°C. 1E9F9, a mouse anti-VSV Indiana G, was also included in this mixture at a concentration of 600 ng/ml (Absolute Antibody, Ab01402-2.0) to neutralize any potential VSV-G carryover virus. The serum-virus mixture was then used to replace the media on VeroE6 TMPRSS2 cells. Twenty-two hours after infection, the cells were washed and fixed with 4% paraformaldehyde before visualization on an S6 FluoroSpot Analyzer (CTL; Shaker Heights, OH). Individual infected foci were enumerated, and the values were compared with control wells without antibody. The focus reduction neutralization titer 50% ($FRNT_{50}$) was measured as the greatest serum dilution at which focus count was reduced by at least 50% relative to control cells that were infected with pseudotype virus in the absence of human serum. $FRNT_{50}$ titers for each sample were measured in at least two technical replicates and were reported for each sample as the geometric mean of the technical replicates.

Detection and phenotyping of SARS-CoV-2–specific memory B cells

Antigen-specific B cells were detected using biotinylated proteins in combination with different streptavidin (SA)–fluorophore conjugates as described (*16*). All reagents are listed in table S4. Biotinylated proteins were multimerized with fluorescently labeled SA for 1 hour at 4°C. Full-length spike protein was mixed with SA-BV421 at a 10:1 mass ratio (200 ng spike with 20 ng SA; ~4:1 molar ratio). Spike RBD was mixed with SA-APC at a 2:1 mass ratio (25 ng RBD with 12.5 ng SA; ~4:1 molar ratio). Biotinylated influenza HA pools were mixed with SA-PE at a 6.25:1 mass ratio (100 ng HA pool with 16 ng SA; ~6:1 molar ratio). Influenza HA antigens corresponding with the 2019 trivalent vaccine (A/Brisbane/02/2018/H1N1, B/Colorado/06/2017) were chosen as a historical antigen and were biotinylated using an EZ-Link Micro NHS-PEG4 Biotinylation Kit (Thermo Fisher) according to the manufacturer's instructions. Excess biotin was subsequently removed from HA antigens using Zebra Spin Desalting Columns 7K MWCO (Thermo Fisher), and protein was quantified with a Pierce BCA Assay (Thermo Fisher). SA-BV711 was used as a decoy probe without biotinylated protein to gate out cells that nonspecifically bind streptavidin. All experimental steps were performed in a 50/50 mixture of PBS + 2% FBS and Brilliant Buffer (BD Bioscience). Antigen probes for spike, RBD, and HA were prepared individually

and mixed together after multimerization with 5 μM free D-biotin (Avidity LLC) to minimize potential cross-reactivity between probes. For staining, 5×10^6 cryopreserved PBMC samples were prepared in a 96-well U-bottom plate. Cells were first stained with Fc block (Biolegend, 1:200) and Ghost 510 Viability Dye for 15 min at 4°C. Cells were then washed and stained with 50 μL antigen probe master mix containing 200 ng spike-BV421, 25 ng RBD-APC, 100 ng HA-PE, and 20 ng SA-BV711 decoy for 1 hour at 4°C. After incubation with antigen probe, cells were washed again and stained with anti-CD3, anti-CD19, anti-CD20, anti-CD27, anti-CD38, anti-CD71, anti-IgD, anti-IgM, anti-IgG, and anti-IgA for 30 min at 4°C. After surface stain, cells were washed and fixed in 1% PFA overnight at 4°C. Antigen-specific gates for B cell probe assays were set based on healthy donors stained without antigen probes (similar to an FMO control) and were kept the same for all experimental runs.

Detection of variant RBD, NTD, and S2-specific memory B cells

Variant RBD, NTD, and S2-specific memory B cells were detected using a similar approach as described above. SARS-CoV-2 nucleocapsid was used as a vaccine-irrelevant antigen control. All reagents are listed in table S4. Probes were multimerized for 1.5 hours at the following ratios (all ~4:1 molar ratios calculated relative to the streptavidin-only component irrespective of fluorophore): 200 ng full-length spike protein was mixed with 20 ng SA-BV421, 30 ng NTD was mixed with 12 ng SA-BV786, 25 ng WT RBD was mixed with 12.5 ng SA-BB515, 25 ng B.1.1.7 RBD was mixed with 12.5 ng SA-BV711, 25 ng B.1.351 RBD was mixed with 12.5 ng SA-PE, 25 ng B.1.617.2 was mixed with 12.5 ng SA-APC, 50 ng S2 was mixed with 12 ng SA-BUV737, and 50 ng nucleocapsid was mixed with 14 ng SA-BV605. 12.5 ng SA-BUV615 was used as a decoy probe. All antigen probes were multimerized separately and mixed together with 5 μM free D-biotin. Before staining, total B cells were enriched from 20×10^6 cryopreserved PBMC samples by negative selection using an EasySep human B cell isolation kit (STEMCELL, no. 17954). B cells were then prepared in a 96-well U-bottom plate and stained with Fc block and Ghost 510 Viability Dye as described above. Cells were washed and stained with 50 μL antigen probe master mix for 1 hour at 4°C. After probe staining, cells were washed again and stained with anti-CD3, anti-CD19, anti-CD27, anti-CD38, anti-IgD, and anti-IgG for 30 min at 4°C. After surface stain, cells were washed and fixed in 1X Stabilizing Fixative (BD Biosciences) overnight at 4°C.

For sorting, pre-enriched B cells were stained with Fc block and Ghost 510 Viability Dye, followed by full-length spike, WT RBD, and B.1.351 RBD probes as described above. Cells were then stained for surface markers with anti-CD19, anti-CD20, anti-CD27, and anti-CD38, and anti-IgD. After surface stain, cells were washed and resuspended in PBS + 2% FBS for acquisition.

In vitro differentiation of memory B cells to antibody-secreting cells

Memory B cells from bulk PBMC samples were differentiated into antibody-secreting cells as described (*39*). Briefly, 1×10^6 cryopreserved PBMCs were seeded in 1 mL of complete RPMI media (RPMI + 10% FBS + 1% Pen/Strep) in 24-well plates. PBMCs were then stimulated with 1000 U/mL recombinant human IL-2 and 2.5 μg/mL R848 for 10 days. Supernatants were collected at the indicated time points. anti-spike IgG was quantified using a Human SARS-CoV-2 spike (Trimer) IgG ELISA Kit (Invitrogen) according to the manufacturer's instructions. RBD-ACE2 binding inhibition was measured using a SARS-CoV-2 Neutralizing Ab ELISA Kit (Invitrogen). For anti-spike IgG experiments, culture supernatants were tested at 1:100 and 1:1000 dilutions. For RBD inhibition experiments, culture supernatants were tested without dilution and at a 1:2 dilution. Pseudovirus neutralization titers were also measured in culture supernatants starting at a 1:2 dilution as described above.

Detection of SARS-CoV-2–specific T cells

SARS-CoV-2–specific T cells were detected using an activation induced marker assay. All reagents are listed in table S5. PBMCs were thawed by warming frozen cryovials in a 37°C water bath and resuspending cells in 10 mL of RPMI supplemented with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin (R10). Cells were washed once in R10, counted using a Countess automated cell counter (Thermo Fisher), and resuspended in fresh R10 to a density of 5×10^6 cells/mL. For each condition, duplicate wells containing 1×10^6 cells in 200 μL were plated in 96-well round-bottom plates and rested overnight in a humidified incubator at 37°C, 5% CO2. After 16 hours, CD40 blocking antibody (0.5 μg/mL final concentration) was added to cultures for 15 min before stimulation. Cells were then stimulated for 24 hours with costimulation (anti-human CD28/CD49d, BD Biosciences) and peptide megapools (CD4-S for all $CD4^+$ T cell analyses, CD8-E for all $CD8^+$ T cell analyses) at a final concentration of 1 ug/mL. Peptide megapools were prepared as previously described (*51*, *52*). Matched unstimulated samples for each donor at each time point were treated with costimulation alone. Twenty hours poststimulation, antibodies targeting CXCR3, CCR7, CD40L, CD107a, CXCR5, and CCR6 were added to the culture along with monensin (GolgiStop, BD Biosciences) for a 4-hour stain at 37°C. After 4 hours, duplicate wells were pooled, and cells were washed in PBS supplemented with 2% FBS [fluorescence-activated cell sorting (FACS) buffer]. Cells were stained for 10 min at room temperature with Ghost Dye Violet 510 and Fc receptor blocking solution (Human TruStain FcX, BioLegend) and washed once in FACS buffer. Surface staining for 30 min at room temperature was then performed with antibodies directed against CD4, CD8, CD45RA, CD27, CD3, CD69, CD40L, CD200, OX40, and 41BB in FACS buffer. Cells were washed once in FACS buffer, fixed and permeabilized for 30 min at room temperature (eBioscience Foxp3 / Transcription Factor Fixation/Permeabilization Concentrate and Diluent), and washed once in 1X Permeabilization Buffer before staining for intracellular interferon-γ (IFN-γ) overnight at 4°C. Cells were then washed again and resuspended in 1% paraformaldehyde in PBS before data acquisition.

All data from AIM expression assays were background-subtracted using paired unstimulated control samples. For memory T cell and helper T cell subsets, the AIM^+ background frequency of non-naïve T cells was subtracted independently for each subset. AIM^+ cells were identified from non-naïve T cell populations. AIM^+ $CD4^+$ T cells were defined by coexpression of CD200 and CD40L. AIM^+ $CD8^+$ T cells were defined by a Boolean analysis identifying cells expressing at least four of five markers: CD200, CD40L, 41BB, CD107a, and intracellular IFN-γ.

Flow cytometry and cell sorting

Samples were acquired on a BD Symphony A5 instrument. Standardized SPHERO rainbow beads (Spherotech) were used to track and adjust photomultiplier tubes over time. UltraComp eBeads (Thermo Fisher) were used for compensation. Up to 5×10^6 cells were acquired per sample. Data were analyzed using FlowJo v10 (BD Bioscience). For Boolean analysis of variant cross-binding, data were imported into SPICE 6 [NIH Vaccine Research Center (*55*)]. Cell sorting was performed on a BD FACS Aria II instrument in low pressure mode using a 70 μm nozzle. Cells were sorted into DNA LoBind Eppendorf tubes containing cell lysis buffer (Qiagen).

B cell receptor sequencing

Library preparation

DNA was extracted from sorted cells using a Gentra Puregene Cell kit (Qiagen, catalog no. 158767). Immunoglobulin heavy-chain family–specific PCRs were performed on genomic DNA samples using primers in FR1 and JH as described previously (*47*, *56*). Two biological replicates were run on all samples. Sequencing was performed in the Human Immunology

Core Facility at the University of Pennsylvania using an Illumina 2× 300-bp paired-end kit (Illumina MiSeq Reagent Kit v3, 600-cycle, Illumina MS-102-3003).

IGH sequence analysis

Reads from an Illumina MiSeq were filtered, annotated, and grouped into clones as described previously (*16*, *57*). Briefly, pRESTO v0.6.0 (*58*) was used to align paired end reads, remove short and low-quality reads, and mask low-quality bases with *N*s to avoid skewing SHM and lineage analyses. Sequences which passed this process were aligned and annotated with IgBLAST v1.17.0 (*59*). The annotated sequences were then imported into ImmuneDB v0.29.10 (*60*, *61*) for clonal inference, lineage construction, and downstream processing. For clonal inference, sequences with the same IGHV gene, IGHJ gene, and CDR3 length from each donor were hierarchically clustered. Sequences with 85% or higher similarity in their CDR3 amino-acid sequence were subsequently grouped into clones. Clones with productive rearrangements and ≥2 copies were filtered for downstream analysis.

Lineage construction and visualization

For each clone, a lineage was constructed with ImmuneDB as described in (*61*). ete3 (*62*) was used to visualize the lineages where each node represents a unique sequence, the size of a node represents its relative copy number fraction in the clone, and the integer next to each node represents the number of mutations from the preceding vertical node.

Overlapping clone SHM analysis

Clones were filtered on the basis of size using a copy number filter such that clones that had a copy number less than 50% of the mean copy number frequency (50% mcf) within the subject were excluded. From this population, only clones that appeared in both WT RBD and cross-binder (RBD^{++}) samples were included. The SHM of each clone was averaged across each unique sequence, weighted by the copies of each sequence, and visualized as categorical variables (pie chart) and as frequencies (boxplots).

Data availability

Raw sequencing data for all donors and subsets are available on SRA under BioProject PRJNA752617. Processed AIRR-seq data will be made available on the AIRR Data Commons via the iReceptor portal (*63*).

Estimating decay rates

To understand and compare the rate of loss of immune responses after vaccination, we tested different statistical models of decay against the data. We first tested whether there was significant decay (i.e., was the decay rate significantly different from zero). We then tested whether there was evidence for a slowing of decay with time (using a two-phase model). This is a heuristic approach to understanding decay and does not imply a mechanism or that the underlying immune dynamics may be more complex. The decay rate after second dose of vaccine was estimated using a censored mixed effect regression framework. Briefly, the dependency of variables of interest on days after vaccine can be modeled by using either one constant decay slope or a decay slope that changes with time (assume a two-phase decay with a fixed break point at T_0). The model of the immune response y for participant i at time t_{ij} can be written as

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij}$$

for a model with a single slope and

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij} + \beta_2 s_{ij}$$

for a model with two different slopes, in which

$$s_{ij} = \begin{cases} 0, t_{ij} < T_0 \\ t_{ij} - T_0, t_{ij} \geq T_0 \end{cases}$$

The parameter β_0 is a constant (global intercept), and b_{0i} is a patient-specific adjustment (random effect) to the global intercept. The slope parameter β_1 is a fixed effect to capture the average decay rate for all individuals before T_0, and b_{1i} is a patient-specific random effect of the decay rate. To fit the model with a two-phase decay slope (with break point at time T_0), an extra parameter β_2 (with a subject-specific random effect b_{2i}) was added to represent the difference between the two slopes. Throughout the manuscript, we chose the median of the time points after the second dose of vaccine as the break point in decay rate (i.e., T_0 = day 89).

To account for values less than the detection threshold in the assay, a censored mixed-effect regression method was used to estimate the parameters in the model. Values less than 10 were censored for the neutralization data. For T cell measurements, this detection threshold varies (see next section for details on how this variable limit of detection was captured). The linear models above were fitted with censoring of values below the limit of detection using lmec library in R (*64*) (with the maximum likelihood algorithm option to fit for the fixed effects). We used a likelihood ratio test to determine whether the response variables were better fit with either the single or two-phase decay models (by testing whether $\beta_2 = 0$) and to test whether the decay rates were different between SARS-CoV-2–naïve and –recovered subjects (this test compares the likelihood value of the nested models and the difference in the number of parameters). These analyses were carried out in R version 4.0.4.

Determining the limit of detection for estimating decay rates

For each individual and at each time point (i.e., each sample), the limit of detection in assays of T cell stimulation varied. This is because the background level is determined by running paired assessment of cells from a given sample in (SARS-CoV-2 peptide) stimulated and unstimulated cultures. The quantify of interest (of which we wish to measure the decay rate) is the difference in the fraction of T cells activated in the stimulated and unstimulated cultures. The variable limit of detection (LOD) for each sample must be considered when determining the decay rate for T cell responses. To determine whether the fraction of activated cells in a stimulated sample was significantly higher than the fraction of activated cells in the corresponding unstimulated sample (i.e., if the sample was above the limit of detection), we used a one-sided two proportion Z test. Formally, we let the proportion of unstimulated and stimulated responses (over total non-naïve cells) be denoted by $U_{i,j}$ and $S_{i,j}$ for patient i at time j, respectively. It follows that we are interested in estimating the decay rate of the quantity $\Delta_{i,j} = S_{i,j} - U_{i,j}$. A one-sided two proportion Z test was used to determine if $S_{i,j} > U_{i,j}$. Briefly, for each patient i at time j, the following quantity was calculated

$$Z_{i,j} = \frac{\Delta_{ij}}{\sqrt{p(1-p)\left(\frac{1}{n_{s_{ij}}} + \frac{1}{n_{u_{ij}}}\right)}}$$

with

$$\Delta_{i,j} = S_{i,j} - U_{i,j}$$

and

$$p = \frac{s_{ij} \times n_{s_{ij}} + u_{ij} \times n_{u_{ij}}}{n_{s_{ij}} + n_{u_{ij}}}$$

where $n_{s_{ij}}$ is the total non-naïve cells in the stimulated group for subject i at time j and $n_{u_{ij}}$ is the total non-naïve cells in the unstimulated group for subject i at time j.

For each subject, we calculated the minimum difference needed to achieve significance by solving the above equation for $\Delta_{i,j}$ (assuming p is constant) at the Z_{critical} level (i.e., with $\alpha = 0.05$, $Z_{\text{critical}} = 1.645$ for a one-sided test). This minimum difference can be written as

$$\Delta_{MIN_{ij}} = 1.645 \times \sqrt{p(1-p)\left(\frac{1}{n_{s_{ij}}} + \frac{1}{n_{u_{ij}}}\right)}$$

We censored subject i if the difference is not statistically significant (i.e., $Z_{i,j} < 1.645$,

with $\alpha = 0.05$). The detection limit for subject i was calculated by taking the maximum value of $\Delta_{MIN_{i,j}}$ across all time points for that subject. The values $\Delta_{i,j}$ were normalized by the maximum $\Delta_{MIN_{i,j}}$ for each subject, hence the limit of detection was set to zero, and the lmec regression models applied to the normalized data to determine the decay rates of T cell responses.

High-dimensional analysis and statistics

All data were analyzed using custom scripts in R and visualized using RStudio. Pairwise correlations between variables were calculated and visualized as a correlogram using corrplot with false discovery rate (FDR) correction as described previously (*65*). For heatmaps, data were visualized with pheatmap. For construction of UMAPs, 12 antigen-specific immune features were selected: anti-spike IgG, anti-RBD IgG, D614G $FRNT_{50}$, B.1.351 $FRNT_{50}$, spike$^+$ memory B, RBD$^+$ memory B, % IgG$^+$ of spike$^+$ memory B, % IgG$^+$ of RBD$^+$ memory B, AIM$^+$ CD4 T, AIM$^+$ CD4 T_{FH}, AIM$^+$ CD4 T_H1, and AIM$^+$ CD8 T. Antibody and cell frequency data were $\log_{10}$ transformed and scaled by column (z-score normalization) before generating UMAP coordinates. Statistical tests are indicated in the corresponding figure legends. All tests were performed two-sided with a nominal significance threshold of $P < 0.05$. Benjamini-Hochberg correction was performed in all cases of multiple comparisons. Unpaired tests were used for comparisons between time points unless otherwise indicated because some participants were missing samples from individual time points. A single asterisk indicates $P < 0.05$, two asterisks indicate $P < 0.01$, three asterisks indicate $P < 0.001$, and four asterisks indicate $P < 0.0001$. Source code and data files are available upon request from the authors.

REFERENCES AND NOTES

1. T. Carvalho, F. Krammer, A. Iwasaki, The first 12 months of COVID-19: A timeline of immunological insights. *Nat. Rev. Immunol.* **21**, 245–256 (2021). doi: 10.1038/s41577-021-00522-1; pmid: 33723416
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
3. L. R. Baden *et al.*, Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021). doi: 10.1056/NEJMoa2035389
4. 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
5. D. Cromer *et al.*, SARS-CoV-2 variants: levels of neutralisation required for protective immunity. *medRxiv* 2021.08.11.21261876 [Preprint] (2021). .doi: 10.1101/2021.08.11.21261876
6. P. B. Gilbert *et al.*, Immune Correlates Analysis of the mRNA-1273 COVID-19 Vaccine Efficacy Trial. *medRxiv* 2021.08.09.21261290 [Preprint] (2021). doi: 10.1101/2021.08.09.21261290
7. N. Doria-Rose *et al.*, mRNA-1273 Study Group, Antibody Persistence through 6 Months after the Second Dose of mRNA-1273 Vaccine for Covid-19. *N. Engl. J. Med.* **384**, 2259–2261 (2021). doi: 10.1056/NEJMc2103916; pmid: 33822494
8. M. Bergwerk *et al.*, Covid-19 Breakthrough Infections in Vaccinated Health Care Workers. *N. Engl. J. Med.* **385**, 1474–1484 (2021). doi: 10.1056/NEJMoa2109072
9. A. Israel *et al.*, Elapsed time since BNT162b2 vaccine and risk of SARS-CoV-2 infection in a large cohort. *medRxiv* 2021.08.03.21261496 [Preprint] (2021). doi: 10.1101/2021.08.03.21261496
10. S. Y. Tartof *et al.*, Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: A retrospective cohort study. *Lancet* **398**, P1407–1416 (2021). doi: 10.1016/S0140-6736(21)02183-8; pmid: 34619098
11. J. B. Griffin *et al.*, SARS-CoV-2 Infections and Hospitalizations Among Persons Aged ≥16 Years, by Vaccination Status — Los Angeles County, California, May 1–July 25, 2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, 1170–1176 (2021). doi: 10.15585/mmwr.mm7034e5
12. S. J. Thomas *et al.*, Six Month Safety and Efficacy of the BNT162b2 mRNA COVID-19 Vaccine. *medRxiv* 2021.07.28.21261159 [Preprint] (2021). doi: 10.1101/2021.07.28.21261159
13. P. S. Arunachalam *et al.*, Systems vaccinology of the BNT162b2 mRNA vaccine in humans. *Nature* **596**, 410–416 (2021). doi: 10.1038/s41586-021-03791-x; pmid: 34252919
14. J. S. Turner *et al.*, SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature* **596**, 109–113 (2021). doi: 10.1038/s41586-021-03738-2; pmid: 34182569
15. K. Lederer *et al.*, Germinal center responses to SARS-CoV-2 mRNA vaccines in healthy and immunocompromised individuals. *medRxiv* 2021.09.16.21263686 [Preprint] (2021). doi: 10.1101/2021.09.16.21263686
16. R. R. Goel *et al.*, Distinct antibody and memory B cell responses in SARS-CoV-2 naïve and recovered individuals after mRNA vaccination. *Sci. Immunol.* **6**, eabi6950 (2021). doi: 10.1126/sciimmunol.abi6950; pmid: 33858945
17. A. Cho *et al.*, Anti- SARS-CoV-2 Receptor Binding Domain Antibody Evolution after mRNA Vaccination. *bioRxiv* 2021.07.29.454333 [Preprint] (2021). doi: 10.1101/2021.07.29.454333
18. A. Mazzoni *et al.*, First-dose mRNA vaccination is sufficient to reactivate immunological memory to SARS-CoV-2 in subjects who have recovered from COVID-19. *J. Clin. Invest.* **131**, e149150 (2021). doi: 10.1172/JCI149150; pmid: 33939647
19. M. M. Painter *et al.*, Rapid induction of antigen-specific CD4$^+$ T cells is associated with coordinated humoral and cellular immunity to SARS-CoV-2 mRNA vaccination. *Immunity* **54**, 2133–2142.e3 (2021). doi: 10.1016/j.immuni.2021.08.001; pmid: 34453880
20. V. Oberhardt *et al.*, Rapid and stable mobilization of CD8$^+$ T cells by SARS-CoV-2 mRNA vaccine. *Nature* **597**, 268–273 (2021). doi: 10.1038/s41586-021-03841-4; pmid: 34320609
21. A. Tarke *et al.*, Impact of SARS-CoV-2 variants on the total CD4$^+$ and CD8$^+$ T cell reactivity in infected or vaccinated individuals. *Cell Rep. Med.* **2**, 100355 (2021). doi: 10.1016/j.xcrm.2021.100355
22. J. Mateus *et al.*, Low dose mRNA-1273 COVID-19 vaccine generates durable T cell memory and antibodies enhanced by pre-existing crossreactive T cell memory. *medRxiv* 2021.06.30.21259787 [Preprint] (2021). doi: 10.1101/2021.06.30.21259787
23. D. Cromer *et al.*, Prospects for durable immune control of SARS-CoV-2 and prevention of reinfection. *Nat. Rev. Immunol.* **21**, 395–404 (2021). doi: 10.1038/s41577-021-00550-x; pmid: 33927374
24. M. Akkaya, K. Kwak, S. K. Pierce, B cell memory: Building two walls of protection against pathogens. *Nat. Rev. Immunol.* **20**, 229–238 (2020). doi: 10.1038/s41577-019-0244-2; pmid: 31836872
25. D. L. Farber, N. A. Yudanin, N. P. Restifo, Human memory T cells: Generation, compartmentalization and homeostasis. *Nat. Rev. Immunol.* **14**, 24–35 (2013). doi: 10.1038/nri3567; pmid: 24336101
26. M. C. Shamier *et al.*, Virological characteristics of SARS-CoV-2 vaccine breakthrough infections in health care workers. *medRxiv* 2021.08.20.21262158 [Preprint] (2021). doi: 10.1101/2021.08.20.21262158
27. R. Ke *et al.*, Longitudinal analysis of SARS-CoV-2 vaccine breakthrough infections reveal limited infectious virus shedding and restricted tissue distribution. *medRxiv* 2021.08.30.21262701 [Preprint] (2021). doi: 10.1101/2021.08.30.21262701
28. J. M. Dan *et al.*, Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, eabf4063 (2021). doi: 10.1126/science.abf4063; pmid: 33408181
29. K. W. Cohen *et al.*, Longitudinal analysis shows durable and broad immune memory after SARS-CoV-2 infection with persisting antibody responses and memory B and T cells. *Cell Rep. Med.* **2**, 100354 (2021). doi: 10.1016/j.xcrm.2021.100354; pmid: 34250512
30. D. A. Collier *et al.*, Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021). doi: 10.1038/s41586-021-03412-7
31. D. Zhou *et al.*, Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348–2361.e6 (2021). doi: 10.1016/j.cell.2021.02.037; pmid: 33730597
32. C. Liu *et al.*, Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* **184**, 4220–4236.e13 (2021). doi: 10.1016/j.cell.2021.06.020; pmid: 34242578
33. W. F. Garcia-Beltran *et al.*, Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372–2383.e9 (2021). doi: 10.1016/j.cell.2021.03.013
34. L. Stamatatos *et al.*, mRNA vaccination boosts cross-variant neutralizing antibodies elicited by SARS-CoV-2 infection. *Science* **372**, 1413–1418 (2021). doi: 10.1126/science.abg9175; pmid: 33766944
35. C. J. Reynolds *et al.*, Prior SARS-CoV-2 infection rescues B and T cell responses to variants after first vaccine dose. *Science* **372**, 1418–1423 (2021). doi: 10.1126/science.abh1282; pmid: 33931567
36. Z. Wang *et al.*, Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature* **595**, 426–431 (2021). doi: 10.1038/s41586-021-03696-9; pmid: 34126625
37. A. H. Ellebedy *et al.*, Defining antigen-specific plasmablast and memory B cell subsets in blood after viral infection or vaccination. *Nat. Immunol.* **17**, 1226–1234 (2016). doi: 10.1038/ni.3533; pmid: 27525369
38. A. Nellore *et al.*, Influenza-specific effector memory B cells predict long-lived antibody responses to vaccination in humans. *bioRxiv* 643973 [Preprint] (2021). doi: 10.1101/643973
39. M. Jahnmatz *et al.*, Optimization of a human IgG B-cell ELISpot assay for the analysis of vaccine-induced B-cell responses. *J. Immunol. Methods* **391**, 50–59 (2013). doi: 10.1016/j.jim.2013.02.009; pmid: 23454005
40. S. Gouma *et al.*, Health care worker seromonitoring reveals complex relationships between common coronavirus antibodies and COVID-19 symptom duration. *JCI Insight* **6**, e150449 (2021). doi: 10.1172/jci.insight.150449; pmid: 34237028
41. K. W. Ng *et al.*, Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* **370**, 1339–1343 (2020). doi: 10.1126/science.abe1107; pmid: 33159009
42. P. Nguyen-Contant *et al.*, S protein-reactive IGG and memory B cell production after human SARS-CoV-2 infection includes broad reactivity to the S2 subunit. *mBio* **11**, 1–11 (2020). doi: 10.1128/mBio.01991-20; pmid: 32978311
43. J. Pallesen *et al.*, Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7348–E7357 (2017). doi: 10.1073/pnas.1707304114; pmid: 28807998
44. K. S. Corbett *et al.*, SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* **586**, 567–571 (2020). doi: 10.1038/s41586-020-2622-0; pmid: 32756549
45. T. J. C. Tan *et al.*, Sequence signatures of two public antibody clonotypes that bind SARS-CoV-2 receptor binding domain. *Nat. Commun.* **12**, 3815 (2021). doi: 10.1038/s41467-021-24123-7; pmid: 34155209
46. H. L. Dugan *et al.*, Profiling B cell immunodominance after SARS-CoV-2 infection reveals antibody evolution to non-neutralizing viral targets. *Immunity* **54**, 1290–1303.e7 (2021). doi: 10.1016/j.immuni.2021.05.001; pmid: 34022127
47. A. M. Rosenfeld *et al.*, Computational evaluation of B-cell clone sizes in bulk populations. *Front. Immunol.* **9**, 1472 (2018). doi: 10.3389/fimmu.2018.01472; pmid: 30008715
48. C. Gaebler *et al.*, Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**, 639–644 (2021). doi: 10.1038/s41586-021-03207-w; pmid: 33461210
49. M. G. de Mattos Barbosa *et al.*, IgV somatic mutation of human anti-SARS-CoV-2 monoclonal antibodies governs neutralization and breadth of reactivity. *JCI Insight* **6**, e147386 (2021). doi: 10.1172/jci.insight.147386; pmid: 33769311
50. D. Geers *et al.*, SARS-CoV-2 variants of concern partially escape humoral but not T cell responses in COVID-19

convalescent donors and vaccine recipients. *Sci. Immunol.* **6**, eabj1750 (2021). doi: 10.1126/sciimmunol.abj1750
51. A. Tarke *et al*., Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep. Med.* **2**, 100204 (2021). doi: 10.1016/j.xcrm.2021.100204; pmid: 33521695
52. A. Grifoni *et al*., Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* **181**, 1489–1501.e15 (2020). doi: 10.1016/j.cell.2020.05.015; pmid: 32473127
53. S. M. Kaech, E. J. Wherry, R. Ahmed, Effector and memory T-cell differentiation: Implications for vaccine development. *Nat. Rev. Immunol.* **2**, 251–262 (2002). doi: 10.1038/nri778; pmid: 12001996
54. D. D. Flannery *et al*., SARS-CoV-2 seroprevalence among parturient women in Philadelphia. *Sci. Immunol.* **5**, eabd5709 (2020). doi: 10.1126/sciimmunol.abd5709; pmid: 32727884
55. M. Roederer, J. L. Nozzi, M. C. Nason, SPICE: Exploration and analysis of post-cytometric complex multivariate datasets. *Cytometry A* **79**, 167–174 (2011). doi: 10.1002/cyto.a.21015; pmid: 21265010
56. W. Meng *et al*., An atlas of B-cell clonal distribution in the human body. *Nat. Biotechnol.* **35**, 879–884 (2017). doi: 10.1038/nbt.3942; pmid: 28829438
57. L. Kuri-Cervantes *et al*., Comprehensive mapping of immune perturbations associated with severe COVID-19. *Sci. Immunol.* **5**, eabd7114 (2020). doi: 10.1126/sciimmunol.abd7114; pmid: 32669287
58. J. A. Vander Heiden *et al*., pRESTO: A toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014). doi: 10.1093/bioinformatics/btu138; pmid: 24618469
59. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: An immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–W40 (2013). doi: 10.1093/nar/gkt382; pmid: 23671333
60. A. M. Rosenfeld, W. Meng, E. T. Luning Prak, U. Hershberg, ImmuneDB: A system for the analysis and exploration of high-throughput adaptive immune receptor sequencing data. *Bioinformatics* **33**, 292–293 (2017). doi: 10.1093/bioinformatics/btw593; pmid: 27616708
61. A. M. Rosenfeld, W. Meng, E. T. Luning Prak, U. Hershberg, ImmuneDB, a Novel Tool for the Analysis, Storage, and Dissemination of Immune Repertoire Sequencing Data. *Front. Immunol.* **9**, 2107 (2018). doi: 10.3389/fimmu.2018.02107; pmid: 30298069
62. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016). doi: 10.1093/molbev/msw046; pmid: 26921390
63. B. D. Corrie *et al*., iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol. Rev.* **284**, 24–41 (2018). doi: 10.1111/imr.12666; pmid: 29944754
64. F. Vaida, L. Liu, Fast Implementation for Normal Mixed Effects Models With Censored Response. *J. Comput. Graph. Stat.* **18**, 797–817 (2009). doi: 10.1198/jcgs.2009.07130; pmid: 25829836
65. D. Mathew *et al*., Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science* **369**, eabc8511 (2020). doi: 10.1126/science.abc8511; pmid: 32669297

ACKNOWLEDGMENTS

We thank the study participants for their generosity in making the study possible. We also thank S. Crotty and members of the Wherry laboratory for helpful discussions and feedback, as well as the Flow Cytometry Core and the Human Immunology Core at the University of Pennsylvania for technical support. **Funding:** This work was supported by NIH grants AI105343, AI082630, AI108545, AI155577, AI149680 (to E.J.W.), AI152236, AI142638 (to P.B.), HL143613 (to J.R.G.), P30-AI0450080 (to E.T.L.P.), R38 HL143613 (to D.A.O.), T32 AR076951-01 (to S.A.A.), T32 CA009140 (to J.R.G., D.A.O., and D.M.), T32 AI055400 (to P.H.), and U19AI082630 (to S.E.H. and E.J.W.); NIH contract no. 75N9301900065 (to D.W. and A.S.); Australian government Medical Research Future Fund awards GNT2002073 (to M.P.D.), MRF2005544 (to M.P.D.), and MRF2005760 (to M.P.D.); an NHMRC program grant GNT1149990 (to M.P.D.); NHMRC Fellowship and Investigator grants (to D.S.K. and M.P.D.); funding from the National Health and Medical Research Council of Australia and the Australian Research Council (to D.S.K.); funding from the Allen Institute for Immunology (to S.A.A. and E.J.W.); a Cancer Research Institute-Mark Foundation Fellowship (to J.R.G.); the Chen Family Research Fund (to S.A.A.); the Parker Institute for Cancer Immunotherapy (to J.R.G. and E.J.W.); funding from Moderna and Janssen (to I.F.); the Penn Center for Research on Coronavirus and Other Emerging Pathogens (to P.B.); the University of Pennsylvania Perelman School of Medicine COVID Fund (to R.R.G. and E.J.W.); the University of Pennsylvania Perelman School of Medicine 21st Century Scholar Fund (to R.R.G.); and a philanthropic gift from J. Lurie, J. Embiid, J. Harris, and D. Blitzer (to S.E.H.). **Author contributions:** R.R.G., M.M.P., and E.J.W. designed the study. R.R.G., M.M.P., S.A.A., D.M., W.M., K.A.L., S.G., L.K.-C., P.H., S.D., M.E.W., C.M.M., M.A., N.T., and E.M.D. carried out experiments. R.R.G., S.A.A., J.D., S.L., and O.K. were involved in clinical recruitment and sample collection. W.M., A.M.R., A.R., D.S.K., D.A.O., J.R.G., M.P.D., and E.T.L.P. provided expertise on statistical analyses. R.R.G., M.M.P., D.M., and A.E.B. contributed to the methodology. R.R.G., M.M.P., A.P., A.H., H.S., S.H., S.K., J.T.H., J.C.W., and S.A. processed peripheral blood samples and managed the sample database. I.F., A.G., D.W., and A.S. provided key samples and/or reagents. E.T.L.P., A.R.G., and E.J.W. supervised the study. All authors participated in data analysis and interpretation. R.R.G., M.M.P., E.T.L.P., and E.J.W. wrote the manuscript. **Competing interests:** A.S. is a consultant for Gritstone, Flow Pharma, CellCarta, Arcturus, Oxfordimmunotech, and Avalia. La Jolla Institute for Immunology has filed for patent protection for various aspects of T cell epitope and vaccine design work. S.E.H. has received consultancy fees from Sanofi Pasteur, Lumen, Novavax, and Merk for work unrelated to this study. E.T.L.P. is consulting for or is an advisor for Roche Diagnostics, Enpicom, the Antibody Society, IEDB, and the American Autoimmune Related Diseases Association. A.R.G. is a consultant for Relation Therapeutics. M.R.B. is a consultant for Interius Biotherapeutics. I.F. receives research funding from Moderna and Janssen. E.J.W. is consulting for or is an advisor for Merck, Marengo, Janssen, Related Sciences, Synthekine, and Surface Oncology. E.J.W. is a founder of Surface Oncology, Danger Bio, and Arsenal Biosciences. The authors declare no other competing interests. **Data and materials availability:** All data, code, and materials used in this manuscript are available from the authors upon reasonable request. 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.

The UPenn COVID Processing Unit
Z. Alam, M. M. Addison, K. T. Byrne, A. Chandra, H. C. Descamps, N. Han, Y. Kaminskiy, S. C. Kammerman, J. Kim, N. Markosyan, J. Han Noll, D. K. Omran, E. Perkey, E. M. Prager, D. Pueschl, A. Rennels, J. B. Shah, J. S. Shilan, N. Wellhausen, A. N. Vanderbeck
University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm0829
Figs. S1 to S7
Tables S1 to S5

26 August 2021; accepted 10 October 2021
Published online 14 October 2021
10.1126/science.abm0829

Your Legacy to Science

AN ESTATE GIFT TO THE
AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Since 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you become a member of our 1848 Society, joining Thomas Edison, Alexander Graham Bell and the many distinguished individuals whose vision led to the creation of AAAS and our world-renowned journal, *Science*, so many years ago.

Unlike many of its peers, *Science* is not for-profit. Your estate gift would provide long-term financial stability and durable annual income that will support operations and competitive innovation for years to come. **This support is vital.**

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

1848 SOCIETY AAAS

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLES

MECHANICAL BONDS

Active mechanisorption driven by pumping cassettes

Liang Feng[1]†, Yunyan Qiu[1]†, Qing-Hui Guo[2,3], Zhijie Chen[1], James S. W. Seale[1], Kun He[4], Huang Wu[1], Yuanning Feng[1], Omar K. Farha[1,5], R. Dean Astumian[6], J. Fraser Stoddart[1,2,3,7]*

Over the past century, adsorption has been investigated extensively in equilibrium systems, with a focus on the van der Waals interactions associated with physisorption and electronic interactions in the case of chemisorption. In this study, we demonstrate mechanisorption, which results from nonequilibrium pumping to form mechanical bonds between the adsorbent and the adsorbate. This active mode of adsorption has been realized on surfaces of metal-organic frameworks grafted with arrays of molecular pumps. Adsorbates are transported from one well-defined compartment, the bulk, to another well-defined compartment, the interface, thereby creating large potential gradients in the form of chemical capacitors wherein energy is stored in metastable states. Mechanisorption extends, in a fundamental manner, the scope and potential of adsorption phenomena and offers a transformative approach to control chemistry at surfaces and interfaces.

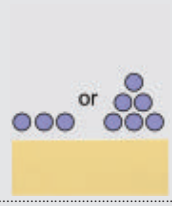
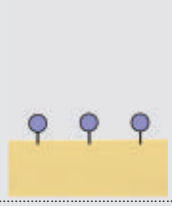
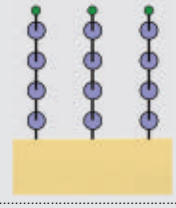
Adsorption-based phenomena play an essential role in present-day approaches to catalysis, energy storage, and environmental remediation. Langmuir (*1*) and Lennard-Jones (*2*) observed (Table 1) in the 1930s that adsorbates interact with surfaces through van der Waals interactions (physisorption) and/or by electronic interactions (chemisorption). Sorption is generally thought of as a passive process in which adsorbates move from regions of high to low concentration such that the amount of adsorbates at surfaces is always changing in the direction determined by approach to equilibrium. Although our understanding of the equilibrium aspects, involving these interactions and their practical applications, has increased drastically (*3*–*7*), little effort has been made to adsorb molecules away from equilibrium. One approach to developing nonequilibrium adsorption (*8*, *9*) is to immobilize artificial molecular machines on surfaces.

Conceptually, molecules can be pumped (*10*) by modulation of energy barriers and wells. A dynamic superstructure can be implemented chemically in the form of a pseudorotaxane (*11*) with a switchable electrostatic stopper and a steric barrier surrounding an addressable recognition site that presents a well in which the binding energy can be modulated. Recently, we have referred to this combination of components as a pumping cassette (*12*–*19*). Artificial molecular pumps (AMPs) use pumping cassettes to induce many rings to reside away from equilibrium on collecting chains. The essential modus operandi of AMPs is (Fig. 1) to recruit rings from bulk solution to the addressable recognition sites within the pumping cassettes in the first phase of an external modulation and force them onto a collecting chain in the second phase. Specific examples of AMPs that rely on redox chemistry for their operation are described in recent reviews (*17*–*20*). There is great flexibility with regard to powering these AMPs where energy can be introduced in the form of chemical fuel (*12*, *13*) or electricity (*14*–*16*) to produce high-energy oligorotaxanes with a precisely controllable number of rings.

Although AMPs have been used to drive nonequilibrium chemistry (*9*, *21*–*26*), they exhibit limited organization in bulk solution on account of their random orientations. By organizing AMPs on surfaces (*27*), it is possible to use the pumping cassette to do exactly what it does in the bulk solution, but in a confined configuration. The result is the transport of rings from one well-defined compartment, the bulk, to another well-defined compartment, the interface, thereby creating a very large chemical potential gradient commensurate with storing energy in a metastable state. Metal-organic frameworks (MOFs) (*28*–*31*) provide the ideal platforms for achieving this goal because they contain (*32*–*34*) building blocks arranged precisely in a periodic framework in which the AMPs can be immobilized and their operations maintained. Numerous examples of robust dynamics (*27*) have been observed in MOFs wherein a component undergoes rotational (*35*–*38*) or linear (*39*, *40*) movements of a nondirectional nature (*35*–*42*). In one instance, Danowski *et al.* (*43*) have described the unidirectional rotary motion of motors in a MOF. None of these examples, however, displays nonequilibrium adsorption.

Here, we report the phenomenon of mechanisorption (*44*), which results from nonequilibrium pumping (Fig. 2A) to form mechanical bonds between the adsorbent and the adsorbate. This phenomenon is associated with molecules that are transported actively to a surface compartment by using pumping cassettes and are retained in a highly nonequilibrium

Table 1. A summary of the features of three sorption types happening on a solid surface—physisorption, chemisorption, and mechanisorption.

Sorption type	Physisorption	Chemisorption	Mechanisorption
Representation			
Thermodynamics	Equilibrium systems	Equilibrium systems	Nonequilibrium systems
Interactions	Van der Waals interactions	Electronic interactions	Mechanical bonds
Selectivity	Not selective	Selective	Highly selective
Adsorption	Spontaneous	Spontaneous	Energy required
Desorption	Energy required	Energy required	Spontaneous
Location	Langmuir monolayer or BET multilayer	Monolayer	Stack

[1]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [2]Stoddart Institute of Molecular Science, Department of Chemistry, Zhejiang University, Hangzhou 310021, China. [3]ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou 311215, China. [4]Northwestern University Atomic and Nanoscale Characterization Experimental Center (NUANCE), Northwestern University, Evanston, IL 60208, USA. [5]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA. [6]Department of Physics and Astronomy, University of Maine, Orono, ME 04469, USA. [7]School of Chemistry, University of New South Wales, Sydney, NSW 2052, Australia.
*Corresponding author. Email: stoddart@northwestern.edu
†These authors contributed equally to this work.

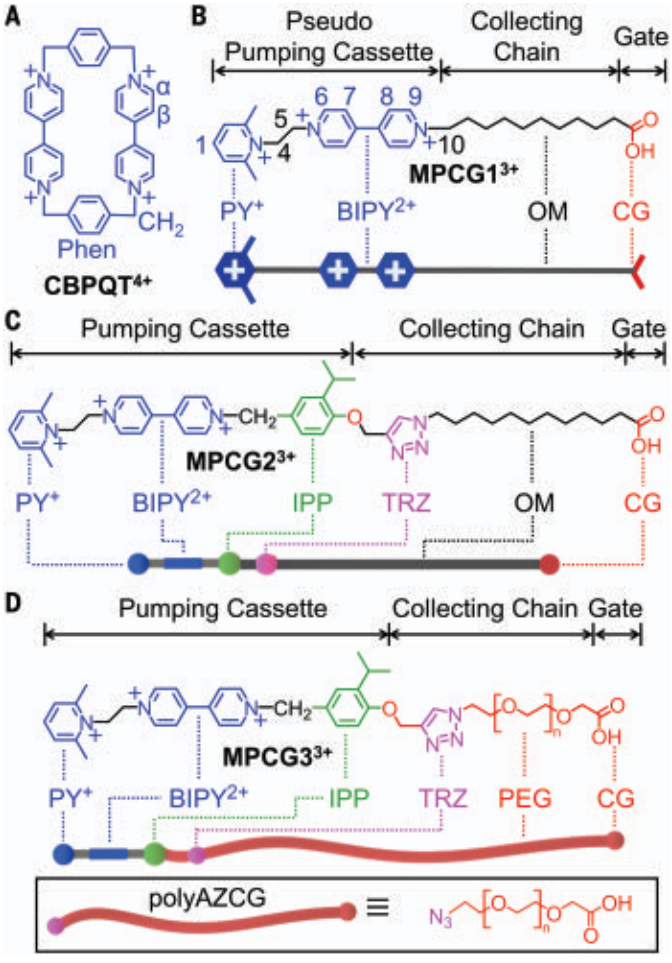


Fig. 1. The structural formulas for the CBPQT^{4+} ring and three generations of molecular pumps (MP)—MPCG1^{3+}, MPCG2^{3+}, and MPCG3^{3+}. (**A**) The structural formula for the CBPQT^{4+} ring. Selected key protons on CBPQT^{4+} (Phen, CH_2, α, and β) are labeled to aid the interpretation of ^{1}H NMR spectra reproduced in Fig. 3. (**B**) The prototype MPCG1^{3+} consists of a PY$^+$ Coulombic barrier and a BIPY^{2+} recognition site at one end (the pseudo pumping cassette), a reversible loading and unloading carboxylate gate (CG) at the other end, and a ring-collecting oligomethylene (OM) chain located in the middle. Selected key protons on MPCG1^{3+} (numbered 1 and 4 to 10) are labeled to aid the interpretation of ^{1}H NMR spectra reproduced in Fig. 3. (**C**) Compared with MPCG1^{3+}, MPCG2^{3+} has an additional steric barrier in the form of an isopropylphenylene (IPP) unit inserted between the BIPY^{2+} recognition site and the collecting chain (OM)—which is attached to IPP by a triazole (TRZ) ring resulting from a click reaction between an azide and an alkyne—allowing for the repeated collection of CBPQT^{4+} rings from solution. (**D**) Compared with MPCG2^{3+}, MPCG3^{3+} consists of a collecting polyethylene glycol (PEG) chain instead of an OM chain located between the steric barrier (IPP) and the carboxylate gate (CG), allowing for the repeated collection of CBPQT^{4+} rings from solution. MPCG3^{3+} is prepared by a click reaction between polyAZCG—which has an azide (AZ) attached to the PEG chain terminated by CG—and an alkyne functionalized with the pumping cassette. The PF_6^- counterions are omitted for the sake of clarity.

steady state indefinitely, before being released on demand into the bulk by a process that involves the breaking of coordinative, mechanical, and noncovalent bonds. Unlike physisorption or chemisorption, we show that, when coupled with an energy source, the passive movement of adsorbates from regions of high to low concentration is reversed, giving rise to active adsorption by which adsorbate concentrations are maintained with large differences in chemical potential on surfaces and in bulk solution. This nonequilibrium pumping is achieved on MOF nanosheets and nanoparticles, which not only function as extended stoppers but also make it possible to align and organize arrays of AMPs. These AMP-grafted MOFs are capable of undergoing repeated and near-quantitative adsorption of rings—cyclobis(paraquat-*p*-phenylene) (CBPQT^{4+}, Fig. 1A)—from solution to form high-energy oligorotaxanes under redox control. These rings, which contain redox-switchable bipyridinium (BIPY^{2+}) units, are sequestered under reducing conditions from solution and transferred on oxidation in a precise manner by pumping cassettes using an energy-ratchet mechanism onto the collecting chains of the oligorotaxanes (*9*, *45*). This redox-controlled active transport (Fig. 2B) transfers rings from bulk solution to much higher local concentrations on surfaces, enabling repeatable mechanisorption to occur in a way that drives an entire integrated system away from equilibrium.

Construction of AMP-grafted MOFs

MOFs have many features—well-defined composition, high stability, controllable stoichiometry, and variable spacing of the surface sites—that made them particularly suitable for our initial investigation of mechanisorption. We selected a robust, ultrathin MOF nanosheet composed of Zr-BTB (*46*, *47*), where BTB stands for 4,4′,4″-benzene-1,3,5-triyl-tris(benzoic acid) (Fig. 3A), as an ideal platform on which to graft AMPs. This two-dimensional MOF is constructed from six-connected $[Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4]$ (Zr_6) clusters and tritopic carboxylate linkers (BTB). The involvement of water during the synthesis hampers (*46*) the packing of the Zr-BTB layers substantially, leading to the formation of nanosheets. Transmission electron microscopy (TEM) of Zr-BTB (fig. S89) reveals that the nanosheets exhibit slightly wrinkled planes, with the Zr_6 clusters appearing as dark spots. These nanosheets can be easily dispersed into solution, forming colloidal suspensions, exhibiting a Tyndall effect (fig. S88) and offering a robust and dispersible platform for the attachment of AMPs. The phase purity of Zr-BTB has been verified by powder x-ray diffraction (PXRD) (fig. S77), and its monolayer structure confirmed by the asymmetric diffraction peaks (*48*). Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy of digested Zr-BTB indicates (figs. S1 to S11) that the Zr_6 clusters, in the form of as-synthesized nanosheets, are capped mainly by benzoates (BA^-). After numerous attempts, we were able to substitute the capping benzoates with OH^-/H_2O ligands by treating the nanosheets with either HCl/H_2O or HCl/MeCN. These exposed $Zr\text{–}OH^-/H_2O$ sites are extremely reactive toward carboxylic acids, even at room temperature (*49*, *50*).

A carboxyl-terminated molecular pump, MPCG1^{3+}, was designed (Fig. 1B). This prototype consists of a Coulombic barrier and a switchable recognition site at one end forming a pseudo pumping cassette, in addition to a reversible loading and unloading carboxylate

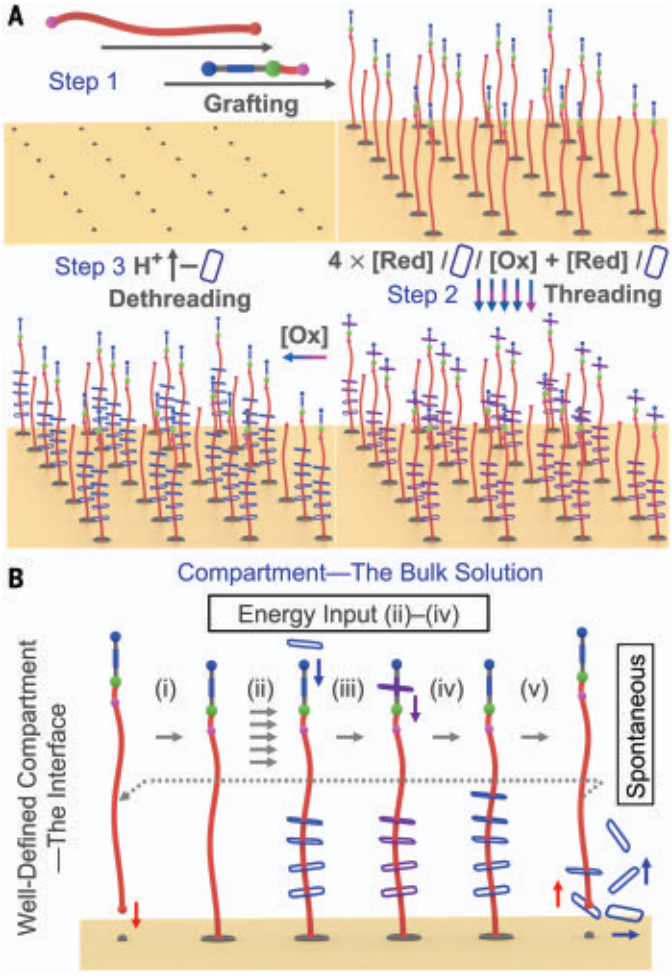


Fig. 2. Mechanisorption on MOF nanosheets with organized molecular pumps. (**A**) A graphical representation of a mechanisorption mechanism summarizing how redox and acid-base chemistry can be used to adsorb and desorb precisely a fleet of rings onto and off the surface of an MOF. Step 1: Grafting of polyAZCG onto the MOF surface followed by a click reaction with an alkyne containing the pumping cassette. Step 2: Adsorption of rings by AMP-grafted MOF nanosheets after multiple redox cycles. Step 3: Acid-driven desorption of rings and $MPCG3^{3+}$. (**B**) The cascade processes taking place at each representative molecular pump, leading to quantitative ring adsorption on organized polymer chains attached to the MOF surface and subsequent acid-triggered ring release: (i) grafting of $MPCG3^{3+}$ on the MOF surface; (ii) repeated redox-driven adsorption of rings; (iii) reduction-driven adsorption of rings from solution via the pumping cassette and the formation of trisradical-tricationic complexes on the basis of radical-radical interactions; (iv) oxidation-driven repulsion of rings from the pumping cassette onto the collecting chain (PEG) and the formation of mechanically interlocked extended oligorotaxanes; (v) acid-triggered desorption of rings associated with the cleavage of coordination bonds between $MPCG3^{3+}$ and surfaces as well as mechanical bonds between rings and surfaces. The conformations and arrangement of $MPCG3^{3+}$ are idealized for the sake of clarity.

gate (CG) at the other end with a ring-collecting chain located in the middle of a pseudo-dumbbell. An acid-base reaction (Fig. 3C and fig. S13) between Zr-BTB and $MPCG1^{3+}$ in MeCN at room temperature for 3 days generated Zr-BTB-MPCG1. Its PXRD pattern demonstrates (fig. S77) that the structure of Zr-BTB is maintained during the grafting of the molecular pump, whereas ^{1}H NMR spectroscopy, carried out on digested (D_2SO_4/CD_3SOCD_3) Zr-BTB-MPCG1, indicates the preservation of $MPCG1^{3+}$ during the grafting process (Fig. 3E and fig. S24). The approximate composition of Zr-BTB-MPCG1 was calculated (table S2) to be $Zr_6(BTB)_{2.4}(MPCG1)_{0.4}$, showing that each Zr_6 cluster is coordinated to 0.4 $MPCG1^{3+}$ and that the Zr-BTB-MPCG1 contains ~10^{17} organized $MPCG1^{3+}$ per milligram or, expressed another way, 10^{13} per square centimeter. The average distance between pairs of molecular pumps in Zr-BTB-MPCG1 is estimated to be 3.1 nm (fig. S48), a separation that is large enough to ensure the independence of the dynamics associated with each $MPCG1^{3+}$, while preserving the essential characteristics and properties of the robust MOF.

The surface modification of Zr-BTB is realized through the partial replacement of the OH^-/H_2O ligands on the MOF surface by $MPCG1^{3+}$ carboxylates. The difference in the local bonding state between Zr-BTB and Zr-BTB-MPCG1 has been probed (figs. S73 and S74) by infrared (IR) spectroscopy. Above 3500 cm^{-1}, a sharp band at 3698 cm^{-1} is observed for Zr-BTB and is assigned to the stretching mode of the OH^-/H_2O coordinated to Zr. The shift of this band to 3648 cm^{-1} indicates the changes of Zr_6 coordinating environments and the successful substitution of OH^-/H_2O ligands on the Zr_6 clusters by the CG of $MPCG1^{3+}$. Additionally, the signals of the coordinating $MPCG1^{3+}$ can also be identified (fig. S70) in the ^{13}C magic-angle spinning (MAS) solid-state (SS) NMR spectra of Zr-BTB-MPCG1: carbons from methylene and methyl groups that resonate at 30 to 40 parts per million (ppm) and carbon from $BIPY^{2+}$/2,6-dimethylpyridinium (PY^+) groups that resonate at 139 ppm can both be identified. Compared with the ^{1}H MAS SSNMR spectrum of Zr-BTB, the Zr-BTB-MPCG1 spectrum shows (fig. S70) an increase in the peak-area intensities for protons from the benzenoid rings and a decrease in the peak-area intensities of terminal OH^-/H_2O groups, demonstrating the successful exchange of the OH^-/H_2O ligands by $MPCG1^{3+}$.

Redox-driven mechanisorption and acid-triggered desorption

The redox-driven mechanisorption (*9*, *45*) of $CBPQT^{4+}$ rings on Zr-BTB-MPCG1 operates by an energy-ratchet mechanism (fig. S57). The redox state of the system influences the height of the electrostatic barriers for threading of the rings. The barrier, arising from PY^+, is much larger for $CBPQT^{4+}$ than for its radical state $CBPQT^{2(\bullet+)}$. The redox state also influences the binding energy strongly because of radical-pairing interactions and Coulombic repulsion. Upon reduction, all the $BIPY^{2+}$ units are reduced to $BIPY^{\bullet+}$ radical cations, leading to the threading (Fig. 3C) of $CBPQT^{2(\bullet+)}$ rings onto the pumping cassettes to form trisradical-tricationic complexes (*51*) based on strong radical-radical interactions. After oxidation, Coulombic repulsion between the $PY^+/BIPY^{2+}$ units and the $CBPQT^{4+}$ ring in the pumping cassette force the ring to move onto the collecting chain of $MPCG1^{3+}$, resulting in the formation of a two-dimensional array of [2]rotaxanes with unreacted dumbbells interspersed among them.

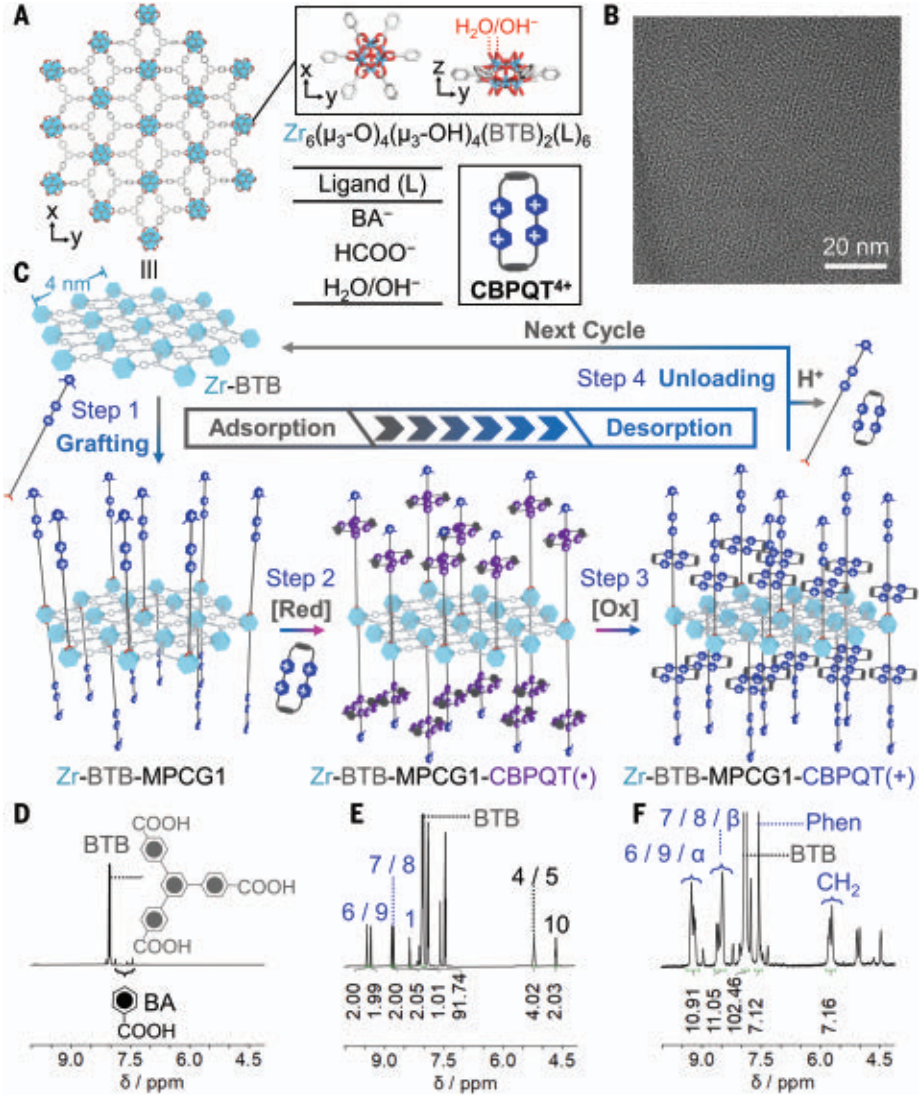


Fig. 3. Redox-driven adsorption and acid-triggered desorption of CBPQT^{4+} rings on Zr-BTB-MPCG1 surface. (**A**) Graphical representations of the structure of the Zr-BTB layer and the CBPQT^{4+} ring. The Zr-BTB layer consists of well-organized six-connected Zr_6 clusters with replaceable terminal BA^-/$HCOO^-$ and H_2O/OH^- ligands. (**B**) TEM image of Zr-BTB-MPCG1-CBPQT(+) nanosheets—where (+) refers to the fully charged states of the CBPQT^{4+} rings and the BIPY^{2+} units on the MOF surfaces. (**C**) The adsorption and desorption mechanisms on the surface of Zr-BTB-MPCG1. The Zr-BTB-MPCG1 is prepared by grafting (step 1) MPCG1^{3+} onto Zr-BTB nanosheets at room temperature. Initially, the CBPQT^{4+} rings and the Zr-BTB-MPCG1 repel each other because of strong Coulombic repulsion. Subsequently, all the BIPY^{2+} units are reduced to their radical cationic states (BIPY$^{\bullet+}$), leading to the threading (step 2) of a collection of CBPQT$^{2(\bullet+)}$ rings, one onto each MPCG1^{3+} end (the pseudo pumping cassette), on Zr-BTB-MPCG1-CBPQT(•)—where (•) refers to the radical states of CBPQT$^{2(\bullet+)}$ rings and BIPY$^{\bullet+}$ units—as a result of the formation of trisradical tricationic complexes. After oxidation, the rings proceed (step 3) onto collecting chains on account of strong Coulombic repulsion between the charged PY$^+$ and BIPY^{2+} units, leading to the formation of Zr-BTB-MPCG1-CBPQT(+). Desorption (step 4) of the rings is achieved by an acid-base reaction between Zr-BTB-MPCG1-CBPQT(+) and H_3O^+ at the carboxylate gates (CG). Meanwhile, this treatment also releases the MPCG1^{3+} and Zr-BTB for subsequent use. (**D**) Partial ^{1}H NMR spectra (600 MHz, D_2SO_4/CD_3SOCD_3, 298 K) of digested Zr-BTB. (**E**) Partial ^{1}H NMR spectra (600 MHz, D_2SO_4/CD_3SOCD_3, 298 K) of digested Zr-BTB-MPCG1. (**F**) Partial ^{1}H NMR spectra (600 MHz, D_2SO_4/CD_3SOCD_3, 298 K) of digested Zr-BTB-MPCG1-CBPQT(+). The ratio between CBPQT^{4+} and MPCG1^{3+} was calculated as 0.90:1.00, revealing a 90% threading efficiency on the Zr-BTB-MPCG1 surface during the redox cycle.

The noncoordinating redox reagents cobaltocene (Cp_2Co, reductant; Cp stands for cyclopentadienyl anion) and ferrocenium hexafluorophosphate ([Cp_2Fe]PF_6, oxidant) have been shown to be effective and innocuous as far as mechanical adsorption on Zr-BTB-MPCG1 is concerned (*52*). We subjected Zr-BTB-MPCG1 and a large excess of CBPQT^{4+} rings to one redox cycle using Cp_2Co and [Cp_2Fe]PF_6 (Fig. 3C). After repeated washings with MeCN to remove unthreaded and physisorbed rings, we isolated Zr-BTB-MPCG1-CBPQT(+) by centrifugation as a solid whose composition—$Zr_6(MPCG1)_{0.4}(CBPQT)_{0.4}$—could be obtained (Fig. 3F) by solution-state NMR spectroscopy after digestion by D_2SO_4/CD_3SOCD_3. The integrated ratio of CBPQT^{4+} to MPCG1^{3+} was calculated to be 0.90:1.00 (fig. S28), corresponding to a 90% threading efficiency on the Zr-BTB surface after one redox cycle. The presence of mechanically adsorbed CBPQT^{4+} rings on Zr-BTB-MPCG1-CBPQT(+) was also verified by ^{13}C MAS SSNMR (fig. S70), where the chemical shift of 65 ppm corresponds to the carbon resonances from CH_2 and the chemical shifts of 136 and 151 ppm correspond to the carbon resonances from BIPY^{2+} units on CBPQT^{4+} rings. Control experiments show (figs. S29 and S30) that there is no uptake of CBPQT^{4+} rings by Zr-BTB or Zr-BTB-MPCG1 in the absence of the redox cycling, ruling out the possibility of physisorption or chemisorption of CBPQT^{4+} rings on the MOF surfaces after washing.

The porosity of Zr-BTB nanosheets before and after mechanisorption was estimated by evaluation of their N_2 isotherms at 77 K (figs. S79 to S81). The Brunauer-Emmett-Teller (BET) surface area of the untreated surface was 436 $m^2\ g^{-1}$ (table S5). After grafting with MPCG1^{3+}, the BET surface area decreased from 436 to 307 $m^2\ g^{-1}$, and the mechanisorption of the CBPQT^{4+} rings reduced the surface area further, to 183 $m^2\ g^{-1}$. No notable change to the PXRD patterns or TEM images of the MOFs after grafting or mechanisorption was observed (Fig. 3B and figs. S77 and S90), which is indictive of their well-maintained crystallinity.

This AMP-grafted MOF also exhibits the capability of desorbing and resorbing rings without compromising the structural integrity of either the robust MOF or the dynamic AMPs for the simple reason that the mechanisorption only involves the cleavage and reformation of noncovalent bonding interactions, for example, mechanical and coordination bonds. Desorption of the threaded rings was achieved by an acid-base reaction (figs. S10 and S11), where a strong acid reacts with the MOF surface at the CG site to generate weak acid MPCG1^{3+}. Mixing Zr-BTB-MPCG1-CBPQT(+) with 1 M HCl aqueous or MeCN solutions at room temperature for

several hours released the mechanisorbed rings, $MPCG1^{3+}$ and Zr-BTB, for subsequent use (*53*). After desorption, the solid was digested and analyzed by ^{1}H NMR spectroscopy (fig. S31), confirming the complete release of $MPCG1^{3+}$ and $CBPQT^{4+}$ rings. The mechanisorbed rings are desorbed by lowering the barrier associated with the gate as a result of acid-triggered cleavage of coordination and mechanical bonds. Rings escape spontaneously from a high-concentration compartment into the bulk compartment with a much lower concentration.

After the successful alignment of AMPs on MOF nanosheets, we extended the construction of AMP-grafted MOFs to three dimensions (figs. S67 to S69). The surfaces of UiO-66 nanoparticles (*54*) can also be decorated with an abundance of $MPCG1^{3+}$. Accordingly, $CBPQT^{4+}$ rings can be adsorbed subsequently onto the surface of the UiO-66-MPCG1 by redox cycling and desorbed by treatment with acid. The approach is generalizable to other types of surfaces.

Repeated and near-quantitative mechanisorption

Taking cues from previous investigations (*12*–*15*), we designed a system to accumulate a couple of rings by incorporating a steric barrier between a $BIPY^{2+}$ unit and a collecting chain to form a full pumping cassette. The steric barrier prevents adsorbed rings from interfering with the subsequent recruiting of rings from solution during the repeated redox cycles. As a proof-of-concept, an isopropylphenylene (IPP) unit was introduced (Fig. 1C) as a steric barrier into the backbone of $MPCG2^{3+}$ by a copper-catalyzed azide–alkyne (click) cycloaddition (*55*). We synthesized Zr-BTB-MPCG2 through sequential acid-base and click reactions performed on these nanosheets (figs. S32 and S33). The success of the click reaction was verified by IR spectroscopy (figs. S75 and S76), where the intensity of an azide stretching band at 2095 cm^{-1} decreased after the click reaction. The ^{1}H NMR spectrum of the digested Zr-BTB-MPCG2 indicates (fig. S41) that 68% of 12-azidododecanoic acid had reacted and that the density of $MPCG2^{3+}$ on the Zr-BTB-MPCG2 surface is close to that of $MPCG1^{3+}$ on Zr-BTB-MPCG1 (fig. S48). Once again, by an energy-ratchet mechanism (Fig. 4), $CBPQT^{4+}$ rings bind and traverse synchronously the pumping cassettes of the dumbbell arrays after a single redox cycle and become localized on the collecting chains. A second redox cycle results in the installation of a second set of rings on the collecting chains of most of the [2]rotaxanes, which thereby become [3]rotaxanes grafted to Zr-BTB-MPCG2-CBPQT(+)-II. Because the amount of BTB in the solid material remains constant throughout the redox cycling, BTB is used as an internal standard for the determination of the $CBPQT^{4+}$ ring-loading efficiency. ^{1}H NMR spectroscopic analysis (fig. S47) of digested (D_2SO_4/CD_3SOCD_3) materials indicates a 96 and 92% threading efficiency after the first and second redox cycles, respectively, in keeping with the efficiency observed in solution (*13*). Notably, these highly energetic and mechanically interlocked arrays maintain (figs. S44 and S45) their chemical stability in MeCN for at least 2 weeks without losing any of the mechanisorbed rings from Zr-BTB-MPCG2. The material also demonstrates high thermal stability, as indicated by thermogravimetric analysis (TGA) coupled with mass spectrometry (figs. S85 to S87). These results suggest that Zr-BTB-MPCG2 acts on $CBPQT^{4+}$ rings to create (Fig. 4) a gradient in their local solid and solution concentrations, moving increasingly away from equilibrium, from 55.4 to 111 mmol/liter after the second redox cycle (fig. S49). The cycled mechanical adsorption of rings on Zr-BTB-MPCG2 was confirmed (figs. S71 and S76) by IR spectroscopy and ^{13}C and ^{1}H MAS SSNMR spectra, all of which demonstrates the stepwise away-from-equilibrium mechanisorption of the $CBPQT^{4+}$ rings toward ever higher local concentrations on the Zr-BTB surface.

Encouraged by these results, we designed and synthesized a high-capacity mechanisorption system—namely, Zr-BTB-MPCG3—constructed (Fig. 1D) from the $MPCG3^{3+}$ with a polyethylene glycol (PEG) collecting chain located between the IPP steric barrier and the CG gate. The PEG chain (the number average molecular weight, M_n = 2000) allows (*15*) for multiple repeated and near-quantitative redox-driven mechanical adsorptions of $CBPQT^{4+}$ rings on Zr-BTB-MPCG3. We produced Zr-BTB-MPCG3-CBPQT(+)-I, -II, -III, -IV, and -V (Fig. 5) by subjecting Zr-BTB-MPCG3 to multiple redox cycles. The cyclical redox-driven processes deliver rings consecutively onto the

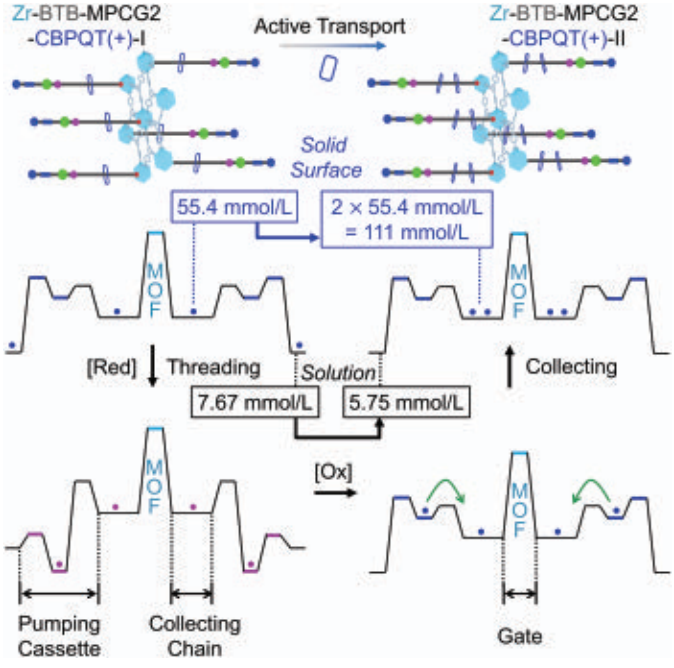


Fig. 4. Flashing energy-ratchet mechanism driving rings away from equilibrium toward a higher local concentration on the Zr-BTB-MPCG2 surface. $CBPQT^{4+}$ concentration analysis in the solid and the solution phases before and after the second pumping cycle of Zr-BTB-MPCG2-CBPQT(+)-I. The pump arrays act on the $CBPQT^{4+}$ rings to create a gradient in their local solid and solution concentrations, moving increasingly away from equilibrium, from 55.4 and 7.67 mmol/liter to 111 and 5.75 mmol/liter after the second redox cycle, respectively. Molar concentration of $CBPQT^{4+}$ rings on surfaces is estimated by dividing the amount of $CBPQT^{4+}$ rings on MOF surfaces (accessed by solution-state NMR) by the estimated volume of the MOF surfaces (the MOF surface area multiplied by the height of collecting chains). Bottom portion of the figure shows energy profiles representing the free energies of the integrated system, as the rings are pumped collectively from the solution phase by pumping cassettes onto the solid surfaces. The mechanisorbed high-concentration rings can be released into the bulk spontaneously after lowering the barrier associated with the gate. The green curved arrows on the energy profiles represent the kinetically favored reaction pathways during the adsorption process. The energy diagram is illustrated with purple as the reduced state and blue as the oxidized state.

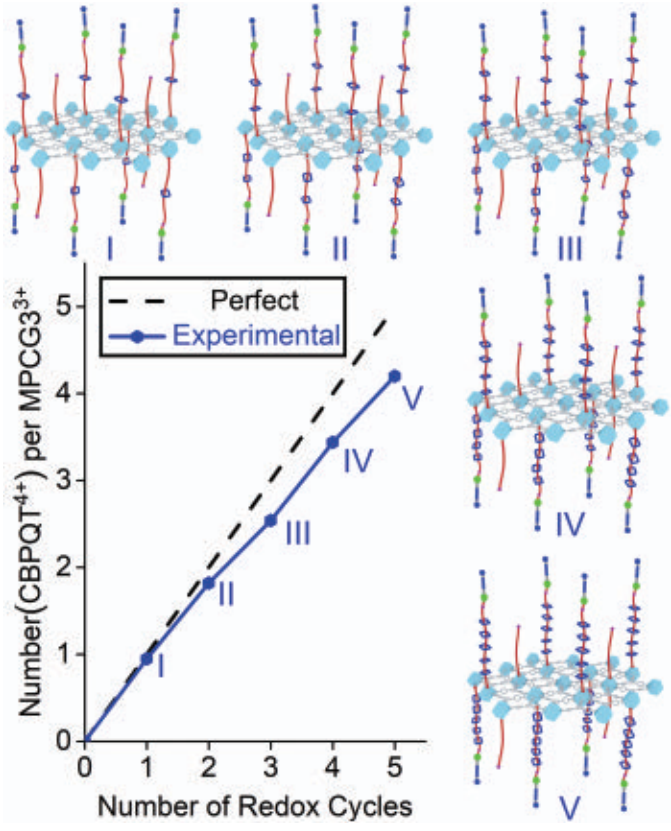


Fig. 5. Sequential and quantitative mechanisorption on the surfaces of a high-capacity Zr-BTB-MPCG3. The relationship between the number of redox cycles and the number of $CBPQT^{4+}$ rings threaded per $MPCG3^{3+}$. Sequential redox reactions drive the formation of a series of highly nonequilibrium robust dynamic systems—robust MOFs decorated with mechanically interlocked oligorotaxanes containing highly controlled numbers of rings threaded onto a PEG collecting chain—that can adsorb incoming guests sequentially and quantitatively upon stimuli. The conformations and arrangements of $MPCG3^{3+}$ are idealized for the sake of clarity. I, II, III, IV, and V refer to the one, two, three, four, and five redox cycles conducted on the surfaces, respectively.

growing [n+1]rotaxanes (n = 1 to 5) aligned on the MOF surfaces, storing increasingly higher-energy structures as the number of rings grows. The surface charge density under oxidizing conditions is $+(4n+3) \times c$, where the number density of $MPCG3^{3+}$ on the surface $c \approx 10^{16}$/mg. These systems containing robust MOFs and high-energy oligorotaxanes have been characterized (figs. S60 to S66) quantitatively by ^{1}H NMR spectroscopy on acid-digested samples. Almost every pumping cassette of $MPCG3^{3+}$ on the surface of Zr-BTB-MPCG3-CBPQT(+)-V operates simultaneously and precisely, affording ~10^{16} out-of-equilibrium [6]rotaxanes, in 1 mg of a high-capacity AMP-grafted material, wherein each [6]rotaxane carries a charge of +23. The pumping efficiency decreases slightly with increasing number of redox cycles (Fig. 5), suggesting a gradual saturation of PEG chains with rings, most likely as a result of Coulombic repulsion between the tetracationic rings. It is worth noting that active transport (*56*, *57*), in which chemical energy is used to pump materials at high concentration into a distinct compartment (e.g., a vacuole), is an essential process in biological cells.

Summary and outlook

We have demonstrated mechanisorption, which allows for energy to be stored on surfaces in the form of chemical gradients between the bulk and the surfaces. In the future, the applicability of mechanisorption could be generalized by using functionalized rings that form complexes with other chemicals being concentrated at the surface. For example, rings functionalized with azides can be clicked with alkyne-substituted molecules, polymers, and nanoparticles so that large concentrations of these guests can be stabilized at interfaces and then released at will. We can think of a chemical capacitor where a nonequilibrium amount of almost any substance can be stored under high chemical potential at a surface and then released and used on demand. The utility of these chemical capacitors in technology will be limited only by the creativity and imagination of chemists and materials scientists.

REFERENCES AND NOTES

1. I. Langmuir, *Chem. Rev.* **6**, 451–479 (1930).
2. J. Lennard-Jones, *Trans. Faraday Soc.* **28**, 333–359 (1932).
3. D. Borodin *et al.*, *Science* **369**, 1461–1465 (2020).
4. D. E. Brown, D. J. Moffatt, R. A. Wolkow, *Science* **279**, 542–544 (1998).
5. Z. Li, J. C. Barnes, A. Bosoy, J. F. Stoddart, J. I. Zink, *Chem. Soc. Rev.* **41**, 2590–2605 (2012).
6. X. Meng, B. Gui, D. Yuan, M. Zeller, C. Wang, *Sci. Adv.* **2**, e1600480 (2016).
7. L.-L. Tan *et al.*, *Chem. Sci.* **6**, 1640–1644 (2015).
8. J. H. van Esch, R. Klajn, S. Otto, *Chem. Soc. Rev.* **46**, 5474–5475 (2017).
9. C. Pezzato, C. Cheng, J. F. Stoddart, R. D. Astumian, *Chem. Soc. Rev.* **46**, 5491–5507 (2017).
10. R. D. Astumian, B. Robertson, *J. Am. Chem. Soc.* **115**, 11063–11068 (1993).
11. P. R. Ashton, D. Philip, N. Spencer, J. F. Stoddart, *J. Chem. Soc. Chem. Commun.* **1991**, 1677–1679 (1991).
12. C. Cheng *et al.*, *Nat. Nanotechnol.* **10**, 547–553 (2015).
13. C. Pezzato *et al.*, *Tetrahedron* **73**, 4849–4857 (2017).
14. Y. Qiu *et al.*, *J. Am. Chem. Soc.* **141**, 17472–17476 (2019).
15. Y. Qiu *et al.*, *Science* **368**, 1247–1253 (2020).
16. Q.-H. Guo *et al.*, *J. Am. Chem. Soc.* **142**, 14443–14449 (2020).
17. Y. Qiu, Y. Feng, Q.-H. Guo, R. D. Astumian, J. F. Stoddart, *Chem* **6**, 1952–1977 (2020).
18. Y. Feng *et al.*, *J. Am. Chem. Soc.* **143**, 5569–5591 (2021).
19. K. Cai, L. Zhang, R. D. Astumian, J. F. Stoddart, *Nat. Rev. Chem.* **5**, 447–465 (2021).
20. C. Cheng, P. R. McGonigal, J. F. Stoddart, R. D. Astumian, *ACS Nano* **9**, 8672–8688 (2015).
21. J. F. Stoddart, *Angew. Chem. Int. Ed.* **56**, 11094–11125 (2017).
22. S. Erbas-Cakmak, D. A. Leigh, C. T. McTernan, A. L. Nussbaumer, *Chem. Rev.* **115**, 10081–10206 (2015).
23. V. Balzani, A. Credi, F. M. Raymo, J. F. Stoddart, *Angew. Chem. Int. Ed.* **39**, 3348–3391 (2000).
24. G. Ragazzon, M. Baroncini, S. Silvi, M. Venturi, A. Credi, *Nat. Nanotechnol.* **10**, 70–75 (2015).
25. S. Erbas-Cakmak *et al.*, *Science* **358**, 340–343 (2017).
26. S. Amano, S. D. P. Fielden, D. A. Leigh, *Nature* **594**, 529–534 (2021).
27. H. Deng, M. A. Olson, J. F. Stoddart, O. M. Yaghi, *Nat. Chem.* **2**, 439–443 (2010).
28. S. Horike, S. Shimomura, S. Kitagawa, *Nat. Chem.* **1**, 695–704 (2009).
29. H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **341**, 1230444 (2013).
30. H. Li, M. Eddaoudi, M. O'Keeffe, O. M. Yaghi, *Nature* **402**, 276–279 (1999).
31. A. M. Rice *et al.*, *Chem. Rev.* **120**, 8790–8813 (2020).
32. M. Eddaoudi *et al.*, *Science* **295**, 469–472 (2002).
33. L. Feng, K.-Y. Wang, J. Willman, H.-C. Zhou, *ACS Cent. Sci.* **6**, 359–367 (2020).
34. S. H. Lo *et al.*, *Nat. Chem.* **12**, 90–97 (2020).
35. J. Perego *et al.*, *Nat. Chem.* **12**, 845–851 (2020).
36. S. L. Gould, D. Tranchemontagne, O. M. Yaghi, M. A. Garcia-Garibay, *J. Am. Chem. Soc.* **130**, 3246–3247 (2008).
37. Y.-S. Su *et al.*, *Nat. Chem.* **13**, 278–283 (2021).
38. V. N. Vukotic, K. J. Harris, K. Zhu, R. W. Schurko, S. J. Loeb, *Nat. Chem.* **4**, 456–460 (2012).

39. Q. Li *et al*., *Science* **325**, 855–859 (2009).
40. K. Zhu, C. A. O'Keefe, V. N. Vukotic, R. W. Schurko, S. J. Loeb, *Nat. Chem.* **7**, 514–519 (2015).
41. S. Krause, B. Feringa, *Nat. Rev. Chem.* **4**, 550–562 (2020).
42. P. Martinez-Bulit, A. J. Stirk, S. J. Loeb, *Trends Chem.* **1**, 588–600 (2019).
43. W. Danowski *et al*., *Nat. Nanotechnol.* **14**, 488–494 (2019).
44. Mechanisorption is not to be confused with the macroscopic concept of mechanosorption, which describes the usage of mechanical forces to enhance physisorption in wood and other materials.
45. R. D. Astumian, I. Derényi, *Eur. Biophys. J.* **27**, 474–489 (1998).
46. Y. Wang *et al*., *Adv. Sci.* **6**, 1802059 (2019).
47. L. Cao *et al*., *Angew. Chem. Int. Ed.* **55**, 4962–4966 (2016).
48. Y. Jiang *et al*., *Inorg. Chem.* **57**, 15123–15132 (2018).
49. L. Feng, G. S. Day, K.-Y. Wang, S. Yuan, H.-C. Zhou, *Chem* **6**, 2902–2923 (2020).
50. P. Deria *et al*., *J. Am. Chem. Soc.* **135**, 16801–16804 (2013).
51. Y. Wang, M. Frasconi, J. F. Stoddart, *ACS Cent. Sci.* **3**, 927–935 (2017).
52. Coordinating oxidants such as nitrosonium hexafluorophosphate ($NOPF_6$) were found to cleave coordination bonds between Zr-BTB and $MPCG1^{3+}$, leading to failed mechanisorption.
53. Y. Peng *et al*., *Nat. Commun.* **9**, 187 (2018).
54. W. Morris, W. E. Briley, E. Auyeung, M. D. Cabezas, C. A. Mirkin, *J. Am. Chem. Soc.* **136**, 7261–7264 (2014).
55. H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **40**, 2004–2021 (2001).
56. J. C. Skou, *Angew. Chem. Int. Ed.* **37**, 2320–2328 (1998).
57. P. Läuger, *Angew. Chem. Int. Ed.* **24**, 905–923 (1985).

ACKNOWLEDGMENTS

We acknowledge Foresight Institute and its 2021 Fellowship Program for the support, and thank K.-Y. Wang, L. Zhang, H. Chen, and Y. Jiao for experimental help and discussions. **Funding:** We thank Northwestern University (NU) for its continued support of this research. It made use of the Integrated Molecular Structure Education and Research Center (IMSERC) NMR Facility and X-Ray Facility at NU, which receives support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-2025633), NIH 1S10OD012016-01/1S10RR019071-01A1, the International Institute for Nanotechnology (IIN), and NU. This work also made use of the EPIC facility and the Keck-II facility of Northwestern University's Atomic and Nanoscale Characterization Experimental Center (NUANCE), which receives support from the SHyNE Resource (NSF ECCS-2025633), the IIN, and NU's MRSEC program (NSF DMR-1720139). O.K.F. acknowledges support from Inorganometallic Catalyst Design Center, an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), under Award DE-SC0012702. **Author contributions:** L.F., Y.Q., and J.F.S. conceived of the study and designed experiments. L.F., Y.Q., Q.-H.G., J.S.W.S., H.W., and Y.F. conducted the synthesis. L.F. conducted NMR, PXRD, and TGA analysis. Z.C., J.S.W.S., and O.K.F. conducted gas sorption studies and IR measurements. K.H. contributed to the TEM imaging studies in this work. L.F., Y.Q., Q.-H.G., J.S.W.S., R.D.A., and J.F.S. commented on the data and the presentation. All authors contributed to data analysis. L.F., R.D.A., and J.F.S. wrote the draft, with input from all authors. **Competing interests:** L.F., Y.Q., and J.F.S. have a patent application lodged with Northwestern University (INVO Reference No. NU 2021-183) based on this work. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk1391
Materials and Methods
Supplementary Text
Figs. S1 to S90
Tables S1 to S5
References (*58*–*63*)

23 June 2021; accepted 1 October 2021
Published online 21 October 2021
10.1126/science.abk1391

HUMAN GENOMICS

Genetic and functional evidence links a missense variant in *B4GALT1* to lower LDL and fibrinogen

May E. Montasser[1]*†, Cristopher V. Van Hout[2,3]†, Lawrence Miloscio[2]†, Alicia D. Howard[1,4], Avraham Rosenberg[5], Myrasol Callaway[5], Biao Shen[5], Ning Li[5], Adam E. Locke[2], Niek Verweij[2], Tanima De[2], Manuel A. Ferreira[2], Luca A. Lotta[2], Aris Baras[2], Thomas J. Daly[5], Suzanne A. Hartford[5], Wei Lin[5], Yuan Mao[5], Bin Ye[2], Derek White[5], Guochun Gong[5], James A. Perry[1], Kathleen A. Ryan[1], Qing Fang[5], Gannie Tzoneva[2], Evangelos Pefanis[5], Charleen Hunt[5], Yajun Tang[5], Lynn Lee[5], Regeneron Genetics Center Collaboration‡, Carole Sztalryd-Woodle[1,6], Braxton D. Mitchell[1,7], Matthew Healy[8], Elizabeth A. Streeten[1,9], Simeon I. Taylor[1], Jeffrey R. O'Connell[1], Aris N. Economides[2,5], Giusy Della Gatta[2]§, Alan R. Shuldiner[2]§

Increased blood levels of low-density lipoprotein cholesterol (LDL-C) and fibrinogen are independent risk factors for cardiovascular disease. We identified associations between an Amish-enriched missense variant (p.Asn352Ser) in a functional domain of beta-1,4-galactosyltransferase 1 (*B4GALT1*) and 13.9 milligrams per deciliter lower LDL-C ($P = 4.1 \times 10^{-19}$) and 29 milligrams per deciliter lower plasma fibrinogen ($P = 1.3 \times 10^{-5}$). *B4GALT1* gene–based analysis in 544,955 subjects showed an association with decreased coronary artery disease (odds ratio = 0.64, $P = 0.006$). The mutant protein had 50% lower galactosyltransferase activity compared with the wild-type protein. N-linked glycan profiling of human serum found serine 352 allele to be associated with decreased galactosylation and sialylation of apolipoprotein B100, fibrinogen, immunoglobulin G, and transferrin. *B4galt1* ^{353}Ser knock-in mice showed decreases in LDL-C and fibrinogen. Our findings suggest that targeted modulation of protein galactosylation may represent a therapeutic approach to decreasing cardiovascular disease.

Cardiovascular disease (CVD) is the leading cause of morbidity and mortality worldwide (*1*). Elevated low-density lipoprotein cholesterol (LDL-C) increases arterial plaque formation and atherosclerosis, and fibrinogen increases the risk for blood clotting and thrombosis. LDL-C is an established risk factor for coronary artery disease (CAD), and fibrinogen also plays a potential role (*2*–*4*). Variations in LDL-C and fibrinogen are governed by both genetic and environmental factors, as well as by the interplay between them (*5*, *6*).

Rare and common genetic variants have been identified for both LDL-C and fibrinogen (*7*–*9*). However, few variants have been found with pleiotropic effects on more than one CAD risk factor. Similarly, therapeutic approaches to mitigating CAD risk have focused on treating individual risk factors. Deeper understanding of the genetic determinants of LDL-C and fibrinogen may unveil new targets for therapy that may be more efficacious and safer to treat or prevent CAD.

Founder populations can facilitate the identification of new disease associations with DNA sequence variants that are enriched to a higher frequency through genetic drift. In homogeneous human populations in Iceland (*10*), Sardinia (*11*), Greenland (*12*), Samoa (*13*), and the Old Order Amish (OOA) (*14*–*20*), enriched variants with large effect sizes associated with complex diseases and traits have been identified. Although such variants are often rare or absent in the general population, they can inform biological mechanisms and provide therapeutic targets relevant to all humans.

Association analyses identify *B4GALT1* p.Asn352Ser as a new LDL-C variant

To identify genetic variants associated with LDL-C, we performed whole-exome sequencing (WES) and association analysis in 6890 OOA subjects (table S1) (*21*). Linear mixed-model association analysis identified previously ascertained loci for LDL-C, as well as a previously unidentified locus on the short arm of chromosome 9 (fig. S1 and table S2). A missense variant (rs551564683, p.Asn352Ser) in *B4GALT1* (Fig. 1A)

[1]Division of Endocrinology, Diabetes and Nutrition and Program for Personalized and Genomic Medicine, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD 21201, USA. [2]Regeneron Genetics Center, LLC, Tarrytown, NY 10591, USA. [3]Laboratorio Internacional de Investigatión sobre el Genoma Humano, Campus Juriquilla de la Universidad Nacional Autónoma de México, Querétaro, Querétaro 76230, México. [4]Center for Biologics Evaluation and Research, US Food and Drug Administration, Silver Spring, MD 20993, USA. [5]Regeneron Pharmaceuticals, Inc., Tarrytown, NY 10591, USA. [6]US Department of Veterans Affairs, Washington, DC 20420 USA. [7]Geriatrics Research and Education Clinical Center, VA Medical Center, Baltimore, MD 21201, USA. [8]Enveda Biosciences, Boulder, CO 80301, USA. [9]Division of Genetics, Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD 21201, USA.
***Corresponding author. Email: mmontass@som.umaryland.edu**
†These authors contributed equally to this work and are co-first authors. ‡A complete list of investigators contributing to the Regeneron Genetics Center Collaboration is provided at the end of this manuscript. §These authors contributed equally to this work and are co-senior authors.

was associated with 13.9 mg/dl lower LDL-C ($P = 4.1 \times 10^{-19}$, all single-variant regression used an additive model, and P values were calculated based on t test unless otherwise specified). This variant has a minor allele frequency (MAF) of 6% in the OOA population but is ultrarare across other human populations: Only eight copies were identified in 140,000 whole-genome sequences (WGS) of non-Amish participants in the National Heart, Lung, and Blood Institute (NHLBI) Trans-Omics for Precision Medicine (TOPMed) program (*22*).

To more exhaustively investigate coding and noncoding variants in this region, we performed association analysis using 1083 OOA subjects with WGS (*21*) as part of TOPMed. Despite the smaller sample size, WGS analysis identified the same variant (rs551564683) as the top association with LDL-C in this region, with a P value of 3.1×10^{-6} and an effect size of –16.9 mg/dl (Fig. 1B). In addition, WGS analysis revealed 20 variants in the region (Fig. 1A and fig. S2) in linkage disequilibrium with rs551564683 ($r^2 = 0.84$ to 0.95), with LDL-C association P values ranging from 6.3×10^{-6} to 2.3×10^{-5} (table S3). Of these 21 variants that comprise a 4-Mb OOA-specific haplotype (fig. S2), rs551564683 was the only protein-coding variant, and is classified as damaging or deleterious by the in silico protein function prediction algorithms SIFT [deleterious (*23*)], Polyphen2 [possibly damaging (*24*)], LRT [deleterious (*25*)], Mutation Taster [disease-causing (*26*)], and PROVEAN [deleterious (*27*)].

Because of its limited sample size ($n = 1083$), WGS was not able to differentiate the top missense variant ($P = 3.1 \times 10^{-6}$) from the 20 other highly correlated variants ($P = 6.3 \times 10^{-6}$ to 2.3×10^{-5}). To further differentiate among these 21 highly linked variants, we imputed genotypes in 5890 OOA subjects with genotype chip data to the TOPMed WGS reference panel (*21*). rs551564683 was the top associated variant, with a P value of 3.6×10^{-15} (Fig. 1C and table S2), two or more orders of magnitude smaller than any of the other variants (P value of 9.4×10^{-13} to 1.6×10^{-9}). Independent direct genotyping for seven of these variants gave similar results (fig. S3).

Conditional analysis adjusting for rs551564683 completely abolished the association of the other 20 variants, whereas conditional analyses adjusting for any of the other 20 variants reduced the association of rs551564683 because of the strong correlation ($r^2 = 0.84$ to 0.95) but did not abolish it ($P = 1.0 \times 10^{-3}$ to 3.0×10^{-7}). Thus, rs551564683 is the most likely causal variant in this region.

rs551564683 has a strong association with total cholesterol and non–high-density lipoprotein cholesterol (HDL-C) levels compared with a moderate association with the HDL-C and total cholesterol:HDL-C ratio and a lack of association with triglycerides, indicating that LDL-C is the main driver of this association (table S4A). Similar results were found under various models, indicating that this robust association is not driven by outliers or body weight (table S4B).

B4GALT1 is a member of the beta-1,4-galactosyltransferase gene family that encodes type II membrane-bound glycoproteins. B4GALT1 is ubiquitously expressed and plays a critical role in the processing of N-linked oligosaccharide moieties in glycoproteins, transferring the galactose from uridine diphosphate galactose (UDP-Gal) to specific glycoprotein substrates (*28*). Thus, impairment of B4GALT1 activity has the potential to alter the structure of N-linked oligosaccharides and introduce aberrations in the glycan structure that may alter glycoprotein function.

A rare homozygous frame shift insertion in *B4GALT1* that predicts a nonfunctional truncated protein was reported to cause congenital disorder of glycosylation type 2 (CDGII) (*28*–*31*). Three of the six reported CDGII patients exhibited, among other traits, abnormal coagulation and very high levels of aspartate transaminase (AST). Interestingly, rs551564683 was associated with lower levels of fibrinogen

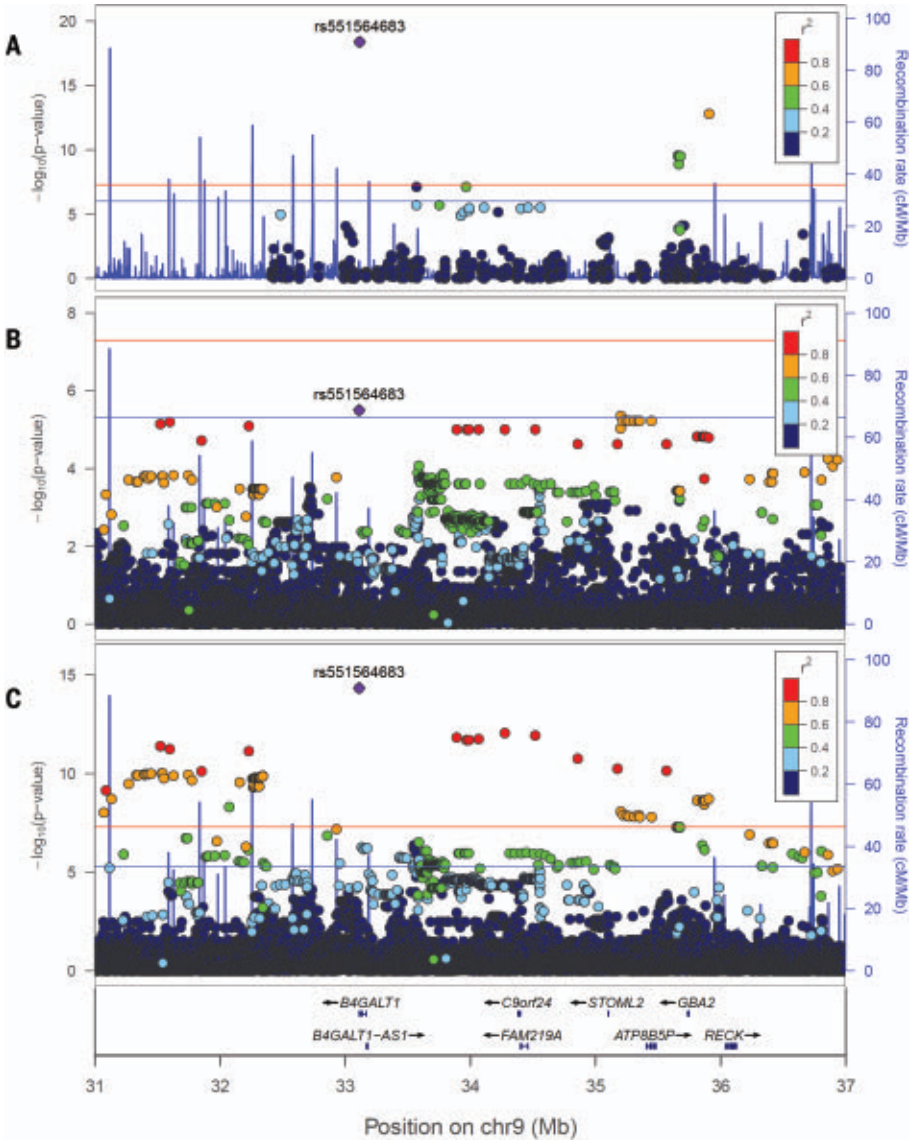


Fig. 1. Single-variant association analyses identify *B4GALT1* p.Asn352Ser as a new LDL-C–lowering variant. (**A**) WES results ($n = 6890$). (**B**) WGS results ($n = 1083$). (**C**) Imputed data from genotyping chip results ($n = 5890$). The blue line marks the suggestive threshold ($P = 5.0 \times 10^{-6}$) and the red line the significance threshold ($P = 5.0 \times 10^{-8}$). All P values are based on t test using the additive genetic model.

(n = 826, beta = −29 mg/dl, $P = 1.3 \times 10^{-5}$). We also found that the mean level of AST in ^{352}Ser homozygotes was twofold higher than that of ^{352}Asn homozygotes (35.8 versus 18.4 U/L, respectively, normal range 10 to 35 U/L; n = 6443, recessive model, $P = 5.9 \times 10^{-17}$), further supporting the functional role of this variant.

Although the higher level of AST might originate from liver, we found no association with other liver dysfunction measures, such as alanine aminotransferase (ALT) and the liver to spleen density ratio (a measure of fatty liver), or with alkaline phosphatase or inflammatory markers (table S4). Moreover, ^{352}Ser homozygotes (n = 12) had normal levels of gamma-glutamyl transpeptidase, activated partial thromboplastin time, prothrombin time, and internationalized normalized ratio (table S5). Creatine kinase levels, a marker of muscle damage, was higher, but was not significantly higher and still within the normal range among 12 ^{352}Ser homozygotes (121 U/L) compared with 13 ^{352}Asn homozygotes (86.5 U/L) and 12 heterozygotes (75.5 U/L) (normal range, 44 to 196 U/L for males and 29 to 143 U/L for females), suggesting that the origin of increased AST levels may be from muscle.

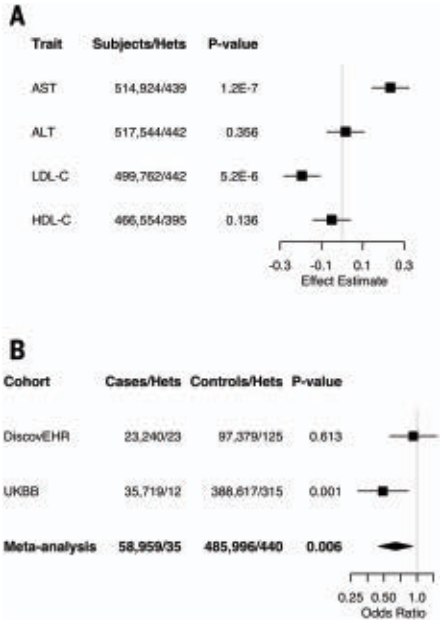


Fig. 2. Gene-based association analyses of rare *B4GALT1* pLOF and predicted deleterious missense variants in DiscovEHR and UK BioBank show CAD protection. (**A**) Association with LDL, HDL, AST, and ALT are similar to OOA results. All P values are based on t test using the additive genetic model. (**B**) Association with CAD protection (n = 544,955). P values are based on the likelihood ratio test using the additive genetic model. Meta-analysis P values are based on Z test.

B4GALT1 is associated with CAD protection

The Amish cohort studied was relatively healthy, with a low prevalence of CAD. Therefore, to test the impact of B4GALT1 on risk of CAD, we leveraged data from two large exome-sequencing datasets (total N > 500 K): Geisinger Health System's DiscovEHR and the UK Biobank (UKBB). Given the rarity of p.Asn352Ser in the general population (MAF < 3.6×10^{-5}), directly testing the variant for association was not possible. Therefore, we focused on aggregating *B4GALT1*-coding variants by creating a multivariant burden test of rare (MAF < 0.01%) predicted loss-of-function (pLOF) or deleterious missense variants, including the small number of p.Asn352Ser carriers, which was then used in analyzing both the DiscovEHR and UKBB data, followed by fixed-effects meta-analysis (*21*). With this burden test, we largely recapitulated the constellation of associations observed for p.Asn352Ser in the OOA, including lower LDL-C [beta = −0.195 (−8 mg/dL),

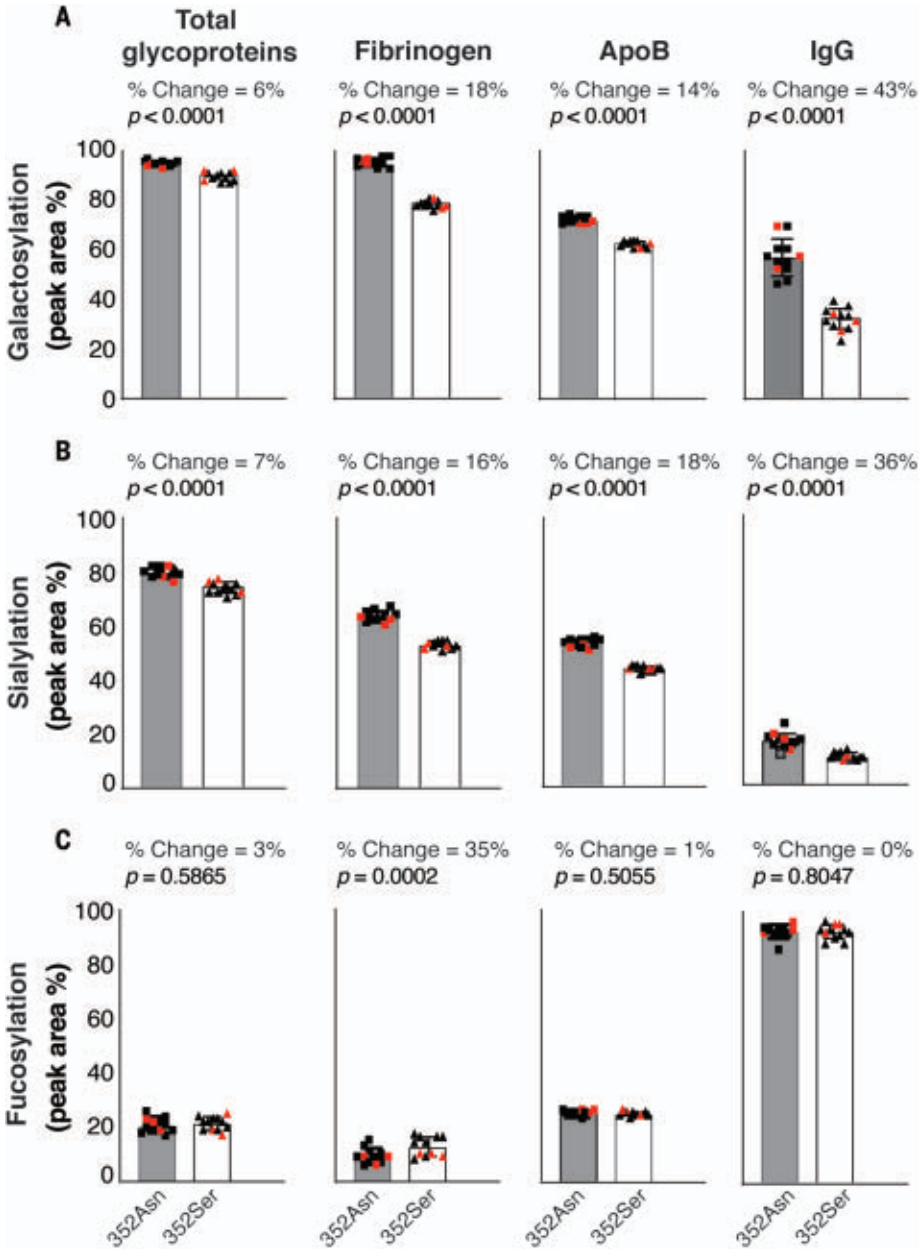


Fig. 3. Human *B4GALT1* ^{352}Ser homozygotes show decreased galactosylation and sialylation. (**A** to **C**) Levels of total galactosylation (A), sialylation (B), and fucosylation (C) in global plasma glycoproteins, fibrinogen, ApoB100, and IgG (mean ± SEM, n = 12 opposite homozygote-matched pairs). P values are based on t test. Red color indicates individuals carrying the known familial hypercholesterolemia p.Arg3527Gln *APOB* variant.

$P = 5.2 \times 10^{-6}$, *t* test] and higher AST [beta = 0.234 (1.7 U/L), $P = 1.2 \times 10^{-7}$, *t* test], but no association with HDL or ALT ($P > 0.05$, *t* test) (Fig. 2A).

We then tested this collection of rare deleterious variants for association with CAD, showing a consistent association with decreased risk (odds ratio = 0.64, $P = 0.006$, *Z* test) (Fig. 2B). The association was protective in both the DiscoverEHR and UKBB data, although the effect was much more significant in the UKBB data ($P = 0.6$ versus 0.001, likelihood ratio test). The rare nature of these variants will require additional large sequence-based cohorts to validate this association, but the results are consistent with the pattern of association for quantitative traits and suggestive of alterations in B4GALT1 function providing protection from CAD.

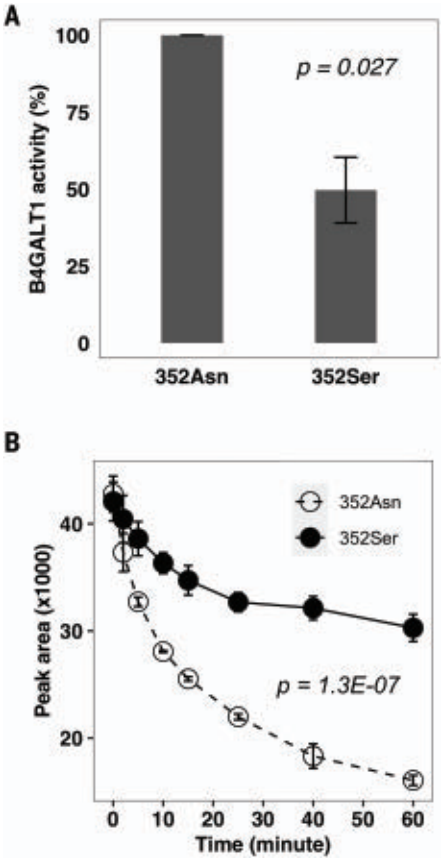


Fig. 4. . *B4GALT1* p.Asn352Ser decreases galactosyltransferase activity. (**A**) Percentage of ^{352}Asn B4GALT1 galactosylation activity of ^{352}Asn B4GALT1 and ^{352}Ser B4GALT1 immunoprecipitated proteins from transfected COS-7 cells (mean ± SEM, $n = 4$). (**B**) Enzymatic activity of recombinant B4GALT1 ^{352}Ser protein (open circles) and ^{352}Asn protein (solid circles). Decrease in substrate UDP-Gal was measured as peak area by hydrophilic interaction chromatography–mass spectroscopy over eight different time points (mean ± SEM, $n = 3$). *P* values are based on *t* test.

B4GALT1 p.Asn352Ser is associated with increased levels of incompletely synthesized glycans on glycoproteins

The carbohydrate-deficient transferrin test (CDT) (*32*), which assesses glycosylation levels of transferrin and apolipoprotein CIII (ApoCIII), is used clinically to diagnose patients with congenital disorders of glycosylation. We performed the CDT in serum samples from 28 participants from the three genotype

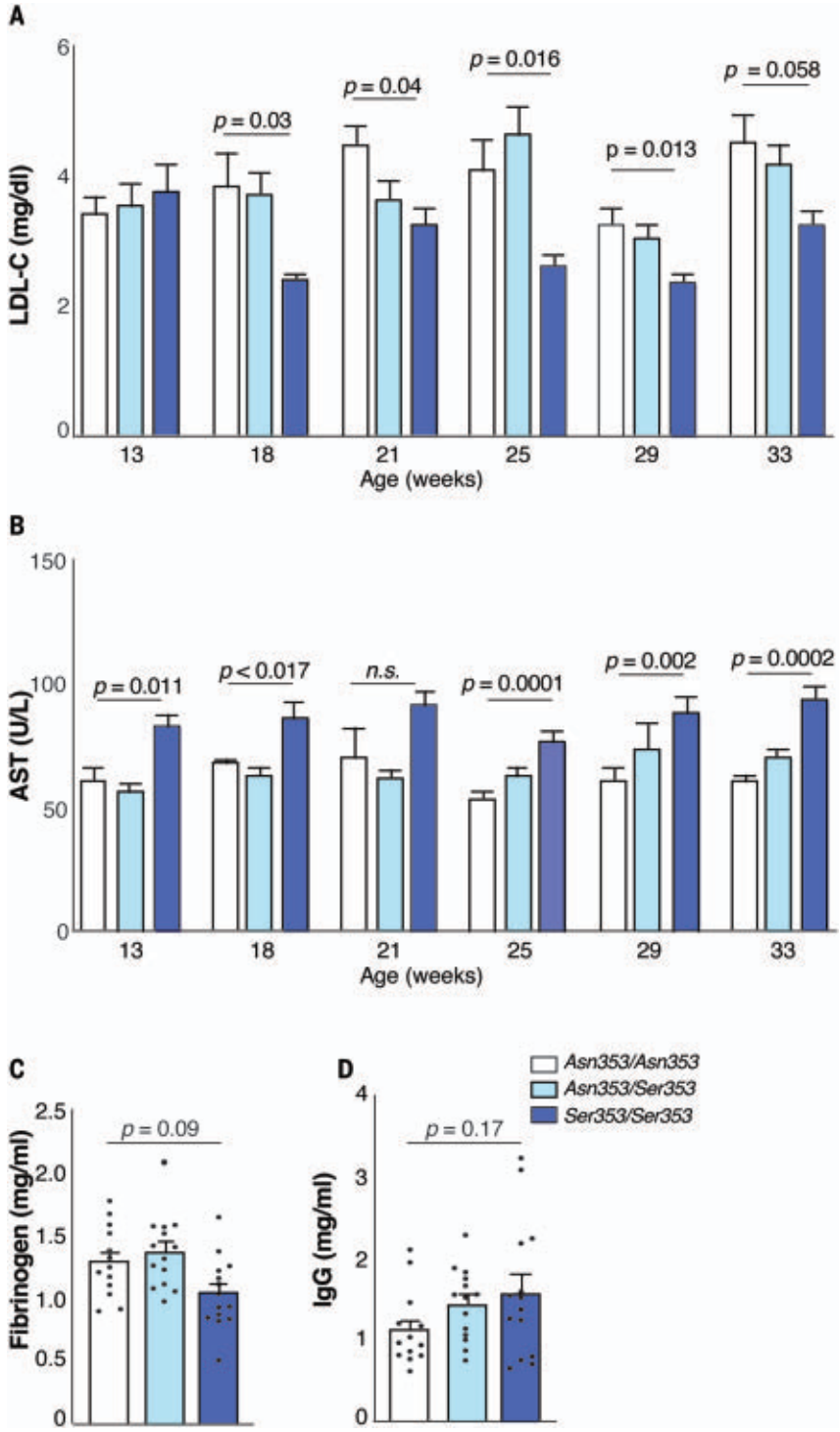


Fig. 5. *B4galt1* p.Asn353Ser germline knock-in mice phenocopy the human phenotype. (**A** and **B**) Heterozygous and homozygous knock-in mice fed a standard chow diet showed lower levels of plasma LDL-C and higher levels of AST compared with wild-type mice. (**C** and **D**) Trends toward decreased fibrinogen and increased IgG were detected in *B4galt1* knock-in mice. *P* values are based on *F* test. Data are represented as mean ± SEM ($n = 14$/group, two biological replicates).

groups (*21*). As expected from B4GALT1's role in processing N-linked carbohydrate, all samples had normal profiles for ApoCIII O-linked carbohydrate (table S6). All samples also had normal mono-oligosaccharide/di-oligosaccharide transferrin and a-oligosaccharide/di-oligosaccharide transferrin ratios. However, whereas all eight ^{352}Asn homozygotes had normal tri-sialo/di-oligosaccharide transferrin ratios, this ratio in all ^{352}Ser homozygotes was abnormal ($P = 9.1 \times 10^{-10}$, *t* test) (fig. S4 and table S6). Because transferrin is mostly tetrasialylated (*33*, *34*), this increase in trisialylated transferrin reflects a paucity of tetrasialylated molecules and indicates that the ^{352}Ser allele is associated with increased levels of carbohydrate-deficient transferrin, likely because of decreased enzymatic activity of B4GALT1.

To determine whether the lower levels of sialylation only affect transferrin or if they also affect other glycoproteins, N-linked glycan profiling was performed using plasma samples from 12 pairs of matched opposite homozygotes for all plasma glycoproteins and selected specific plasma glycoproteins: apolipoprotein B100 (ApoB100), fibrinogen, and immunoglobulin G (IgG) (*21*). There was significantly lower galactosylation and sialylation ($P < 0.0001$, all by *t* test) for global glycoproteins, ApoB100, fibrinogen, and IgG among ^{352}Ser homozygotes compared with ^{352}Asn homozygotes. There was no difference in fucosylation (Fig. 3, A to C), which is consistent with the role of B4GALT1 in adding galactose moieties that then are capped by sialic acid while having no role in the addition of fucose (*35*). Glycans released from all plasma glycoproteins of ^{352}Ser homozygotes had increased levels of incompletely synthesized oligosaccharides, as evidenced by increased percentages of glycans devoid of galactoses and sialic acids (G0F, $P = 5.4 \times 10^{-11}$) and glycans with only one galactose and one sialic acid (G1S1, $P = 5.1 \times 10^{-16}$). Reciprocally, ^{352}Ser homozygotes had significantly lower levels of glycans with two galactose and two sialic acid residues (G2S2, $P = 2.6 \times 10^{-8}$) (table S7). The results for plasma ApoB100, fibrinogen, and IgG were similar in that serum from ^{352}Ser homozygotes had significantly increased levels of glycans deficient in galactose and sialic acid moieties and decreased levels of more mature glycans (tables S8 to S10).

Furthermore, in these same plasma samples, ^{352}Ser homozygotes had lower levels of ApoB100 ($P = 0.0075$) and increased IgG ($P = 0.0004$) compared with ^{352}Asn homozygotes (fig. S5, A and B). Fibrinogen also trended downward ($P = 0.06$) (fig. S5C), consistent with the larger sample size in the initial discovery cohort. In summary, the ^{352}Ser allele is associated with significantly increased levels of incompletely synthesized N-linked glycans on glycoproteins, thus indicating defective protein glycosylation. Altered levels of these proteins may be secondary to the observed defects in protein glycosylation, which could affect protein synthesis, stability, or turnover.

B4GALT1 p.Asn352Ser causes reduced enzymatic activity

To compare the in vitro enzymatic activities of wild-type (^{352}Asn) and mutant (^{352}Ser) B4GALT1 protein, we transiently transfected COS-7 cells, which express very low endogenous levels of B4GALT1, with complementary DNAs (cDNAs) encoding myc-FLAG epitope-tagged versions of the proteins (*21*). We immunoprecipitated with anti-FLAG antibody and measured B4GALT1 activities in the immune complexes (fig. S6). Compared with ^{352}Asn B4GALT1, ^{352}Ser B4GALT1 showed, on average, a 50% decrease in galactosyltransferase specific activity (Fig. 4A).

As a complementary approach, we synthesized recombinant human ^{352}Asn and ^{352}Ser B4GALT1 protein to test the reduction rate of the substrate UDP-Gal incubated with each protein (*21*). Recombinant ^{352}Ser B4GALT1 showed a 2.5-fold decrease of substrate consumption compared with the wild-type protein ($P = 1.3 \times 10^{-7}$, *t* test), indicating a decrease in the enzymatic activity of ^{352}Ser B4GALT1 (Fig. 4B).

B4galt1 Asn353Ser knock-in mouse model phenocopies the human LDL-C and fibrinogen phenotype

Human *B4GALT1* and its mouse ortholog share 92% similarity of their coding sequence, and the ^{352}Asn residue (corresponding to ^{353}Asn in the murine protein sequence) is conserved between the two species (fig. S7). To further validate the findings from the human studies, we generated *B4galt1 353Ser* germline knock-in mice (*21*), which were born at Mendelian frequencies, survived perinatally, and showed normal postnatal development without apparent behavioral or visible physical abnormalities. On a standard chow diet, *B4galt1* ^{353}Ser mice showed significantly lower body weights compared with wild-type mice [average 10%

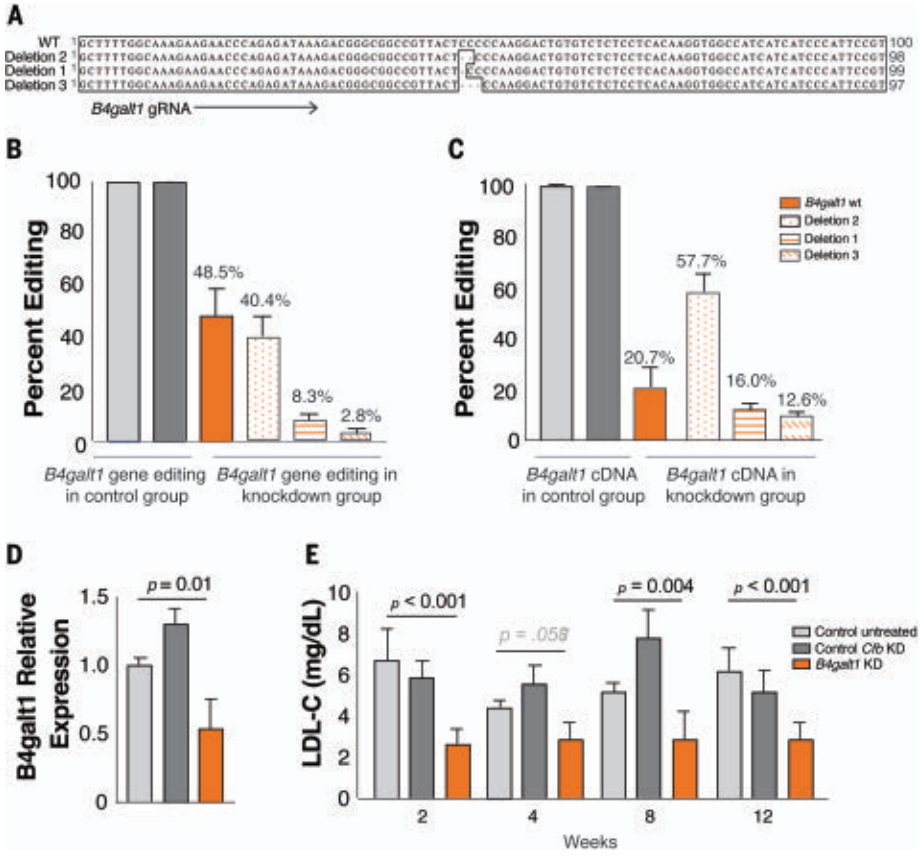


Fig. 6. Liver-specific *B4galt1* knockdown decreases circulating levels of LDL-C in adult mice. (**A** to **C**) Percentage of editing of *B4galt1* genomic [(A) and (B)] and cDNA (C) sequence in liver. The number of reads containing *B4galt1* INDELs was compared with the number of reads with *B4galt1* wild-type sequence. (**D**) *B4galt1* total mRNA levels in liver by TaqMan analysis. The expression of *B4galt1* was calculated relative to *Gapdh*. Values represent the mean of four technical replicates per condition ± SD. (**E**) LDL-C levels were lower in the livers of *B4galt1*-knockdown mice compared with controls. *P* values are based on *F* test. Data are represented as mean ± SD ($n = 6$/group).

decrease in body weight; $0.001 < P < 0.04$ and $0.004 < P < 0.03$ for heterozygotes and homozygotes, respectively; analysis of variance (ANOVA) was used for all animal group comparisons, and P values were calculated based on F test] (fig. S9A). Serum chemistry analysis revealed a 38% decrease of LDL-C in ^{353}Ser/^{353}Ser mice ($0.01 < P < 0.06$; Fig. 5A), and this reduction remained significant at 18 and 25 weeks after adjusting for body weight. Consistent with the human phenotype, an increase of AST, but not ALT, was observed in both ^{353}Asn/^{353}Ser and ^{353}Ser/^{353}Ser mice (14 and 50%, respectively, $P < 0.02$; Fig. 5B and fig. S9F). No significant differences were observed in total cholesterol, ApoB-100, HDL-C, triglycerides, or non-esterified fatty acids (figs. S9, B to E, and S10). Finally, protein analysis of plasma showed a 20% trending decrease of fibrinogen ($P < 0.09$) and a trend toward increased circulating IgG levels in ^{353}Ser/^{353}Ser mice ($P = 0.2$; Fig. 5, C and D).

B4galt1 knockdown in liver of adult mice decreases circulating LDL-C levels

We further investigated whether down-regulation of *B4galt1* in mouse liver could recapitulate the human phenotype. CRISPR-Cas9 was used to knock down endogenous expression of *B4galt1* in livers of adult mice (*21*). Adult mice constitutively expressing Cas9 enzyme were transduced with an AAV8 vector containing guide RNAs (gRNAs) targeting exon 2 of *B4galt1*. This approach resulted in an ~50% gene-editing rate of *B4galt1* in the liver, where edits included deletions of 2 bp (40.4%), 1 bp (8.3%), and 3 bp (2.8%) in exon 2 (Fig. 6, A and B). Further sequencing analysis of the mRNA from the liver revealed that the majority of the persisting transcripts (57.7%) had the 2-bp deletion seen in the genomic DNA, and only 20.7% of the transcripts were wild-type for B4galt1 (Fig. 6C). These results suggested that the knockdown of *B4galt1* in the liver was greater than the 50% suggested by the genomic DNA editing rate. Indeed, TaqMan analysis of B4galt1 in liver showed a 50% decrease in mRNA levels (Fig. 6D), with no changes in *B4galt1* mRNA expression in the spleen (fig. S11). Minimal *B4galt1* gene editing was observed in the spleen (2%), a secondary target of AAV8 (fig. S12). We then measured circulating LDL-C levels starting at 2 weeks from viral transduction and throughout a 12-week study period. An overall 50% decrease of LDL-C was detected throughout the duration of the study ($0.001 < P < 0.06$; Fig. 6E). No significant changes in HDL-C or total cholesterol were observed (fig. S13, A to D). Additionally, there was a trend toward increased circulating AST enzyme activity and no consistent changes in ALT (fig. S13, E and F). These findings were confirmed by two other independent gRNAs designed against exon 2 of *B4galt1* (figs. S14 and S15).

Finally, to ascertain whether the decrease in the LDL-C was mediated specifically by *B4galt1* knockdown, we generated a liver knockdown of complement factor B (*Cfb*). *Cfb* was chosen as control because of its high expression in liver and no known function in LDL-C and cholesterol metabolism or glycosylation. Although we achieved similar levels of liver-specific gene editing (~50% in liver versus <1% in spleen) (fig. S16), LDL-C levels were not affected by *Cfb* liver knockdown (Fig. 6D). These results further corroborate the functional link between *B4galt1* and LDL-C metabolism and suggest that liver-specific modulation of *B4galt1* expression may be a useful approach to reducing LDL-C and perhaps CAD.

Discussion

Large genome-wide analyses of ~600,000 individuals have identified 386 loci associated with lipid traits, none of which identified the *B4GALT1* gene (*36*). The identified p.Asn352Ser variant is conserved across 100 vertebrate species [Genomic Evolutionary Rate Profiling (GERP) score = 5.9] and is located in the flexible long C-terminal region of the protein that undergoes conformational changes to allow for the exchange of the sugar molecule during glycosylation (*37*). Therefore, as supported by our enzymatic assays, a mutation in this region may impede the necessary conformational change and affect glycosylation efficiency. Several studies support a role of protein glycosylation in lipid metabolism (*7*, *38*–*46*) and CVD (*46*, *47*). Our results suggest a previously uncharacterized role of *B4GALT1* in LDL metabolism.

We also observed decreased fibrinogen levels in ^{352}Ser individuals. Although Mendelian randomization studies have not provided evidence for a causal effect of fibrinogen on CVD (*48*), increased fibrinogen can increase CVD risk through multiple pathways (*4*), and a synergistic effect for LDL and fibrinogen on CVD cannot be ruled out.

Our *B4galt1 353Ser* germline knock-in mouse model phenocopied the LDL-C and fibrinogen phenotype identified in humans. Additional investigation of this mouse model and deeper human phenotyping studies will be necessary to better understand the mechanisms underlying these observations.

The p.Asn352Ser *B4GALT1* missense variant does not appear to be associated with any severe phenotype. OOA ^{352}Ser homozygotes, UKBB heterozygotes for pLOF and deleterious variants, and ^{353}Ser *b4galt1* knock-in mice all had increased serum AST levels without elevations in other liver dysfunction biomarkers. Further investigation of the source and consequences of this observation will be required when *B4GALT1* is considered further as a potential therapeutic target.

Previous mass spectrometry studies of N-glycans from plasma and hepatic membrane glycoproteins of *B4galt1*-knockout mice revealed a higher level of sialylation and galactosylation in *B4galt1*$^{-/-}$ and a shift from type 2 to type 1 glycan chains (*49*, *50*), which was likely caused by enhanced activity of B3galt proteins (*51*). Our preliminary investigation showed that the level of type 1 chain (corresponding to the beta-galactose 1,3 linkage) was low and comparable between ^{352}Asn and ^{352}Ser homozygotes (table S11). This observation suggests that our missense mutation did not cause a glycan chain type switch, as was observed in mice (*51*), and that the compensatory action of *B3GALT1* may differ between mouse and human.

We provide human genetic and mouse modeling evidence for an important role of B4GALT1 and protein glycosylation in the regulation of lipid metabolism and fibrinogen levels. Our data suggest that modulation of B4GALT1 expression and/or activity may have pleiotropic effects for cardioprotection. Further understanding of the underlying mechanisms, including potential adverse consequences, may divulge new targets and pathways for the treatment and prevention of CVD.

REFERENCES AND NOTES

1. E. J. Benjamin *et al*., *Circulation* **137**, e67–e492 (2018).
2. R. H. Nelson, *Prim. Care* **40**, 195–211 (2013).
3. T. W. Meade, *J. Clin. Pathol.* **50**, 13–15 (1997).
4. R. Vilar, R. J. Fish, A. Casini, M. Neerman-Arbez, *Haematologica* **105**, 284–296 (2020).
5. L. A. Lange *et al*., *Am. J. Hum. Genet.* **94**, 233–245 (2014).
6. S. Su *et al*., *Atherosclerosis* **200**, 213–220 (2008).
7. I. Surakka *et al*., *Nat. Genet.* **47**, 589–597 (2015).
8. T. J. Hoffmann *et al*., *Nat. Genet.* **50**, 401–413 (2018).
9. P. S. de Vries *et al*., *Hum. Mol. Genet.* **25**, 358–370 (2016).
10. A. Helgadottir *et al*., *Nat. Genet.* **48**, 634–639 (2016).
11. C. Sidore *et al*., *Nat. Genet.* **47**, 1272–1281 (2015).
12. I. Moltke *et al*., *Nature* **512**, 190–193 (2014).
13. R. L. Minster *et al*., *Nat. Genet.* **48**, 1049–1054 (2016).
14. T. I. Pollin *et al*., *Science* **322**, 1702–1705 (2008).
15. H. Shen *et al*., *Arch. Intern. Med.* **170**, 1850–1855 (2010).
16. R. B. Horenstein *et al*., *Arterioscler. Thromb. Vasc. Biol.* **33**, 413–419 (2013).
17. E. Daley *et al*., *J. Bone Miner. Res.* **25**, 247–261 (2010).
18. J. S. Albert *et al*., *N. Engl. J. Med.* **370**, 2307–2315 (2014).
19. F. K. Welty, J. Ordovas, E. J. Schaefer, P. W. Wilson, S. G. Young, *Circulation* **92**, 2036–2040 (1995).
20. M. E. Montasser *et al*., *Circulation* **138**, 1343–1355 (2018).
21. Materials and methods are available as supplementary materials.
22. D. Taliun *et al*., *Nature* **590**, 290–299 (2021).
23. R. Vaser, S. Adusumalli, S. N. Leng, M. Sikic, P. C. Ng, *Nat. Protoc.* **11**, 1–9 (2016).
24. I. Adzhubei, D. M. Jordan, S. R. Sunyaev, *Curr. Protoc. Hum. Genet.* **Chapter 7**, Unit 7.20 (2013).
25. J. H. Sul, B. Han, E. Eskin, *J. Comput. Biol.* **18**, 1611–1624 (2011).
26. J. M. Schwarz, D. N. Cooper, M. Schuelke, D. Seelow, *Nat. Methods* **11**, 361–362 (2014).
27. Y. Choi, A. P. Chan, *Bioinformatics* **31**, 2745–2747 (2015).
28. B. Hansske *et al*., *J. Clin. Invest.* **109**, 725–733 (2002).
29. M. Guillard *et al*., *J. Pediatr.* **159**, 1041–1043.e2 (2011).

30. O. Staretz-Chacham *et al*., *Clin. Genet.* **97**, 920–926 (2020).
31. C. Pérez-Cerdá *et al*., *J. Pediatr.* **183**, 170–177.e1 (2017).
32. O. Iourin *et al*., *Glycoconj. J.* **13**, 1031–1042 (1996).
33. A. Helander, A. Husa, J. O. Jeppsson, *Clin. Chem.* **49**, 1881–1890 (2003).
34. C. Kranz *et al*., *Hum. Mol. Genet.* **16**, 731–741 (2007).
35. G. Lauc *et al*., *PLOS Genet.* **9**, e1003225 (2013).
36. D. Klarin *et al*., *Nat. Genet.* **50**, 1514–1523 (2018).
37. P. K. Qasba, B. Ramakrishnan, E. Boeggeman, *Curr. Drug Targets* **9**, 292–309 (2008).
38. I. Filipovic, G. Schwarzmann, W. Mraz, H. Wiegandt, E. Buddecke, *Eur. J. Biochem.* **93**, 51–55 (1979).
39. R. J. Preston, O. Rawley, E. M. Gleeson, J. S. O'Donnell, *Blood* **121**, 3801–3810 (2013).
40. A. Bowie, D. Owens, P. Collins, A. Johnson, G. H. Tomkin, *Atherosclerosis* **102**, 63–67 (1993).
41. C. Cheng *et al*., *Cancer Cell* **28**, 569–581 (2015).
42. I. Filipovic, *J. Biol. Chem.* **264**, 8815–8820 (1989).
43. M. A. W. van den Boogert *et al*., *J. Inherit. Metab. Dis.* **43**, 611–617 (2020).
44. D. Liu *et al*., *J. Transl. Med.* **16**, 235 (2018).
45. T. M. Teslovich *et al*., *Nature* **466**, 707–713 (2010).
46. P. Nioi *et al*., *N. Engl. J. Med.* **374**, 2131–2141 (2016).
47. C. Menni *et al*., *Circ. Res.* **122**, 1555–1564 (2018).
48. C. K. Ward-Caviness *et al*., *PLOS ONE* **14**, e0216222 (2019).
49. M. Asano *et al*., *EMBO J.* **16**, 1850–1857 (1997).
50. Q. Lu, P. Hasty, B. D. Shur, *Dev. Biol.* **181**, 257–267 (1997).
51. N. Kotani, M. Asano, Y. Iwakura, S. Takasaki, *Biochem. J.* **357**, 827–834 (2001).
52. J. R. O'Connell, "Mixed model analysis for pedigree and population (MMAP)" (GitHub, 2021); https://github.com/MMAP.

ACKNOWLEDGMENTS

We thank the Amish community and the Amish Research Clinic staff. WGS for the TOPMed program was supported by the NHLBI. WGS for "NHLBI TOPMed: Genetics of Cardiometabolic Health in the Amish" (phs000956) was performed at the Broad Institute of MIT and Harvard. 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 no. HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). MMAP (*52*) was used for single-variant genetic analyses. **Funding:** This work was supported in part by NIH grants U01HL137181, U01HL072515, R01AG18728, R01HL121007, HHSN268201500014C, 3R01HL-117626-02S1, 3R01HL-120393-02S1, and P30DK072488 and by American Heart Association grant 17GRNT33661168 and Regeneron Pharmaceuticals, Inc. **Author contributions:** Conceived, designed and supervised the work: A.R.S., A.N.E., M.E.M. Data collection: A.D.H., A.R., A.R.S., B.S., C.H., Ch.H., D.W., E.P., G.D.G., G.T., J.R.O., L.A.L., L.L., L.M., M.E.M., N.L., Q.F., S.A.H., W.L., Y.M., Y.T. Data analysis: A.E.L., B.S., B.Y., C.H., C.V.H., D.W., G.D.G., G.G., G.T., J.A.P., J.R.O., K.A.R., L.A.L., L.L., L.M., M.E.M., M.F., M.H., N.L., N.V., S.A.H., T.D., Y.T. Results interpretation: A.B., A.D.H., A.N.E., A.R.S., B.D.M., B.S., C.S., C.V.H., E.S., G.D.G., J.R.O., M.E.M., M.H., N.L., S.I.T., T.J.D. Manuscript preparation: A.R.S., G.D.G., M.E.M.. **Competing interests:** A.E., A.R., A.R.S., B.S., B.Y., C.H., C.V.H., D.W., E.P., G.D.G., G.G., G.T., L.A.L., L.L., L.M., M.C., M.H., N.L., Q.F., S.A.H., T.J.D., W.L., Y.M., and Y.T. are current or former employees of Regeneron Pharmaceuticals, Inc. A.R.S., C.V.H., G.D.G., M.E.M., and M.H. are inventors on US Patent Number 10,738,284 "B4GALT1 Variants And Uses Thereof" and international patent applications filed in Europe, Canada, New Zealand, Singapore, Israel, India, and Japan. A.R.S. and G.D.G. are inventors on a separate patent application related to this work filed by Regeneron Pharmaceuticals, Inc., US Application No. 17/190,650, on 3 March 2021, which is pending. M.E.M. and B.D.M. receive sponsored research support from Regeneron Pharmaceuticals, Inc. A.R.S. is a part-time faculty member at the University of Maryland School of Medicine, in addition to his employment at Regeneron. **Data availability:** Amish WGS, phenotypes, and covariates are available through dbGaP (phs000956). WES, chip genotypes, imputed data, and glycoprotein data are available to academic investigators through a data use agreement with UMB. DiscovEHR data are available through a data transfer agreement with Regeneron. Regeneron materials (mice and data) described in this manuscript may be available to qualified, academic, noncommercial researchers upon request through the Regeneron portal (https://regeneron.envisionpharma.com/vt_regeneron/). Please send any questions about materials sharing to preclinical.collaborations@regeneron.com. UKBB data are available by application to UK Biobank as described at https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abe0348
Materials and Methods
Figs. S1 to S16
Tables S1 to S11
References (*53*–*75*)
MDAR Reproducibility Checklist

Regeneron Genetics Center Collaborators

Goncalo Abecasis, Aris Baras, Michael Cantor, Giovanni Coppola, Andrew Deubler, Aris Economides, Katia Karalis, Luca A. Lotta, John D. Overton, Jeffrey G. Reid, Alan R. Shuldiner, Nilanjana Banerjee, Dadong Li, Deepika Sharma, Xiaodong Bai, Suganthi Balasubramanian, Andrew Blumenfeld, Gisu Eom, Lukas Habegger, Alicia Hawes, Shareef Khalid, Evan K. Maxwell, William Salerno, Jeffrey C. Staples, Josh Backman, Mathew Barber, Christian Benner, Shan Chen, Amy Damask, Lee Dobbyn, Manuel A. R. Ferreira, Arkopravo Ghosh, Lauren Gurski, Eric Jorgenson, Bindu Kalesan, Jack A. Kosmicki, Hyun Min Kang, Alexander Li, Nan Lin, Daren Liu, Adam E. Locke, Jonathan Marchini, Anthony Marcketta, Joelle Mbatchou, Arden Moscati, Colm O'Dushlaine, Charles Paulding, Jonathan Ross, Eli Stahl, Dylan Sun, Cristopher Van Hout, Kyoko Watanabe, Bin Ye, Andrey Ziyatdinov, Marcus B. Jones, Michelle G. LeBlanc, Jason A. Mighty, Lyndon J. Mitnaul, Ariane Ayer, Kavita Praveen, Regeneron Genetics Center, LLC, Tarrytown, NY 10591, USA.

27 July 2020; resubmitted 25 February 2021
Accepted 19 October 2021
10.1126/science.abe0348

METABOLISM

Fumarate is a terminal electron acceptor in the mammalian electron transport chain

Jessica B. Spinelli[1,2]*, Paul C. Rosen[1,2,3], Hans-Georg Sprenger[1,2], Anna M. Puszynska[1,2], Jessica L. Mann[1,2,3], Julian M. Roessler[1,2,3], Andrew L. Cangelosi[1,2,3], Antonia Henne[1], Kendall J. Condon[1,2,3], Tong Zhang[1,2,3], Tenzin Kunchok[1], Caroline A. Lewis[1], Navdeep S. Chandel[4], David M. Sabatini[3]†

For electrons to continuously enter and flow through the mitochondrial electron transport chain (ETC), they must ultimately land on a terminal electron acceptor (TEA), which is known to be oxygen in mammals. Paradoxically, we find that complex I and dihydroorotate dehydrogenase (DHODH) can still deposit electrons into the ETC when oxygen reduction is impeded. Cells lacking oxygen reduction accumulate ubiquinol, driving the succinate dehydrogenase (SDH) complex in reverse to enable electron deposition onto fumarate. Upon inhibition of oxygen reduction, fumarate reduction sustains DHODH and complex I activities. Mouse tissues display varying capacities to use fumarate as a TEA, most of which net reverse the SDH complex under hypoxia. Thus, we delineate a circuit of electron flow in the mammalian ETC that maintains mitochondrial functions under oxygen limitation.

The flow of electrons through the mitochondrial electron transport chain (ETC) supports a diverse set of cellular processes, such as the synthesis of central metabolites and the regulation of signaling and cell death pathways (*1*–*7*). Electrons enter the ETC through the activities of enzymes such as dihydroorotate dehydrogenase (DHODH) and complex I, move between complexes via the electron carrier ubiquinol (UQH_2), and exit by reducing a terminal electron acceptor (TEA). The canonical view is that in mammalian cells, oxygen (O_2) serves as the sole TEA and its reduction is necessary for the reoxidation of UQH_2 into ubiquinone (UQ) and thus the continuous input of electrons into the ETC. However, under a variety of physiological states, mammalian cells can exist in hypoxic niches (*8*–*16*) while maintaining functions that require the flow of electrons into the ETC, including de novo pyrimidine biosynthesis and oxidation of reduced nicotinamide adenine dinucleotide (NADH) (*17*–*22*). Thus, we sought to clarify the extent to which mitochondrial functions that necessitate electron input into the ETC also require the use of O_2 as a TEA.

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. [2]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA.
***Corresponding author. Email: spinelli@wi.mit.edu**
†David M. Sabatini is no longer affiliated with the Whitehead Institute or the Howard Hughes Medical Institute. To ensure execution of the duties of corresponding author, Jessica B. Spinelli has taken on this role.

Results

In mammalian cells, the flow of electrons into the ETC does not require O_2 reduction

To ask how limitations in O_2 reduction affect mitochondrial functions that depend on the

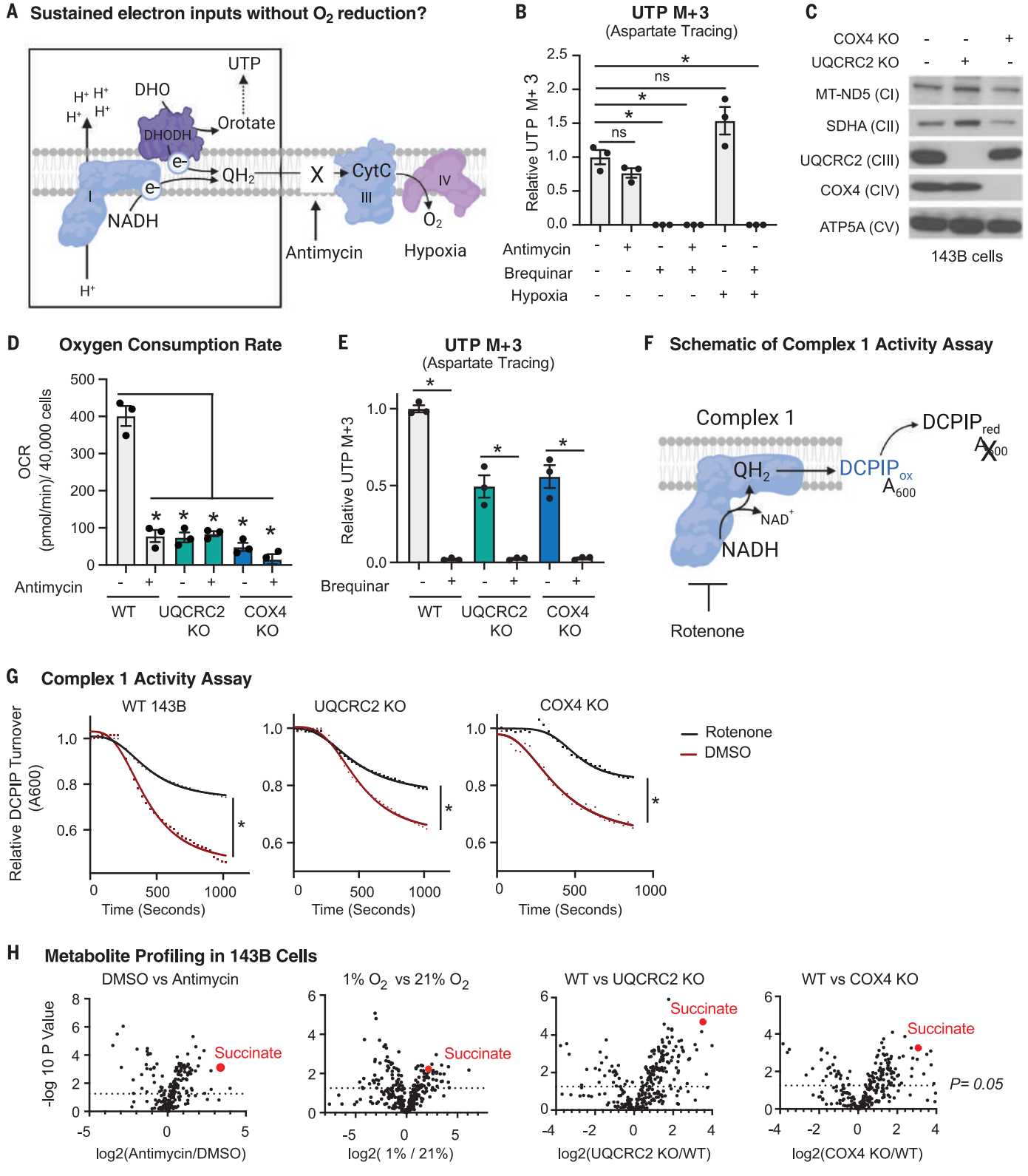


Fig. 1. Cells deficient in O_2 reduction retain the capacity to input electrons into the ETC. (**A**) Schematic depicting the electron transport chain (ETC) and the deposition of electrons (e–) onto a terminal electron acceptor. CytC, cytochrome c. (**B**) DHODH activity as determined by stable isotope tracing with 10 mM $^{13}C_4$-aspartate, which generates $^{13}C_3$-UTP if DHODH is active. Tracing was performed for 8 hours in 143B cells treated with vehicle [dimethyl sulfoxide (DMSO)], 500 nM antimycin, 2 μM brequinar, or in combination or cultured in 1% O_2 for 24 hours (mean ± SEM, n = 3 biological replicates per condition). *P < 0.05. P values were calculated using a two-way analysis of variance (ANOVA). (**C**) Immunoblot analyses of mitochondrial proteins in wild-type (WT), UQCRC2 (complex III) knockout (KO), and COX4 (complex IV) KO 143B cells. (**D**) O_2 consumption rate (OCR) of WT, UQCRC2 KO, and COX4 KO 143B cells treated with DMSO or 500 nM antimycin

for 1 hour (mean ± SEM, $n = 3$ biological replicates per condition). *$P < 0.05$. P values were calculated using a two-way ANOVA. (**E**) DHODH activity as determined using stable isotope tracing of 10 mM $^{13}C_4$- aspartate, which generates $^{13}C_3$-UTP if DHODH is active. Tracing was performed for 8 hours in WT, UQCRC2 KO, and COX4 KO 143B cells treated with DMSO or 2 μM brequinar (mean ± SEM, $n = 3$ biological replicates per condition). *$P < 0.05$. P values were calculated using a parametric t test. (**F**) Schematic depicting the complex I activity assay on purified mitochondria. NADH initiates the reaction, and the absorbance (A_{600}) of the oxidized electron acceptor 2,6-dichlorophenolindophenol (DCPIP) is measured over time. (**G**) Complex I activity in mitochondria purified from WT, UQCRC2 KO, and COX4 KO 143B cells in the presence or absence of 1 μM rotenone (complex I inhibitor). *$P < 0.05$. P values were calculated using an extra sum of squares F test in GraphPad Prism. (**H**) Polar metabolite profiling of 143B cells treated with DMSO versus 500 nM antimycin for 8 hours, grown in 21% versus 1% O_2, WT versus UQCRC2 KO 143B cells, or WT versus COX4 KO 143B cells, $n = 3$ biological replicates per condition. P values were calculated using a parametric t test.

input of electrons into the ETC, we monitored the activity of DHODH. DHODH oxidizes dihydroorotate into orotate and deposits these electrons into the ETC (Fig. 1A). As this reaction is a step in de novo pyrimidine biosynthesis, DHODH activity can be monitored by tracing $^{13}C_4$-aspartate into $^{13}C_3$–uridine 5′-triphosphate (UTP) (fig. S1A). Because ETC inhibition reduces aspartate levels (*23*–*25*), we used excess aspartate as the stable isotope tracer to ensure its availability is not limiting for DHODH activity. Upon treatment of human 143B osteosarcoma cells with antimycin, which inhibits complex III and prevents the transfer of electrons to O_2, $^{13}C_3$-UTP levels were unchanged (Fig. 1B). However, in both vehicle- and antimycin-treated cells, the DHODH inhibitor brequinar ablated $^{13}C_3$-UTP production (Fig. 1B). Similarly, under hypoxia (1% O_2), cells sustained pyrimidine biosynthesis in a DHODH-dependent manner (Fig. 1B). Thus, even under conditions that reduce the transfer of electrons to O_2, DHODH can still deposit electrons into the ETC.

Because complex IV has a high affinity for O_2 and has partial activity even under hypoxia (*26*–*28*), we generated 143B cells that lack a key component of complex IV (COX4) or complex III (UQCRC2), which rendered them incapable of reducing O_2 in the ETC. Loss of these genes did not induce the expression of paralogs, reduce the amounts or assembly of other ETC complexes, or strongly affect mitochondrial DNA (mtDNA) copy number (Fig. 1C and fig. S1, B and C). Although O_2 consumption was greatly reduced in UQCRC2 and COX4 knockout cells (Fig. 1D), de novo pyrimidine biosynthesis was far less impaired. DHODH enzyme activity was unaffected by the loss of UQCRC2 or COX4 in cell lysates (fig. S1D) and dropped ~50% from wild-type levels in live cells, as measured by $^{13}C_4$-aspartate incorporation into $^{13}C_3$-UTP (Fig. 1E). The DHODH inhibitor brequinar ablated $^{13}C_3$-UTP biosynthesis in UQCRC2 and COX4 knockout cells despite having no effect on their O_2 consumption (Fig. 1E and fig. S1, E and F), indicating that DHODH maintains activity independently of the cells' ability to reduce O_2 (Fig. 1E). Notably, the difference in oxygen consumption rate upon brequinar treatment in wild-type cells between our two assays warrants follow-up in future studies (fig. S1, E and F). Furthermore, supplementation of culture media with aspartate, an essential precursor in de novo pyrimidine biosynthesis, increased the proliferation of antimycin-treated cells, which was ablated by the DHODH inhibitor brequinar (fig. S1G). Similarly, brequinar treatment reduced the proliferation of UQCRC2 and COX4 knockout cells (fig. S1H). Thus, both pharmacological and genetic experiments reveal that DHODH maintains electron input into the ETC when O_2 cannot be used as a TEA and that an adaptive mechanism (or mechanisms) must exist to sustain electron flow into the ETC in this context.

Next, we examined complex I activity because it deposits electrons into the ETC during the oxidation of NADH. We purified mitochondria from wild-type, UQCRC2-, and COX4-knockout cells and measured complex I enzymatic activity (*29*) (Fig. 1F). All mitochondria, including those genetically incapable of O_2 reduction, had reduced complex I activity upon rotenone treatment, indicating that complex I enzymatic activity is intact (Fig. 1G). The activity of complex I was slightly lower in mitochondria that lacked UQCRC2 or COX4 (Fig. 1G), consistent with previous findings that hypoxia reduces but does not ablate complex I activity in cells (*30*–*34*). Inhibition of complex I by piericidin did not affect O_2 consumption rate in UQCRC2 and COX4 knockout cells (fig. S1, E and F) but did reduce their proliferation, although to a lesser extent than in wild-type cells (fig. S1I). This is expected from complex I having partial activity upon a block in O_2 reduction. Thus, as with DHODH, complex I can still deposit electrons into the ETC when O_2 cannot be reduced.

Inhibition of O_2 reduction stimulates a net reversal of the succinate dehydrogenase complex, enabling fumarate reduction

Given that DHODH and complex I deposit electrons into the ETC when O_2 reduction is not possible, we sought to determine the fate of these electrons (Fig. 2A). Upon exposure to hypoxia, electrons can be transferred to nicotinamide adenine dinucleotide (NAD^+) through reversal of complex I activity, formally known as reverse electron transfer (*35*, *36*). The combined inhibition of complex III with antimycin and complex I with piericidin had no effect on DHODH activity as measured by the incorporation of $^{13}C_4$-aspartate into $^{13}C_3$-UTP, indicating that an alternative electron removal pathway must sustain nucleotide biosynthesis in the absence of O_2 reduction (fig. S2A). Under hypoxia, lower eukaryotes use fumarate as a TEA, generating succinate as a by-product (*37*). Succinate also accumulates in cancer cells exposed to hypoxia (*38*, *39*), ischemic hearts (*40*), and postexercise muscle (*41*). In 143B cells, we observed an increase in succinate upon hypoxia exposure, antimycin treatment, and depletion of UQCRC2 or COX4 (Fig. 1H).

Although there is agreement that succinate accumulates in mammalian cells under hypoxia, its source is contentious (*42*). Stable isotope tracing studies demonstrate that most of the succinate pool in hypoxic cells derives from α-ketoglutarate through oxidative tricarboxylic acid (TCA) cycle flux (*43*, *44*). However, numerous studies find that a large fraction of the succinate pool comes from fumarate upon a block in O_2 reduction (*38*–*40*, *45*). This reaction is likely catalyzed by the succinate dehydrogenase (SDH) protein complex (complex II), although, given the electrophilicity of fumarate, it is also possible that fumarate is reduced in an unregulated, nonenzymatic fashion.

Fumarate reduction can be monitored in cells by the stable isotope tracing of either $^{13}C_4$-aspartate or $^{13}C_5{}^{15}N_2$-glutamine. $^{13}C_4$-aspartate contributes to the fumarate pool via oxaloacetate and malate and so should lead to the production of $^{13}C_4$-succinate upon fumarate reduction (Fig. 2B). $^{13}C_5{}^{15}N_2$-glutamine contributes to the fumarate pool through the reductive arm of the TCA cycle via glutamate, α-ketoglutarate, isocitrate, citrate, oxaloacetate, and malate (*38*, *39*). A key distinction between the two labeling approaches is that $^{13}C_5{}^{15}N_2$-glutamine enriches the fumarate pool more upon antimycin treatment owing to enhanced reductive carboxylation flux (*38*, *39*), whereas $^{13}C_4$-aspartate labels the fumarate pool to equivalent extents in vehicle- and antimycin-treated cells (Fig. 2C). Therefore, with stable isotope tracing of $^{13}C_4$-aspartate, the amount of $^{13}C_4$-succinate is a direct measure of fumarate reduction, whereas the stable isotope

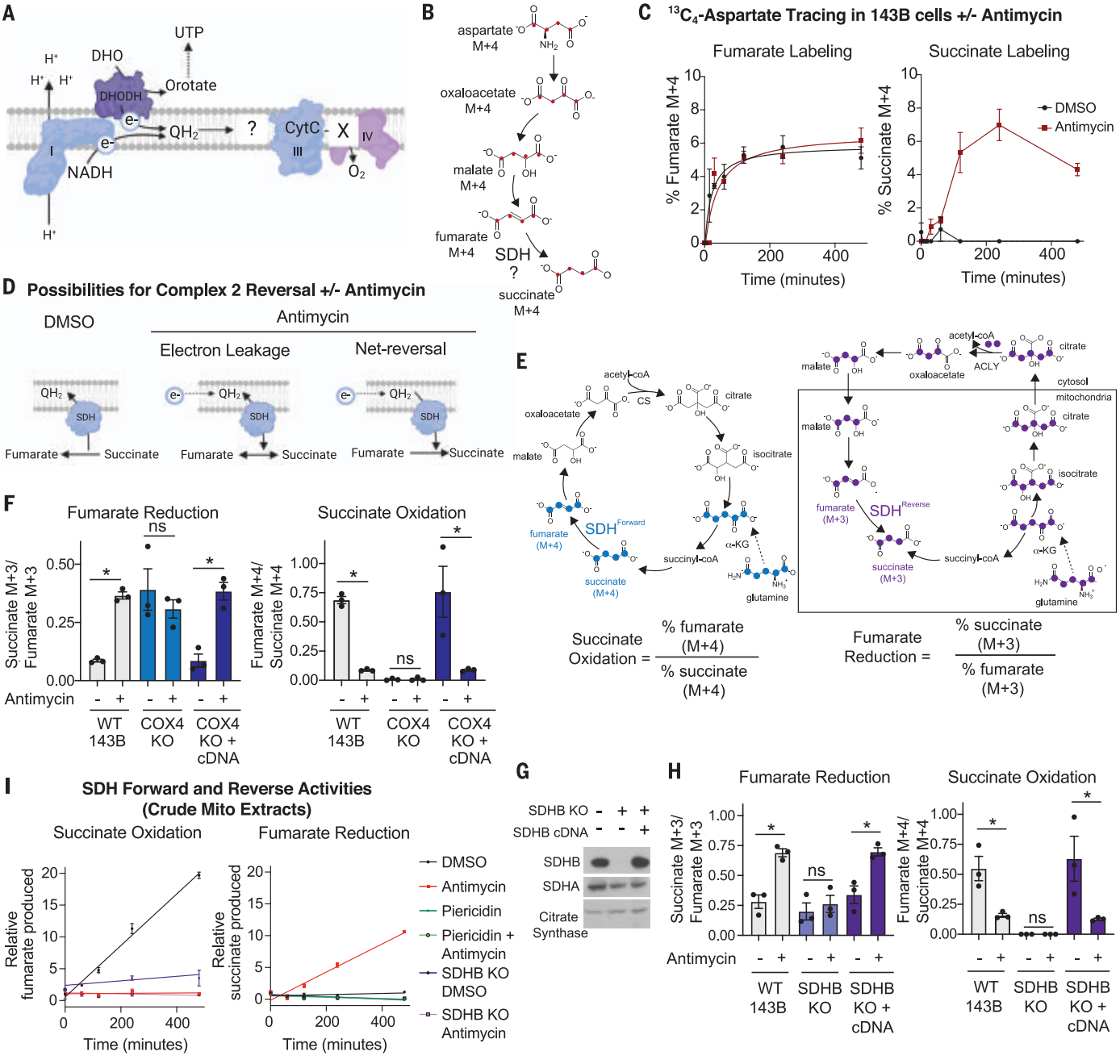


Fig. 2. Upon inhibition of O_2 reduction, fumarate accepts electrons through net reversal of the SDH complex. (**A**) Schematic depicting the question "what is the fate of electrons in the ETC when O_2 cannot be reduced?" (**B**) Schematic showing the expected isotopologues produced during $^{13}C_4$-aspartate tracing if succinate is generated from fumarate. (**C**) Percent labeled fumarate and succinate from a stable isotope tracing experiment using 3 mM $^{13}C_4$-aspartate. WT 143B cells were treated with DMSO or 500 nM antimycin for the indicated times (mean ± SEM, $n = 3$ biological replicates per time point). (**D**) Schematic depicting the reduction of fumarate from either electron leakage onto fumarate or net reversal of SDH upon antimycin treatment. (**E**) Schematic demonstrating the expected isotopologues of TCA cycle metabolites produced during $^{13}C_5{}^{15}N_2$-glutamine tracing. The forward direction of the SDH reaction can be monitored with the ratio of percent labeled fumarate M + 4 to percent labeled succinate M + 4. The reverse direction of the SDH reaction can be monitored with the ratio of percent labeled succinate M + 3 to percent labeled fumarate M + 3. ACLY, ATP citrate lyase; CoA, coenzyme A; CS, citrate synthase; α-KG, α-ketoglutarate. (**F**) Fumarate reduction and succinate oxidation as determined using stable isotope tracing of 2 mM $^{13}C_5{}^{15}N_2$-glutamine. Tracing was performed for 8 hours in WT, COX4 KO, and COX4 KO 143B cells expressing the COX4 cDNA and treated with DMSO or 100 nM antimycin for 8 hours (mean ± SEM, $n = 3$ biological replicates per condition). ns, not significant. (**G**) Immunoblot analyses for indicated proteins in SDHB KO and SDHB cDNA addback 143B cells. (**H**) Fumarate reduction and succinate oxidation as determined using stable isotope tracing of 2 mM $^{13}C_5{}^{15}N_2$-glutamine. Tracing was performed for 8 hours in WT, SDHB KO, and SDHB KO 143B cells expressing the SDHB cDNA and treated with DMSO or 100 nM antimycin for 8 hours (mean ± SEM, $n = 3$ biological replicates per condition). (**I**) SDH activity in purified mitochondria from WT and SDHB KO 143B cells. The succinate oxidation reaction was initiated by adding 10 mM succinate and monitored by the production of fumarate over time. The fumarate reduction reaction was initiated with 10 mM fumarate and 1 mM NADH and monitored through the production of succinate over time; 1 μM antimycin and 1 μM piericidin were included as indicated (mean ± SEM, $n = 3$ biological replicates per time point). Data points were fitted using linear regression. *$P < 0.05$ for all experiments. P values were calculated using an unpaired parametric t test.

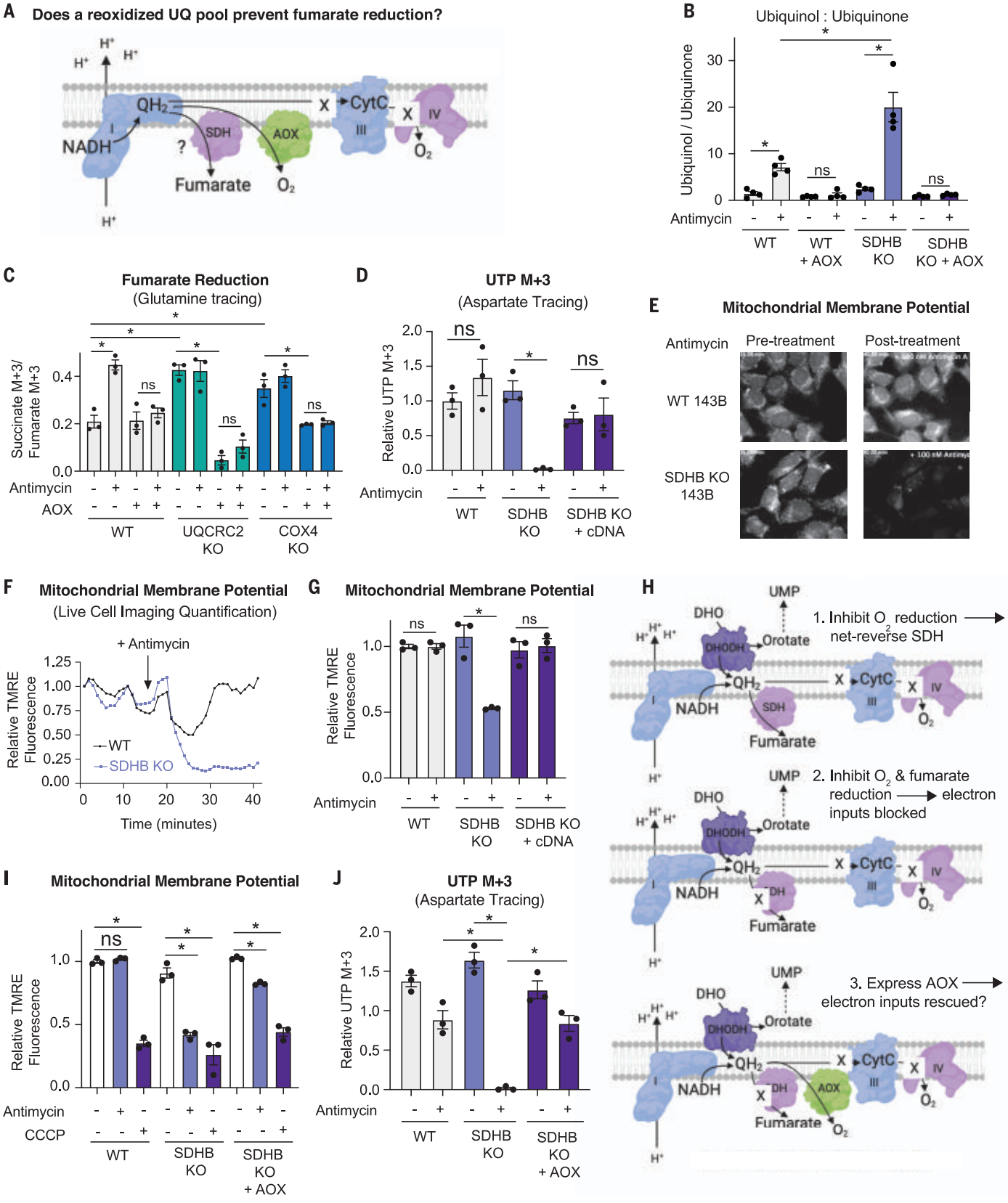


Fig. 3. Fumarate reduction is required to maintain nucleotide biosynthesis and the mitochondrial membrane potential in cells deficient in O_2 reduction. (**A**) Schematic depicting the potential impact of alternative oxidase (AOX) on the accumulation of QH_2 in cells deficient for complex III or IV activity and the consequences for fumarate reduction. (**B**) Ratio of ion counts of ubiquinol to ubiquinone as measured by LC-MS on mitochondria isolated from WT and SDHB KO 143B cells expressing or not expressing AOX and treated with DMSO or 500 nM antimycin for 3 hours (mean ± SEM, *n* = 4 biological replicates per condition).

*P < 0.05. P values were calculated using a two-way ANOVA. (**C**) Relative fumarate reduction as determined using stable isotope tracing of 2 mM $^{13}C_5{}^{15}N_2$-glutamine and the ratio of percent succinate M + 3 to percent fumarate M + 3, representing fumarate reduction in a stable isotope tracing experiment using 2 mM $^{13}C_5{}^{15}N_2$-glutamine. Tracing was performed for 8 hours in WT, UQCRC2 KO, and COX4 KO 143B cells expressing or not expressing AOX and treated with DMSO or 500 nM antimycin (mean ± SEM, n = 3 biological replicates per condition). *P < 0.05. P values were calculated using a two-way ANOVA. (**D**) DHODH activity as measured by stable isotope tracing with 10 mM $^{13}C_4$- aspartate, which generates $^{13}C_3$-UTP if DHODH is active. Tracing was for 8 hours in WT, SDHB KO, and KO 143B cells with the SDHB cDNA expressed and treated with DMSO or 500 nM antimycin (mean ± SEM, n = 3 biological replicates per condition). *P < 0.05. P values were calculated using a parametric t test. (**E**) First and last images from a live-cell imaging video of WT and SDHB KO 143B cells treated with DMSO or 100 nM antimycin. (**F**) Quantification of the mitochondrial membrane potential using live-cell imaging of WT and SDHB KO 143B cells treated with 100 nM antimycin, which was added at the time point indicated with the arrow. (**G**) Mitochondrial membrane potential of WT, SDHB KO, and SDHB KO cells expressing the SDHB cDNA treated with either DMSO or 500 nM antimycin for 1 hour (mean ± SEM, n = 3 biological replicates per condition). *P < 0.05. P values were calculated using a parametric t test. (**H**) Schematic depicting the hypothesis that expression of AOX will rescue complex I and DHODH activities in SDH KO cells treated with antimycin. UMP, uridine 5′-monophosphate. (**I**) Mitochondrial membrane potential in WT, SDHB KO, and SDHB KO 143B cells expressing AOX and treated with DMSO, 500 nM antimycin for 1 hour (mean ± SEM, n = 3 biological replicates per condition). *P < 0.05. P values were calculated using a two-way ANOVA. (**J**) DHODH activity as measured via stable isotope tracing with 10 mM $^{13}C_4$- aspartate, which generates $^{13}C_3$-UTP if DHODH is active. Tracing was performed for 8 hours in WT, SDHB KO, and SDHB KO 143B cells expressing AOX and treated with DMSO or 500 nM antimycin (mean ± SEM, n = 3 per biological replicates condition). *P < 0.05. P values were calculated using a two-way ANOVA.

tracing of $^{13}C_5{}^{15}N_2$-glutamine requires a ratiometric analysis of labeled succinate to labeled fumarate to normalize for differences in the extent of fumarate labeling.

Antimycin robustly stimulated the conversion of fumarate into succinate, as monitored by the production of $^{13}C_4$-succinate over time and the ratio of percent $^{13}C_4$-succinate to percent $^{13}C_4$-fumarate when $^{13}C_4$-aspartate labeling was in the steady state (Fig. 2C and fig. S2, B to D). Given this, we reasoned that even if—as has been argued (*43*, *44*)—fumarate is not the major source of succinate accumulation in hypoxic cells, it could still serve as a TEA when O_2 reduction is limiting. However, there could be two explanations for the increase in fumarate reduction caused by hypoxia and antimycin treatment: minor electron leakage onto the electrophilic molecule fumarate, similar to how electrons leak out of complexes I and III nonspecifically to generate reactive oxygen species (ROS), or net reversal of the SDH complex, in which the rate of fumarate reduction exceeds that of succinate oxidation (Fig. 2D). In both cases, hypoxia would stimulate fumarate to succinate conversion, but only net reversal would enable efficient reoxidation of UQH_2 and thereby sustain the input of electrons into the ETC from complex I and DHODH.

To formally test whether hypoxia triggers net reversal of the SDH complex, we used the $^{13}C_5{}^{15}N_2$-glutamine tracer, which enables simultaneous quantification of the forward and reverse activities of SDH through the generation of specific isotopologues upon flux through the oxidative arm (α-ketoglutarate to succinate) rather than through the reductive arm (α-ketoglutarate, through citrate, to fumarate) of the TCA cycle (Fig. 2E). In this assay, the forward (succinate oxidation) reaction is the ratio of percent labeled $^{13}C_4$-fumarate to its precursor $^{13}C_4$-succinate (Fig. 2E), and the reverse (fumarate reduction) reaction is the ratio of percent labeled $^{13}C_3$-succinate to its precursor $^{13}C_3$-fumarate (Fig. 2E). This analysis was performed when the labeling is in the steady state after 4 to 8 hours of incubation with $^{13}C_5{}^{15}N_2$-glutamine (fig. S2, E and F). Notably, this ratiometric analysis of the succinate and fumarate isotopologues eliminates biases caused by higher $^{13}C_3$-fumarate labeling in antimycin-treated cells (fig. S2E).

Inhibition of O_2 reduction by antimycin or hypoxia decreased succinate oxidation and increased fumarate reduction, which are the SDH forward and reverse activities, respectively (Fig. 2F and fig. S2G). Because the ratio of isotopologues representing fumarate reduction exceeded those for succinate oxidation by approximately fourfold, we conclude that antimycin treatment and hypoxia exposure cause higher levels of fumarate reduction than succinate oxidation. Likewise, the UQCRC2 and COX4 knockout cells had approximately six- and eightfold higher levels of fumarate reduction than succinate oxidation, respectively, whereas the opposite was true in the control cells (Fig. 2F and fig. S3, A to E). Expression in the knockout cells of the UQCRC2 or COX4 cDNA rescued O_2 consumption and restored succinate oxidation and fumarate reduction reactions and their sensitivity to antimycin treatment to close to wild-type levels (Fig. 2F and fig. S3, A to E). In cells lacking either the SDHA or SDHB component of the SDH complex, the fumarate reduction and succinate oxidation reactions were not altered by antimycin treatment (Fig. 2, G and H, and fig. S3, F and G). Expression in the knockout cells of the respective cDNAs restored the increase in fumarate reduction and decrease in succinate oxidation caused by antimycin treatment (Fig. 2, G and H, and fig. S3, F and G). Similarly, the complex II inhibitor malonic acid almost completely ablated fumarate reduction in UQCRC2 and COX4 knockout cells and had no effect on SDHB knockout cells (fig. S3, H and I). Dimethyl succinate treatment slightly suppressed antimycin-induced fumarate reduction, which is consistent with high succinate levels inhibiting SDH activity (fig. S3J). Taken together, these data demonstrate that the SDH complex catalyzes more fumarate reduction than succinate oxidation when electrons cannot be transferred to O_2, suggesting a net reversal of its activity.

In the untreated wild-type, SDHA, and SDHB knockout cells, we detected similar background levels of fumarate reduction, which are likely caused by the nonenzymatic reduction of fumarate into succinate, consistent with its electrophilic nature. To test whether electron leakage out of the ETC contributes to background fumarate reduction, we treated 143B cells with the mitochondrial-targeted antioxidant MitoTEMPO in the presence or absence of antimycin (fig. S3K). MitoTEMPO caused a significant decrease in fumarate reduction in vehicle-treated cells but did not affect fumarate reduction in antimycin-treated cells (fig. S3K), suggesting that electron leakage may contribute to baseline levels of fumarate reduction.

Consistent with results in 143B cells, antimycin treatment increased fumarate reduction in a panel of human cancer cell lines (SW1353, U87, DLD1, and HCT116), in the mouse myoblast cell line C2C12, in primary dermal fibroblasts, and in mink lung epithelial cells (fig. S4, A to D). These data generalize the conclusion that inhibition of O_2 reduction leads to a rewiring of electron flow in the ETC to enable fumarate reduction.

To ask whether net reversal can also occur in a cell-free system, we examined SDH activity in permeabilized purified mitochondria. The rate of succinate oxidation was measured by monitoring fumarate production over time after initiating the reaction with succinate. The rate of fumarate reduction was measured in a separate assay by monitoring the rate of succinate production over time after initiating the reaction with fumarate and NADH. Consistent with the stable isotope tracing experiments in live cells, vehicle-treated mitochondria

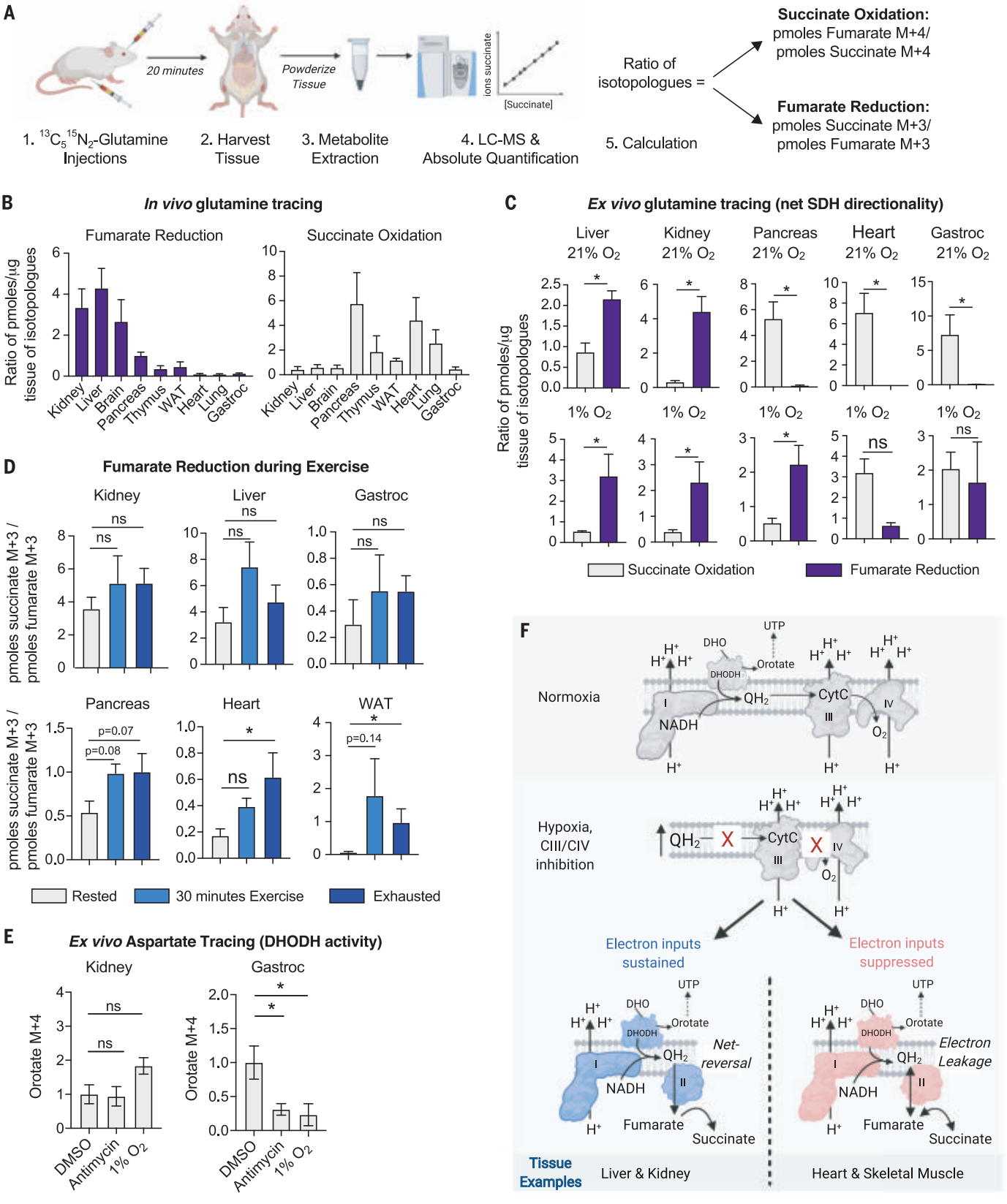


Fig. 4. Fumarate reduction supports mitochondrial functions in tissues capable of net reversal of the SDH reaction. (**A**) Depiction of the workflow for the in vivo $^{13}C_5{}^{15}N_2$-glutamine stable isotope tracing experiment to measure the net directionality of the SDH complex in tissues. (**B**) In vivo stable isotope tracing of $^{13}C_5{}^{15}N_2$-glutamine in indicated tissues. Mice were euthanized 20 min after retroorbital and intraperitoneal injections. Succinate oxidation was calculated as the ratio of picomoles fumarate M + 4 to picomoles succinate M + 4. Fumarate reduction was calculated as the ratio of picomoles succinate

M + 3 to picomoles fumarate M + 3 (mean ± SEM, n = 4 per condition). (**C**) Tissue-autonomous succinate oxidation or fumarate reduction as determined with ex vivo 2 mM $^{13}C_5{}^{15}N_2$-glutamine stable isotope tracing for 24 hours in indicated tissues kept in a tissue culture incubator at 21% O_2 or a hypoxia incubator (1% O_2). Succinate oxidation and fumarate reduction were calculated as described in (B) (mean ± SEM, n = 4 per condition). *P < 0.05. P values were calculated using a parametric t test. (**D**) In vivo $^{13}C_5{}^{15}N_2$-glutamine tracing in female mice 12 weeks old through intraperitoneal and intramuscular injections. Mice were rested, exercised for 30 min, or exercised until exhaustion for ~1.5 hours and then injected with $^{13}C_5{}^{15}N_2$-glutamine. The rested mice were euthanized 15 min after injection with no exercise, and the exercised mice continued to run on the treadmill for 15 min before being euthanized. Tissues were harvested for metabolite isolation and mass spectrometry. Absolute quantification was performed to calculate the concentration of succinate M + 3, succinate M + 4, fumarate M + 3, and fumarate M + 4 in picomoles per microgram tissue protein. The reported ratio representing fumarate reduction was calculated by the picomoles succinate M + 3 per microgram tissue protein to the picomoles fumarate M + 3 per microgram tissue protein. The reported ratio representing succinate oxidation was calculated by the picomoles fumarate M + 4 per microgram tissue protein to the picomoles succinate M + 4 per microgram tissue protein. Data represent the mean ± SEM, n = 5 mice per time point. *P < 0.05. P values were calculated using a two-way ANOVA. (**E**) Ex vivo 3 mM $^{13}C_4$-aspartate stable isotope tracing for 16 hours in indicated tissues kept in an incubator at 21 or 1% O_2 or treated with 2 μM antimycin. Orotate M + 4 levels reflect DHODH activity (mean ± SEM, n = 4 biological replicates per condition). *P < 0.05. P values were calculated using a two-way ANOVA. (**F**) Model in which net reversal of SDH supports certain mitochondrial functions in tissues under conditions that reduce electron transfer to O_2.

had a higher rate of succinate oxidation than fumarate reduction, whereas the opposite was the case in antimycin-treated mitochondria (Fig. 2I). Mitochondria that lacked the SDH complex did not exhibit any succinate oxidation or fumarate reduction, and complex I inhibition by piericidin suppressed the fumarate reduction caused by antimycin treatment (Fig. 2I), demonstrating that electrons enter the ETC from complex I and exit onto fumarate via SDH.

Ubiquinol accumulation is required for SDH reversal

Our data so far suggest that upon inhibition of O_2 reduction, the reduced electron carrier UQH_2 transfers electrons to fumarate, a reversal of the normal reaction catalyzed by SDH in which succinate deposits electrons onto the oxidized electron carrier UQ. Net reversal of the mammalian SDH complex has been considered thermodynamically unfavorable because the standard reduction potential of UQ is slightly greater than that of fumarate (*46*). Moreover, unlike lower eukaryotes, mammals do not appear to have a distinct electron carrier with a lower reduction potential that could facilitate fumarate reduction (*37*). Because the reduction potential of UQ and fumarate are very close to each other (~10 mV apart), we considered the possibility that UQH_2 accumulation drives the net reversal of the SDH complex in mammalian cells upon suppression of O_2 reduction (Fig. 3A). To test this, we took advantage of the enzyme alternative oxidase (AOX), which oxidizes UQH_2 to UQ in an antimycin-insensitive manner (*47*, *48*). If UQH_2 accumulation is necessary to drive SDH in reverse, AOX expression should prevent net reversal of SDH when O_2 reduction is blocked by maintaining an oxidized UQ pool (Fig. 3A).

We targeted AOX to the mitochondrial inner membrane (fig. S5A), where, consistent with its role in maintaining an oxidized UQ pool, it blunted NADH and UQH_2 accumulation upon antimycin treatment (Fig. 3B and fig. S5B). Notably, the levels of UQH_2 and the ratio of UQH_2 to UQ were greater in antimycin-treated SDHB knockout cells than in antimycin-treated wild-type cells, which is consistent with the SDH complex playing a critical role in UQH_2 reoxidation upon antimycin treatment (Fig. 3B). Expression of AOX in SDHB knockout cells blunted the accumulation of UQH_2 upon antimycin treatment (Fig. 3B and fig. S5, C and D). Similarly, expression of AOX in UQCRC2 and COX4 knockout cells reduced the UQH_2/UQ ratio and the levels of UQH_2 (fig. S5, E to I).

To determine whether UQH_2 accumulation is required to reverse the SDH complex, we used stable isotope tracing of both $^{13}C_5{}^{15}N_2$-glutamine and $^{13}C_4$-aspartate to measure fumarate reduction in AOX-expressing wild-type cells as well as those lacking UQCRC2 or COX4 (Fig. 3C and fig. S6, A and B). Consistent with the idea that UQH_2 accumulation is required to reverse SDH upon inhibition of O_2 reduction, AOX expression fully suppressed the increase in fumarate reduction caused by antimycin treatment in wild-type cells (Fig. 3C and fig. S6, A and B). Similarly, AOX expression in the UQCRC2 and COX4 knockout cells also suppressed fumarate reduction and almost completely restored succinate oxidation to wild-type levels (Fig. 3C and fig. S6, A and B).

We corroborated the impact of AOX expression on SDH directionality by using permeabilized purified mitochondria from AOX-expressing 143B cells treated with vehicle or antimycin. As before (Fig. 2I), we monitored the SDH forward (succinate oxidation) and reverse (fumarate reduction) reactions over time. Consistent with the stable isotope tracing results in live cells, AOX expression prevented fumarate reduction upon antimycin treatment (fig. S6C). Taken together, these data demonstrate that when O_2 reduction is blocked, UQH_2 accumulation is required for SDH reversal.

Fumarate reduction sustains electron inputs into the ETC when O_2 reduction is suppressed

To understand the potential importance of fumarate reduction, we asked whether it is required for cells that are incapable of using O_2 as a TEA to sustain mitochondrial functions, such as de novo pyrimidine biosynthesis through DHODH, which requires the deposition of electrons into the ETC. If fumarate reduction sustains DHODH activity upon inhibition of O_2 reduction, we expect that simultaneous loss of both TEAs—O_2 and fumarate—will suppress this reaction. Indeed, antimycin ablated DHODH activity (as read out by $^{13}C_3$-UTP production) in SDHB-deficient cells but had no effect in wild-type cells, knockout cells complemented with the SDHB cDNA, or knockout cells complemented with a class 1 DHODH that directly deposits electrons on fumarate as opposed to UQ (Fig. 3D and fig. S6D).

Next, we tested whether fumarate reduction sustains complex I activity in cells incapable of using O_2 as a TEA. To do so, we measured both the NAD^+/NADH ratio, which is an indicator of complex I–mediated NADH reoxidation, and the mitochondrial membrane potential ($\Delta\Psi^{Mito}$), to which complex I contributes via proton pumping. As with DHODH activity, we reasoned that simultaneous inhibition of fumarate and O_2 reduction would reduce complex I activity, causing a decrease in the NAD^+/NADH ratio and depolarization of the $\Delta\Psi^{Mito}$. Consistent with this idea, SDHB-null cells treated with antimycin, which cannot use fumarate or O_2 as a TEA, had a lower NAD^+/NADH ratio than wild-type cells treated with antimycin (fig. S6, E to G). Moreover, the $\Delta\Psi^{Mito}$, which we monitored using the fluorescent dye tetramethylrhodamine ethyl ester (TMRE), was substantially more depolarized in SDHB-null cells after 30 min of antimycin treatment than in wild-type cells (Fig. 3, E and F), in a fashion complemented by the SDHB cDNA (Fig. 3G). Notably, treatment with 250 nM carbonyl cyanide 3-chlorophenylhydrazone (CCCP), which specifically uncouples the $\Delta\Psi^{Mito}$ without affecting the plasma membrane potential, reduced fluorescence, indicating that the TMRE dye was not in quench mode (fig. S6H) (*49*). Upon antimycin treatment, wild-type cells displayed an initial reduction

in the $\Delta\Psi^{Mito}$, which was restored to pretreatment levels with a similar timing as it takes for antimycin to increase fumarate reduction (Fig. 3F and fig. S6I). The effects of antimycin treatment on $\Delta\Psi^{Mito}$ were further corroborated in the contexts of pharmacologic and genetic suppression of SDHA activity (fig. S6, J and K), and notably, antimycin treatment did not alter mtDNA copy number in the SDHA and SDHB knockout cells (fig. S6, L and M). Thus, fumarate reduction supports partial complex I activity, specifically NADH reoxidation and $\Delta\Psi^{Mito}$, in cells incapable of reducing O_2 in the ETC.

Thus far, our data establish that the SDH complex is required to maintain DHODH and complex I activities when O_2 reduction is impeded. To determine whether these defects are caused by a lack of electron removal from the ETC onto fumarate and thus an inability to reoxidize UQH_2 (*47*), we tested whether expression of AOX could restore DHODH activity and the $\Delta\Psi^{Mito}$ in cells that are unable to use both fumarate and O_2 as TEAs (Fig. 3H). To do so, we expressed AOX in SDHB knockout cells and measured $\Delta\Psi^{Mito}$ and DHODH activity upon antimycin treatment. Indeed, expression of AOX in SDHB knockout cells almost completely prevented depolarization of $\Delta\Psi^{Mito}$ and the reduction in DHODH activity upon antimycin treatment (Fig. 3, I and J). Thus, we conclude that simultaneous loss of electron transfer to fumarate and O_2 ablates mitochondrial functions that are dependent on electron input into the ETC and that these functions can be restored by reoxidation of UQH_2 to UQ.

An expected consequence of complete loss of electron flow in the ETC is reduced proliferation rate because the ETC supports many biosynthetic pathways, including pyrimidines. Therefore, so long as there are sufficient nutrients to supply precursors for these biosynthetic pathways, we expect that fumarate reduction will support proliferation in cells incapable of using O_2 as a TEA. Consistent with this idea, treatment of SDHB-null cells treated with antimycin in media containing high pyruvate and aspartate reduced proliferation more than wild-type cells (fig. S6N). Similarly, treatment of UQCRC2 and COX4 knockout cells with the complex II inhibitor malonic acid significantly reduced their proliferation (fig. S6, O and P). Thus, in the context of sufficient metabolic precursors, fumarate reduction can support the proliferation of cells incapable of using O_2 as a TEA.

Fumarate is a terminal electron acceptor in mouse tissues

To understand the physiological relevance of fumarate reduction, we traced $^{13}C_5{}^{15}N_2$-glutamine in 143B cells cultured at O_2 concentrations that fall within the 1 to 15% range observed in tissues in vivo (*8*). Notably, although fumarate reduction was undetectable at 20% O_2, even a decrease to 15% was sufficient to stimulate fumarate reduction, which continued to increase as O_2 concentrations were lowered until it reached a maximum at 3% O_2 (fig. S7A).

To investigate fumarate reduction in vivo, we applied the $^{13}C_5{}^{15}N_2$-glutamine tracing technique to mouse tissues. Mice were injected with $^{13}C_5{}^{15}N_2$-glutamine, followed by absolute quantification of the succinate and fumarate isotopologues in tissues. SDH activities were quantified by calculating the ratio of the picomoles of M + 4 isotopologues, representing succinate oxidation, and of M + 3 isotopologues, representing fumarate reduction (Fig. 4A). Notably, because these in vivo tracing experiments were performed with a bolus injection, the labeling was not in the steady state, and therefore, the forward and reverse SDH activities in a given tissue could not be compared with one another (*50*).

The lung, heart, pancreas, thymus, white adipose tissue (WAT), and gastrocnemius muscle catalyzed little to no detectable fumarate reduction and high levels of succinate oxidation, whereas the kidney, liver, and brain appeared to catalyze high levels of fumarate reduction (Fig. 4B), suggesting that tissues may have differing capacities to utilize fumarate as a TEA at physiological O_2 concentrations. The fumarate reduction and succinate oxidation reactions among mouse tissues did not correlate with their total levels or ratios of succinate and fumarate, nor their ratio of UQH_2 to UQ (fig. S7, B to D). Notably, adenosine 5′-triphosphate (ATP) citrate lyase, an enzyme required for reductive carboxylation, was low in the heart and high in the liver (fig. S7E). Although this positively correlates with their capacity to do fumarate reduction, this correlation did not extend to other tissues. Moreover, all tissues sufficiently enriched the $^{13}C_3$-fumarate pool upon injection with $^{13}C_5{}^{15}N_2$-glutamine, and across three different time points (10, 20, and 30 min) after injection, the ratio of isotopologues representing fumarate reduction and succinate oxidation remained similar, corroborating that differences exist among tissues in their ability to reduce fumarate at physiological O_2 concentrations (fig. S8).

To determine whether the observed labeling in vivo is tissue autonomous and not from interorgan transfer of labeled metabolites (*51*), we performed ex vivo $^{13}C_5{}^{15}N_2$-glutamine tracing on mouse tissues cultured in incubators set to either 21 or 1% O_2 (Fig. 4C and fig. S9, C and D). Notably, ex vivo $^{13}C_5{}^{15}N_2$-glutamine tracing enables steady-state labeling of the fumarate and succinate isotopologues after 16 hours (fig. S9, A and B), allowing us to directly compare forward and reverse SDH activities in each tissue. Similar to the retinas cultured ex vivo (*45*), even when cultured in atmospheric O_2, the liver, kidney, and brain displayed higher levels of fumarate reduction than succinate oxidation, indicating that the SDH complex is intrinsically operating in reverse in these tissues (Fig. 4C and fig. S9, C and D). Similar to cultured cells, the pancreas, thymus, lung, WAT, heart, and gastrocnemius muscle all favor the succinate oxidation SDH activity over the fumarate reduction SDH activity when cultured in atmospheric O_2 (Fig. 4C and fig. S9, C and D). Upon hypoxia exposure, all of these tissues undertook some level of fumarate reduction, but only a subset, including the liver, kidney, brain, pancreas, WAT, thymus, and lung, exhibited net reversal of SDH, in which fumarate reduction was greater than succinate oxidation (Fig. 4C and fig. S9, C and D). The heart and gastrocnemius muscle modestly increased fumarate reduction when exposed to hypoxia but did not net reverse the SDH complex (Fig. 4C and fig. S9, C and D).

Our results were corroborated with $^{13}C_4$-aspartate tracing on mouse tissues cultured ex vivo in 21 and 1% O_2, in which the production of $^{13}C_4$-succinate was used as a proxy for fumarate reduction when labeling was in the steady state (fig. S10, A to C). Consistent with the apparently constitutive fumarate reduction in the brain, liver, and kidney and a lack of fumarate reduction in the heart and gastrocnemius muscle, we detected abundant $^{13}C_4$-succinate labeling in the former, but not the latter, tissues cultured in atmospheric O_2 (fig. S10B). Upon exposure to hypoxia, $^{13}C_4$-succinate increased in all tissues (fig. S10C). Notably, incorporation of $^{13}C_4$-aspartate into $^{13}C_2$-succinate occurred in all tissues except the liver, kidney, and lung (fig. S10B), which may be partially driven by differences in citrate synthase levels and oxidative TCA cycle flux among tissues. Taken together, these data confirm—by using an orthogonal tracing approach—that hypoxia induces fumarate reduction in mouse tissues.

We next tested whether physiological perturbations that reduce tissue O_2 concentrations can likewise lead to an increase in fumarate reduction. Exercise causes tissue hypoxia, and it has been observed that succinate levels increase in exercising humans and mice (*41*, *51*). To test whether exercise causes an increase in fumarate reduction, we challenged mice to a short (30 min) or long (90 min) exercise regimen and then injected them with $^{13}C_5{}^{15}N_2$-glutamine and monitored the $^{13}C_3$-succinate and $^{13}C_3$-fumarate isotopologues in the kidney, liver, pancreas, gastrocnemius, heart, and WAT (Fig. 4D and fig. S11). Upon exercise challenge, $^{13}C_3$-succinate significantly increased in the gastrocnemius, heart, and WAT, but not in the kidney, liver, or pancreas (fig. S11). To determine whether this increase in $^{13}C_3$-succinate is

driven by an increase in fumarate reduction, we calculated the ratio of the absolute concentration of $^{13}C_3$-succinate to that of $^{13}C_3$-fumarate, revealing that the heart and WAT, but not the kidney, liver, pancreas, and gastrocnemius, increase fumarate reduction upon exercise (Fig. 4D). Notably, labeling is not in the steady state in this experiment, and therefore, the net directionality of the SDH complex upon exercise challenge cannot be determined by using this protocol. Nevertheless, these data clearly demonstrate that fumarate reduction increases in a subset of tissues upon exercise challenge.

Because net reversal of the SDH complex supports DHODH activity in cultured cells when O_2 reduction is blocked (Fig. 3), we tested whether this was also true in mouse tissues. To do so, we measured DHODH activity in tissues that are capable (the liver and kidney) or incapable (the heart and gastrocnemius muscle) of net reversing the SDH complex when O_2 reduction is limited. Antimycin treatment ablated O_2 consumption in the liver, kidney, heart, and gastrocnemius muscle ex vivo, but only the liver and kidney maintained net reversal of the SDH complex (fig. S12, A and B). DHODH activity was assessed via $^{13}C_4$-aspartate incorporation into $^{13}C_4$-orotate because labeling of the $^{13}C_3$-UTP pool was undetectable in most tissues. Antimycin treatment or hypoxia exposure reduced the levels of $^{13}C_4$-orotate in the heart and gastrocnemius muscle, whereas they remained unchanged in the liver and kidney, which net reverse the SDH complex (Fig. 4E and fig. S12C). These data suggest that the ability to maintain DHODH activity correlates with the ability of a tissue to net reverse the SDH complex upon inhibition of O_2 reduction.

Discussion

Here, we have elucidated a circuit of electron flow in the ETC of mammalian mitochondria that does not require O_2 as a TEA. Although the O_2 consumption rate is classically used as a metric for electron flow through the ETC, our study suggests caveats in directly equating O_2 consumption with ETC flux and overall mitochondrial function. In adapting to O_2 limitation, mammalian mitochondria use fumarate as a TEA. The accumulation of UQH_2 in hypoxia or upon inhibition of complexes III or IV drives the SDH complex in reverse to enable electron deposition onto fumarate. Fumarate reduction sustains the input of electrons into the ETC by complex I and DHODH, enabling NADH reoxidation and de novo pyrimidine biosynthesis. Although in vivo, all tested tissues perform fumarate reduction upon hypoxia exposure, only a subset net reverse the SDH reaction, and only these can maintain electron inputs into the ETC (Fig. 4F).

The surprising differences among tissues in their ability to reduce fumarate likely come from the distinctive roles of mitochondria in each tissue. For example, the heart and skeletal muscle could favor the forward SDH activity over the reverse SDH activity to maximize ATP production, whereas the kidney and brain may reduce fumarate to minimize the electron leakage out of complex III that generates ROS. Moreover, similar to the retina within the eye (*45*), we expect interesting differences in fumarate reduction among cell types within a tissue, such as the thymus, that on a bulk tissue scale, display some degree of both the forward and reverse SDH activities at physiological O_2 concentrations. The full extent to which fumarate reduction contributes to normal physiology and the ways in which fumarate reduction is regulated remain to be understood.

Our data also provide clarity to the long-standing observation that mtDNA-deficient cells require uridine to proliferate (*52*). The lack of DHODH activity to support pyrimidine biosynthesis in these cells has always been attributed to a deficiency in their ability to transfer electrons to oxygen as a TEA. Notably, fumarate reduction by the SDH complex, which is encoded by the nuclear genome, can support pyrimidine biosynthesis, albeit to a lesser extent than oxygen does (Fig. 1E). This is likely caused by less efficient electron transfer from ubiquinol onto fumarate compared with cytochrome c. Thus, mtDNA-deficient cells have the potential to use fumarate as a TEA to support pyrimidine biosynthesis. However, given their dependence on exogenous uridine for rapid proliferation, these cells are likely missing a different component of this pathway, and this warrants further investigation.

Beyond the fundamental role of fumarate reduction in mammalian mitochondria, there are many disease contexts to which fumarate reduction likely contributes. Fumarate reduction is likely important in diseases that cause tissue hypoxia such as ischemia, diabetes, obesity, and cancer. Cancer is an area of particular interest, given that some tumors have mutations in the SDH complex (*53*), the tumor microenvironment is hypoxic, and the ETC supports tumor growth in multifaceted ways (*18*, *54*–*58*). For example, given the importance of DHODH-mediated pyrimidine biosynthesis for tumor growth (*21*), fumarate reduction might sustain this process in the hypoxic tumor microenvironment. However, it is also possible that the nutrient limitations in the tumor microenvironment, particularly those of glucose and aspartate (*23*, *24*, *59*), may limit de novo pyrimidine biosynthesis. It will be critical for future work to investigate the role of fumarate reduction in the context of diseases such as cancer and to dissociate the forward and reverse activities of the SDH complex in each of these systems.

REFERENCES AND NOTES

1. N. S. Chandel *et al.*, *J. Biol. Chem.* **275**, 25130–25138 (2000).
2. D. V. Titov *et al.*, *Science* **352**, 231–235 (2016).
3. T. Nakagawa *et al.*, *Nature* **434**, 652–658 (2005).
4. J. B. Spinelli, M. C. Haigis, *Nat. Cell Biol.* **20**, 745–754 (2018).
5. H. Zhang, K. J. Menzies, J. Auwerx, *Development* **145**, dev143420 (2018).
6. E. Mick *et al.*, *eLife* **9**, e49178 (2020).
7. K. J. Condon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022120118 (2021).
8. T. Ast, V. K. Mootha, *Nat. Metab.* **1**, 858–860 (2019).
9. A. J. Peacock, *BMJ* **317**, 1063–1066 (1998).
10. M. C. Simon, B. Keith, *Nat. Rev. Mol. Cell Biol.* **9**, 285–296 (2008).
11. L. Zheng, C. J. Kelly, S. P. Colgan, *Am. J. Physiol. Cell Physiol.* **309**, C350–C360 (2015).
12. G. L. Semenza, *Nat. Rev. Cancer* **3**, 721–732 (2003).
13. B. Keith, M. C. Simon, *Cell* **129**, 465–472 (2007).
14. T. Kalogeris, C. P. Baines, M. Krenz, R. J. Korthuis, *Int. Rev. Cell Mol. Biol.* **298**, 229–317 (2012).
15. F. Palm, *Clin. Exp. Pharmacol. Physiol.* **33**, 997–1001 (2006).
16. H. K. Eltzschig, P. Carmeliet, *N. Engl. J. Med.* **364**, 656–665 (2011).
17. S. Vyas, E. Zaganjor, M. C. Haigis, *Cell* **166**, 555–566 (2016).
18. I. Martínez-Reyes *et al.*, *Nature* **585**, 288–292 (2020).
19. R. Dumollard, M. Duchen, J. Carroll, in *The Mitochondrion in the Germline and Early Development*, J. C. St. John, Ed. (Academic Press, 2007), vol. 77, pp. 21–49.
20. J. Nunnari, A. Suomalainen, *Cell* **148**, 1145–1159 (2012).
21. M. Bajzikova *et al.*, *Cell Metab.* **29**, 399–416.e10 (2019).
22. E. Ansó *et al.*, *Nat. Cell Biol.* **19**, 614–625 (2017).
23. J. Garcia-Bermudez *et al.*, *Nat. Cell Biol.* **20**, 775–781 (2018).
24. L. B. Sullivan *et al.*, *Nat. Cell Biol.* **20**, 782–788 (2018).
25. S. Cardaci *et al.*, *Nat. Cell Biol.* **17**, 1317–1326 (2015).
26. D. F. Wilson, W. L. Rumsey, T. J. Green, J. M. Vanderkooi, *J. Biol. Chem.* **263**, 2712–2718 (1988).
27. R. Fukuda *et al.*, *Cell* **129**, 111–122 (2007).
28. P. Lee, N. S. Chandel, M. C. Simon, *Nat. Rev. Mol. Cell Biol.* **21**, 268–283 (2020).
29. M. Spinazzi, A. Casarin, V. Pertegato, L. Salviati, C. Angelini, *Nat. Protoc.* **7**, 1235–1246 (2012).
30. I. H. Jain *et al.*, *Cell* **181**, 716–727.e11 (2020).
31. D. Tello *et al.*, *Cell Metab.* **14**, 768–779 (2011).
32. S. Y. Chan *et al.*, *Cell Metab.* **10**, 273–284 (2009).
33. M. T. Frost, Q. Wang, S. Moncada, M. Singer, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **288**, R394–R400 (2005).
34. I. H. Jain *et al.*, *Science* **352**, 54–61 (2016).
35. P. C. Hinkle, R. A. Butow, E. Racker, B. Chance, *J. Biol. Chem.* **242**, 5169–5173 (1967).
36. E. L. Robb *et al.*, *J. Biol. Chem.* **293**, 9869–9879 (2018).
37. G. M. Aloysius, A. G. M. Tielens, J. J. Van Hellemond, *Biochim. Biophys. Acta Bioenerg.* **1365**, 71–78 (1998).
38. A. R. Mullen *et al.*, *Nature* **481**, 385–388 (2011).
39. C. M. Metallo *et al.*, *Nature* **481**, 380–384 (2011).
40. E. T. Chouchani *et al.*, *Nature* **515**, 431–435 (2014).
41. A. Reddy *et al.*, *Cell* **183**, 62–75.e17 (2020).
42. C. Chinopoulos, *Int. J. Biochem. Cell Biol.* **115**, 105580 (2019).
43. J. Zhang *et al.*, *Cell Rep.* **23**, 2617–2628 (2018).
44. A. R. Mullen *et al.*, *Cell Rep.* **7**, 1679–1690 (2014).
45. C. M. Bisbach *et al.*, *Cell Rep.* **31**, 107606 (2020).
46. H. R. Pershad, J. Hirst, B. Cochran, B. A. C. Ackrell, F. A. Armstrong, *Biochim. Biophys. Acta* **1412**, 262–272 (1999).
47. A. Guarás *et al.*, *Cell Rep.* **15**, 197–209 (2016).
48. E. Perales-Clemente *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18735–18739 (2008).
49. S. W. Perry, J. P. Norman, J. Barbieri, E. B. Brown, H. A. Gelbard, *Biotechniques* **50**, 98–115 (2011).
50. C. Jang, L. Chen, J. D. Rabinowitz, *Cell* **173**, 822–837 (2018).
51. P. W. Hochachka, R. H. Dressendorfer, *Eur. J. Appl. Physiol. Occup. Physiol.* **35**, 235–242 (1976).
52. M. P. King, G. Attardi, *Science* **246**, 500–503 (1989).
53. M. A. Selak *et al.*, *Cancer Cell* **7**, 77–85 (2005).
54. A. Viale *et al.*, *Nature* **514**, 628–632 (2014).
55. F. Weinberg *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8788–8793 (2010).
56. S. Boukalova *et al.*, *Mol. Cancer Ther.* **15**, 2875–2886 (2016).

57. K. K. Brown, J. B. Spinelli, J. M. Asara, A. Toker, *Cancer Discov.* **7**, 391–399 (2017).
58. J. Rohlena, L.-F. Dong, S. J. Ralph, J. Neuzil, *Antioxid. Redox Signal.* **15**, 2951–2974 (2011).
59. A. N. Lau, M. G. Vander Heiden, *Annu. Rev. Cancer Biol.* **4**, 17–40 (2020).

ACKNOWLEDGMENTS

We thank all members of the Sabatini lab for their thoughtful comments, especially G. Frenkel, A. Armani, and N. Kory for thoughtful feedback on the manuscript and experiments. We thank the Whitehead Institute Metabolite Profiling core for their assistance with experimental design and data interpretation and the Keck Imaging facility for help with microscopy experiments. **Funding:** This work was funded by an R01 application granted to D.M.S. (R01CA219859). J.B.S. is funded by the NCI F99/K00 predoctoral to postdoctoral transition fellowship (K00CA234839) and P.C.R. by a predoctoral fellowship from the NCI (F31 CA254162-01). H.G.S. is funded by the DFG (Deutsche Forschungsgemeinschaft; SP 1897/1-1) and is a Hope Funds for Cancer Research Fellow supported by the Hope Funds for Cancer Research (HFCR-20-03-01). A.M.P. is funded by a William N. and Bernice E. Bumpus Fellowship, A.L.C. by an NIH F31 predoctoral fellowship (F31 5F31DK113665), K.J.C. by an MIT School of Science Fellowship in Cancer Research and an NSF fellowship (2016197106), and N.S.C. by NIH grants (R35CA197532 and 5P01AG049665). D.M.S. is formerly an investigator of the Howard Hughes Medical Institute. **Author contributions:** J.B.S. initiated the project, designed and analyzed most experiments, and interpreted experimental results with guidance from D.M.S.; J.B.S., P.C.R., T.Z., and A.H. purified mitochondria and performed biochemical assays. J.B.S., J.M.R., H.-G.S., A.M.P., and A.L.C performed mouse experiments. J.L.M. performed proliferation experiments. J.B.S. and K.J.C. performed fluorescence-activated cell sorting experiments. J.B.S., T.K., J.L.M., and C.A.L. designed and analyzed LC-MS experiments. N.S.C. guided experimental design and data interpretation. J.B.S. and D.M.S. wrote the manuscript and acquired funding. **Competing interests:** All authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. To ensure sustainable access to data and materials associated with this study, the institution has committed to assuring long-term access and has designated sabadmin@wi.mit.edu as a contact. Access to reagents not found on Addgene will be facilitated by sabadmin@wi.mit.edu.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi7495
Materials and Methods
Figs. S1 to S12
References (*60*–*64*)
MDAR Reproducibility Checklist

28 March 2021; accepted 2 November 2021
10.1126/science.abi7495

TOPOLOGICAL MATTER

Realizing topologically ordered states on a quantum processor

K. J. Satzinger[1]*, Y.-J Liu[2,3], A. Smith[2,4,5], C. Knapp[6,7]†, M. Newman[1], C. Jones[1], Z. Chen[1], C. Quintana[1], X. Mi[1], A. Dunsworth[1], C. Gidney[1], I. Aleiner[1], F. Arute[1], K. Arya[1], J. Atalaya[1], R. Babbush[1], J. C. Bardin[1,8], R. Barends[1], J. Basso[1], A. Bengtsson[1], A. Bilmes[1], M. Broughton[1], B. B. Buckley[1], D. A. Buell[1], B. Burkett[1], N. Bushnell[1], B. Chiaro[1], R. Collins[1], W. Courtney[1], S. Demura[1], A. R. Derk[1], D. Eppens[1], C. Erickson[1], L. Faoro[9], E. Farhi[1], A. G. Fowler[1], B. Foxen[1], M. Giustina[1], A. Greene[1,10], J. A. Gross[1], M. P. Harrigan[1], S. D. Harrington[1], J. Hilton[1], S. Hong[1], T. Huang[1], W. J. Huggins[1], L. B. Ioffe[1], S. V. Isakov[1], E. Jeffrey[1], Z. Jiang[1], D. Kafri[1], K. Kechedzhi[1], T. Khattar[1], S. Kim[1], P. V. Klimov[1], A. N. Korotkov[1,11], F. Kostritsa[1], D. Landhuis[1], P. Laptev[1], A. Locharla[1], E. Lucero[1], O. Martin[1], J. R. McClean[1], M. McEwen[1,12], K. C. Miao[1], M. Mohseni[1], S. Montazeri[1], W. Mruczkiewicz[1], J. Mutus[1], O. Naaman[1], M. Neeley[1], C. Neill[1], M. Y. Niu[1], T. E. O'Brien[1], A. Opremcak[1], B. Pató[1], A. Petukhov[1], N. C. Rubin[1], D. Sank[1], V. Shvarts[1], D. Strain[1], M. Szalay[1], B. Villalonga[1], T. C. White[1], Z. Yao[1], P. Yeh[1], J. Yoo[1], A. Zalcman[1], H. Neven[1], S. Boixo[1], A. Megrant[1], Y. Chen[1], J. Kelly[1], V. Smelyanskiy[1], A. Kitaev[1,6,7], M. Knap[2,3,13], F. Pollmann[2,3]*, P. Roushan[1]*

The discovery of topological order has revised the understanding of quantum matter and provided the theoretical foundation for many quantum error–correcting codes. Realizing topologically ordered states has proven to be challenging in both condensed matter and synthetic quantum systems. We prepared the ground state of the toric code Hamiltonian using an efficient quantum circuit on a superconducting quantum processor. We measured a topological entanglement entropy near the expected value of −ln2 and simulated anyon interferometry to extract the braiding statistics of the emergent excitations. Furthermore, we investigated key aspects of the surface code, including logical state injection and the decay of the nonlocal order parameter. Our results demonstrate the potential for quantum processors to provide insights into topological quantum matter and quantum error correction.

Different phases of matter can commonly be distinguished in terms of spontaneous symmetry breaking and local order parameters. However, several exotic quantum phases have been discovered in recent decades that defy this simple classification, instead exhibiting topological order (*1*, *2*). These phases are characterized by their long-range quantum entanglement and the emergence of quasiparticles with anyonic exchange statistics. Moreover, they have energetically gapped ground states with degeneracies that depend on their boundary conditions. The nonlocal nature of these states makes them particularly attractive platforms for fault-tolerant quantum computation because quantum information encoded in locally indistinguishable ground states is robust to local perturbations (*3*, *4*). This is the underlying principle of topological quantum error–correcting codes, in which the logical codespace corresponds to the degenerate ground-state subspace of a lattice model (*5*–*7*).

An archetypical topological two-dimensional (2D) lattice model is the toric code, which exhibits so-called $\mathbb{Z}_2$ topological order (*3*). The realization of the toric code on a plane—the surface code—has emerged as one of the most promising stabilizer codes for quantum error correction owing to its amenable physical requirements (*8*, *9*). Given both its inherent richness and applications in quantum computing, experimentally realizing $\mathbb{Z}_2$ topological order has sparked extensive interest, resulting in several proposals and experimental studies with comparatively small-scale synthetic quantum systems (*10*–*21*). Despite these efforts, the experimental realization of topologically ordered states remains a major challenge, requiring the generation of long-range entanglement. This can be achieved by identifying suitable quantum systems with topologically ordered ground states or by constructing a topologically ordered state in an engineered quantum system. Probing the nonlocal topological properties of such a state on an array of qubits requires high-fidelity gates and a sufficiently large 2D lattice.

In this work, we developed an efficient quantum circuit to prepare the toric code ground state on a lattice of 31 superconducting qubits. We then experimentally established the

[1]Google Quantum AI, Mountain View, CA, USA. [2]Department of Physics, Technical University of Munich, 85748 Garching, Germany. [3]Munich Center for Quantum Science and Technology (MCQST), Schellingstraße 4, 80799 München, Germany. [4]School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK. [5]Centre for the Mathematics and Theoretical Physics of Quantum Non-Equilibrium Systems, University of Nottingham, Nottingham NG7 2RD, UK. [6]Department of Physics and Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA, USA. [7]Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA, USA. [8]Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA, USA. [9]Laboratoire de Physique Theorique et Hautes Energies, Sorbonne Université, 75005 Paris, France. [10]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [11]Department of Electrical and Computer Engineering, University of California, Riverside, CA, USA. [12]Department of Physics, University of California, Santa Barbara, CA, USA. [13]Institute for Advanced Study, Technical University of Munich, 85748 Garching, Germany.
***Corresponding author. Email: frank.pollmann@tum.de (F.P); pedramr@google.com (P.R.); ksatz@google.com (K.J.S.)**
†Present address: Station Q, Microsoft, Santa Barbara, CA, USA.

topological nature of the state by measuring the topological entanglement entropy. By simulating interferometry of toric code excitations, we fully determined their associated braiding statistics. Furthermore, we prepared logical states of the distance-5 surface code on 25 qubits and demonstrated error correction of logical measurements. Although a meaningful implementation of active error correction on these states is beyond current experimental capabilities, we realized these states without stabilizer circuitry, providing a scheme to characterize and understand errors of logical qubits.

Preparation of the ground state

We realized the toric code ground state (Fig. 1A) by implementing a shallow quantum circuit on a Sycamore quantum processor (*22*). The toric code Hamiltonian

$$H = -\sum_s A_s - \sum_p B_p \qquad (1)$$

is defined in terms of qubits living on the edges of a square lattice. The "star" operators $A_s = \prod_{i \in s} Z_i$ are products of Pauli Z operators touching each star (Fig. 1A, "+," blue). The "plaquette" operators $B_p = \prod_{j \in p} X_j$ are products of Pauli X operators on each plaquette (Fig. 1A, square shape, purple). For the boundary conditions shown in Fig. 1A, there is only one toric code ground state $|G\rangle$, with parity +1 for all star and plaquette operators: $A_s|G\rangle = B_p|G\rangle = +1|G\rangle$.

Rather than realizing the Hamiltonian explicitly, we directly prepared the ground state using the algorithm depicted in Fig. 1B. This algorithm is motivated by the observation that the ground state is an equal superposition of all possible "plaquette configurations" and can be written as

$$|G\rangle = \frac{1}{\sqrt{2^{12}}} \prod_p (\mathbb{I} + B_p)|0\rangle^{\otimes 31} \qquad (2)$$

where $|0\rangle^{\otimes 31}$ is the product of single-qubit states $|0\rangle$, and the product is over the 12 plaquettes. We began in the trivial state $|0\rangle^{\otimes 31}$, where all $\langle A_s \rangle = 1$ and $\langle B_p \rangle = 0$. For each plaquette B_p, we performed a Hadamard gate on the upper qubit, preparing the state $(|0\rangle + |1\rangle)/\sqrt{2}$, and then performed CNOT gates on the other qubits on the plaquette, effectively realizing $\mathbb{I} + B_p$. These operations were carefully ordered, starting in the middle and working outward, to avoid conflict between plaquettes while minimizing circuit depth. The 12 Hadamard gates create a superposition of 2^{12} bitstrings, and the CNOT gates transform each of those bitstrings into a configuration in which the Z parity on each star is +1; the final superposition has X parity +1 on each plaquette. This circuit exhibits optimal scaling, with depth linear in system width (*23*), specifically $3 + 2\lfloor (N-1)/2 \rfloor$ nearest-neighbor CNOT layers for a lattice N plaquettes wide (*24*).

Measurement of topological entanglement entropy

Topologically ordered states in 2D systems exhibit long-range quantum entanglement, characterized by the topological entanglement entropy S_{topo} (*24*, *26*). Ground states of 2D gapped Hamiltonians typically satisfy the "area law" scaling of the entanglement entropy: The leading-order contribution to the entanglement entropy S_A of a subsystem A results from local interactions that scale with the boundary length of the subsystem. Topological ground states have an additional universal constant contribution $S_{topo} < 0$ that arises from their intrinsic long-range entanglement. To extract S_{topo}, a linear combination of subsystem entropies can be constructed so that the local contributions cancel. For the subsystems depicted in Fig. 2A

$$S_{topo} = S_A + S_B + S_C - S_{AB} - S_{BC} - S_{AC} + S_{ABC} \qquad (3)$$

where AB indicates the union of A and B. Because the correlation length of the toric code eigenstates is zero, S_{topo} can be inferred from small subsystems. For the toric code ground state, $S_{topo} = -\ln 2$, reflecting the total quantum dimension of $\mathbb{Z}_2$ topological order (*27*), whereas $S_{topo} = 0$ is the absence of topological order.

The structure of the toric code Hamiltonian results in entanglement characterized by integer multiples of ln2, scaling with the number of star operators A_s intersecting the subsystem boundary (Fig. 2B) (*28*). To compute S_{topo}, one can measure the second Rényi entropy $S^{(2)} = -\ln[\mathrm{Tr}(\rho^2)]$, where ρ is the density matrix, for each subsystem in Eq. 3. Recently introduced randomized methods enable efficient measurement of Rényi entanglement entropies, requiring a smaller number of measurements for large subsystems as compared with full quantum state tomography (*29*–*31*). This enables accurate entropy measurement when tomography is intractable, such as the nine-qubit subsystem in Fig. 2A. We applied random single-qubit unitary gates to the subsystem of interest and sampled the probability distribution of the bitstrings. Analyzing statistical correlations across many random instances

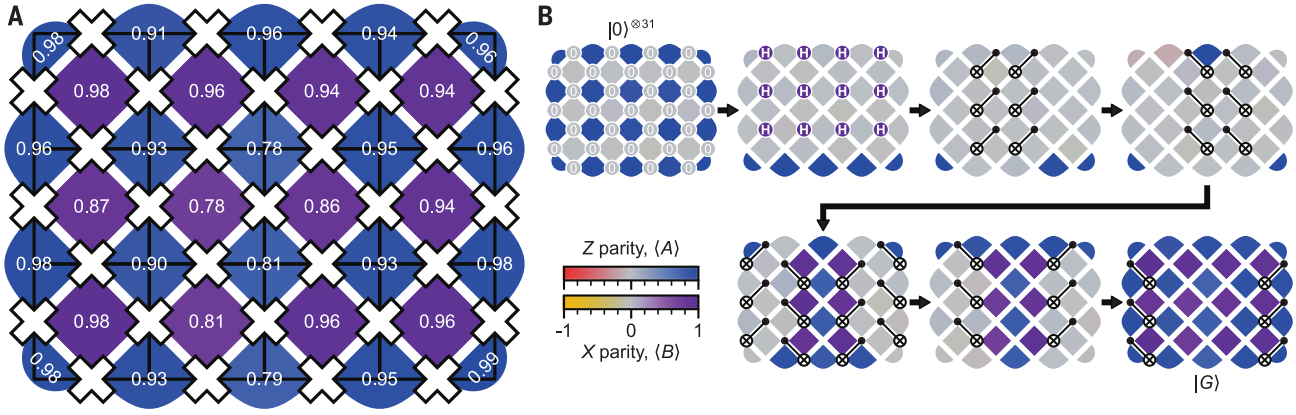


Fig. 1. Toric code ground state. (**A**) Experimentally measured parity values for a 31-qubit lattice in the toric code ground state $|G\rangle$. Qubits ("×") are drawn on the standard toric code lattice, touching star (A_s, "+," blue tile) and plaquette (B_p, square shape, purple tile) operators. We computed each parity from a measured probability distribution (measuring each A_s and B_p separately, 10^4 repetitions), which we corrected for readout error using iterative Bayesian methods (*31*) [(*32*), section II]. Mean parity, 0.92 ± 0.06 (1σ). (**B**) Quantum circuit to prepare $|G\rangle$, with quantum gates superimposed on experimentally measured parity values after each step. The circuit consists of Hadamard (H) and CNOT gates, which we compiled into CZ gates.

allowed us to compute the second Rényi entropy. We used an iterative Bayesian scheme (*32*) to mitigate measurement errors and remove undersampling bias [we also compare randomized measurement with tomography results in (*24*), section III].

Distributions of the measured entanglement entropies are shown in Fig. 2C for subsystems of 2 by 2, 2 by 3, and 3 by 3 qubits within the toric code ground state. For a subsystem with n qubits, the entanglement entropy ranges from 0 for a product state to $n\ln 2$. In the toric code, subsystems with no interior have the maximum value $n\ln 2$; in those cases, we measured a narrow distribution centered just below the ideal value. For subsystems with an interior, we measured a wider distribution centered slightly above the predicted value. This is consistent with unitary error and decoherence slightly mixing the system with its environment, which increases entanglement entropies that are not yet at their maximal value.

We computed S_{topo} from the subsystem entropies using Eq. 3 for 14 different 2-by-2 arrays, 20 different 2-by-3 arrays, and three different 3-by-3 arrays. Each randomized measurement on the qubit array yields several S_{topo} estimates from different orientations of the partitions A, B, and C. Distributions of measured S_{topo} are shown in Fig. 2D, with mean values $S_{\text{topo}}/\ln 2 = -0.89$, -0.90, and -0.95 for the 2-by-2, 2-by-3, and 3-by-3 qubit arrays, respectively. The distributions provide strong evidence for the nontrivial topological nature of the state, approaching the ideal value of $S_{\text{topo}} = -\ln 2$, which is incompatible with the trivial state value of zero.

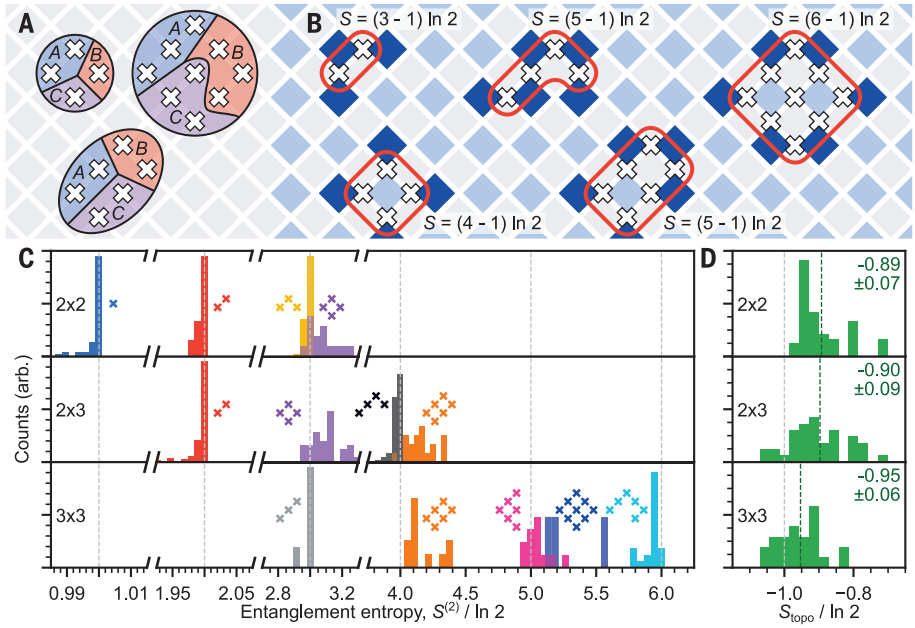


Fig. 2. Topological entanglement entropy. (**A**) Illustration of subsystems A, B, and C used to measure topological entanglement entropy S_{topo} on four-, six-, and nine-qubit systems within the toric code lattice. (**B**) Expected entanglement entropy S for groups of qubits in the toric code. We drew a red perimeter around each group and counted the number k of star operators (blue tiles) that it crosses: $S = k\ln 2 + S_{\text{topo}} = (k - 1)\ln 2$. (**C**) Second Rényi entropy $S^{(2)}$ distributions measured on the 31-qubit toric code ground state. A histogram is shown for each subsystem shape. Dashed gray lines indicate the predicted integer values for $|G\rangle$. (**D**) Topological entanglement entropy $S_{\text{topo}}/\ln 2$ (ideal value −1) computed from the entropies in (C). We evaluated each dataset in all possible orientations of the subsystems in (A) (2 by 2, 4; 2 by 3, 2; 3 by 3, 8). In (D), top to bottom, the numbers at top right show the mean (also indicated by the dark green line) and the standard deviation of the distribution.

Simulation of braiding statistics

Our approach of directly preparing the toric code ground state also allows us to simulate the exotic braiding statistics of its quasiparticle excitations (anyons). We used a mapping between the adiabatic evolution of toric code excitations and strings of Pauli operators applied to the ground state. Within this framework, a controlled Pauli string implements an interferometry protocol, through which we experimentally extracted the mutual and exchange statistics corresponding to all combinations of excitations.

The quasiparticle excitations of the toric code are commonly denoted as "electric" e with $\langle A_s \rangle = -1$, and "magnetic" m with $\langle B_p \rangle = -1$, in connection to lattice gauge theory. The four distinct anyons of the toric code are $\mathbb{1}$ (the absence of an e or m), e, m, and ψ (an emergent fermion resulting from the combination of e and m). In the toric code, the mutual statistics are encoded in the phase accumulated when dragging one anyon around another anyon of a different type, and the exchange statistics are phases that arise from the spatial interchange of two identical anyons. The toric code excited states can be created by applying a string of Pauli operators to the ground state: An X-string will result in the state with e excitations at each end, whereas a Z-string prepares the state with m excitations at each end. We visualize an example of $e - m$ mutual braiding in Fig. 3A, with snapshots of experimentally measured parity values, $\langle A_s \rangle$ and $\langle B_p \rangle$. We moved an e around m with an X-string, eventually returning to its initial position. The initial and final states have the same parity values but differ by an overall phase—in this case, π, which is not directly detectable.

To experimentally determine this phase, we used multiqubit Ramsey interferometry (*33*). This protocol provides a scalable way to measure the overlap between the initial and final states, allowing experimental access to the accumulated phase θ. A key step in this protocol is the use of an auxiliary qubit and a controlled operation, effectively creating a superposition of the braided and nonbraided states (Fig. 3, B and C). This sequence imparts θ into a measurable rotation of the auxiliary qubit. We efficiently compiled the multiqubit controlled operations into CZ gates. Because the measured phases are sensitive to coherent and non-Markovian errors, we used randomized compiling to mitigate these errors (*34*, *35*). Details are available in (*24*), section IV.

Illustrated in Fig. 3, D and E, are two examples of braiding interferometry by use of minimal-length paths, which we chose because of the zero correlation length of the states. In Fig. 3D, we extract the $e - m$ mutual statistics, where the braiding path is a Pauli string $XXXX$, moving e around the plaquette that contains an m. A similar example is shown in Fig. 3E for the exchange statistics of two identical ψ excitations by using a path of intertwining Pauli strings of $XXXX$ and $ZZZZ$, simplifying to $XXYYZZ$ [(*24*), section VI]. The parity measurements show consistent values before and after the controlled-braiding operation, slightly fading owing to decoherence and gate error. We measured the phases for the other mutual and exchange combinations and provide the results in Fig. 3F, in which the phases are plotted alongside their corresponding braid diagrams, with the expected values 0 and π indicated with dashed gray lines.

Our measurements illuminate the nontrivial mutual and exchange statistics of the toric code. Braiding e around m results in a π phase, which does not occur for local bosons or fermions. Moreover, although e and m both satisfy

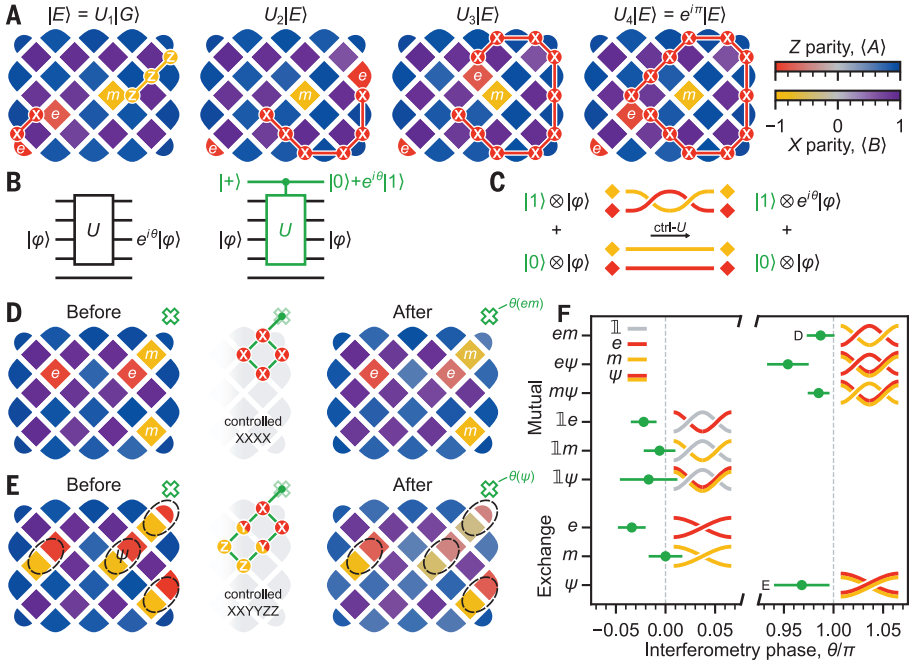


Fig. 3. Extracting braiding statistics by using Ramsey interferometry. (**A**) Visualizing braiding with a toric code excited state $|E\rangle = U_1|G\rangle$ [excitations *e* (red) and *m* (yellow), experimentally measured parities]. We applied additional *X* gates (U_2, U_3, and U_4) to visualize braiding an *e* around the *m*. (**B**) Quantum circuits with unitary *U* and an eigenstate $|\phi\rangle$. (Left) Direct application. (Right) Extracting the phase θ by using an auxiliary qubit (green). (**C**) Illustration of Ramsey interferometry for the case of braiding an *e* and *m* (state $|\phi\rangle$) by using an operator *U*. We visualized the superposition of two paths, with the braid operation *U* controlled by an auxiliary qubit in $|+\rangle$. (**D**) Extracting the mutual statistics for *e* and *m*. (Left) Initial excited eigenstate [similar to that in (A)]. We implemented controlled-*XXXX* with an auxiliary control qubit (green) starting in $|+\rangle$. (Right) Parity measurements after controlled-*XXXX*. (**E**) Extracting the fermion exchange statistics, analogous to (D). We created two pairs of ψ (neighboring *e* and *m*) and implemented controlled-*XXYYZZ* to measure the exchange phase. (**F**) Measured mutual and exchange phases, with braiding diagrams. Phases are from tomography on the auxiliary qubit, 18,000 total repetitions per compiled instance. Standard error was estimated with jackknife resampling over instances.

bosonic exchange statistics, their combination ψ exhibits fermionic exchange statistics. The mutual and exchange statistics of the anyons, conventionally summarized in the modular $\mathcal{S}$ and $\mathcal{T}$ matrices, fully characterize the $\mathbb{Z}_2$ topological order (*29*).

Preparation of logical qubit states

Distinct topologically ordered ground states are locally indistinguishable, making them attractive logical qubits thanks to this immunity to local perturbations. The lattice of Fig. 1A has only one ground state under Eq. 1, but as shown in Fig. 4A, we used different boundary conditions in which the toric code admits a ground-state degeneracy, as proposed for the surface code (*5*, *6*, *36*). We introduced logical operators Z_L and X_L, which span across the lattice and commute with the Hamiltonian but anticommute with each other.

We generalized the state preparation circuit of Fig. 1B to create the logical states $|0_L\rangle$ and $|+_L\rangle$, where $Z_L|0_L\rangle = +1|0_L\rangle$ and $X_L|+_L\rangle = +1|+_L\rangle$, on both 5-by-5 (distance-5) and 3-by-3 (distance-3) arrays. The $|0_L\rangle$ and $|+_L\rangle$ preparations are closely related, connected by a logical Hadamard gate. We then used the logical operators, which are simply products of single-qubit gates, to realize $|1_L\rangle = X_L|0_L\rangle$ and $|-_L\rangle = Z_L|+_L\rangle$. Details on state preparation and logical operations are provided in (*24*), section I.

The logical states are resilient to local errors, which we demonstrated with logical measurement with error correction (Fig. 4B). Following surface code proposals, we performed a logical measurement by projectively measuring all the qubits in *Z* or *X* basis (for Z_L or X_L, respectively). Naively evaluating the parity of the logical operator is vulnerable to errors on any qubit along the operator, but errors can be detected by also evaluating the local parities (A_s or B_p) from the individual qubit measurements. By construction, we expected the local parities to be +1, so any −1 values indicate nearby errors. We found a minimal set of qubits to flip in order to recover +1 parities before evaluating the logical operator. This correction decreases the logical error substantially. Averaging logical preparation and measurement error over X_L and Z_L eigenstates, without correction, we observed 0.17 for distance-5 and 0.090 for distance-3, whereas with correction, we observed 0.030 for both, which is lower than the average physical qubit preparation and measurement error, 0.034. This is a simplified form of error correction compared with the repetitive stabilizer measurements of surface code proposals, in which parity changes are matched together over space and time.

The logical subspace also admits arbitrary superposition states $\alpha|0_L\rangle + \beta|1_L\rangle$, which we realized with state injection, encoding a single physical qubit state into the logical qubit. For 5-by-5 state injection, we prepared the central qubit in $\alpha|0\rangle + \beta|1\rangle$ and then created a GHZ-like state $(\alpha\mathbb{I} + \beta X_L)0^{\otimes 25}$ using three CZ layers. The toric code preparation circuit maps $|0\rangle^{\otimes 25} \rightarrow |0_L\rangle$ and $X_L|0\rangle^{\otimes 25} \rightarrow |1_L\rangle$, giving $\alpha|0_L\rangle + \beta|1_L\rangle$. For example, the states depicted in Fig. 4A are logical *T* states $|T_L\rangle = (|0_L\rangle + e^{i\pi/4}|1_L\rangle)/\sqrt{2}$, which is of interest for non-Clifford operations. We characterized injected states using logical tomography. Measurement of Z_L and X_L was straightforward and robust. We measured Y_L by performing another logical gate, $X_L^{1/2} = (\mathbb{I} - iX_L)/\sqrt{2}$, decomposed into five CZ layers, and then measuring Z_L. We plot the resultant Bloch vectors for 128 injected states across the Bloch sphere in Fig. 4C. By measuring these nonlocal order parameters, we illuminate the logical degree of freedom that was invisible to the local parity measurements of Fig. 4A. In contrast to surface code proposals, our state-preparation methods are not fault tolerant, although in principle, states prepared in this way can be further purified through distillation (*7*, *8*, *37*, *38*).

Last, we investigated decoherence of Z_L and X_L eigenstates by plotting logical error versus wait time *t* (Fig. 4D). We continued our focus on measurement error correction through comparison of the raw and corrected data. Although distance-5 has substantially worse raw error, after correction it is modestly better than distance-3 for $|1_L\rangle$. However, $|+_L\rangle$ decays much more quickly than $|1_L\rangle$, owing to its sensitivity to *Z* errors (dephasing). We dynamically decoupled the qubits from low-frequency noise with a simple sequence that executed an *X* gate on each qubit at *t*/4 and 3*t*/4, which brought $|+_L\rangle$ error slightly below $|1_L\rangle$ error, with distance-3 remaining slightly lower-error compared with that of distance-5. $|1_L\rangle$ and $|0_L\rangle$ are not appreciably affected by this dynamical decoupling [(*24*), section V]. Overall, the logical error increases linearly at 0.06 per microsecond. For active error correction with the surface code, we expected a few percent logical error per cycle at threshold (*9*). Typical cycle durations are hundreds of nanoseconds (*39*), in which the logical state suffers the decoherence studied here as well as

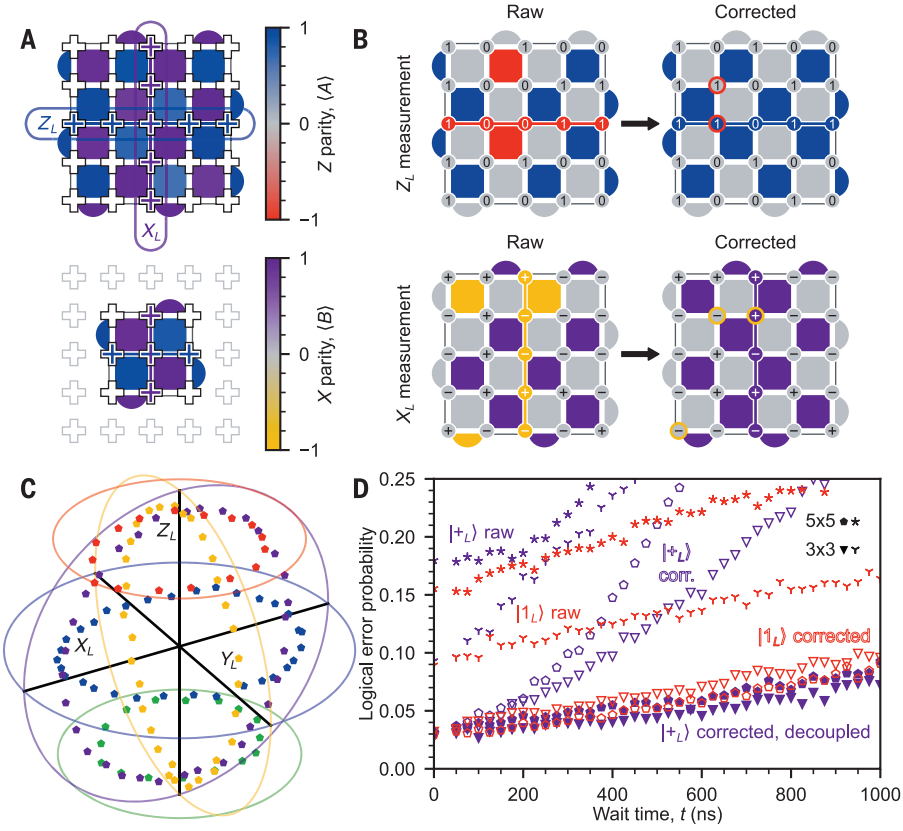


Fig. 4. Surface code logical qubit states. (**A**) Measured parity values for surface code logical qubit states $|T_L\rangle = \left(|0_L\rangle + e^{i\pi/4}|1_L\rangle/\sqrt{2}\right)$ on 5-by-5 and 3-by-3 qubit arrays. Logical operators Z_L and X_L span across each array. (**B**) Logical measurement with error correction. We measured a 25-qubit bitstring in X or Z basis and evaluated the local parities of the same basis. Negative parities indicate an error. We flipped the circled qubits to restore positive parities. (**C**) Experimental logical qubit tomography immediately after state injection for 128 states (sweeping the initial state of the center qubit $\alpha|0\rangle + \beta|1\rangle$), plotted in the Bloch sphere (5 by 5). The ideal states lie on five planes: $x = 0$ (yellow), $y = 0$ (purple), $z = 1/\sqrt{2}$ (red), $z = 0$ (blue), and $z = -1/\sqrt{2}$ (green). Mean Bloch vector length, 0.6 ± 0.1 (1σ). (**D**) We prepared logical states, waited for a time t, and then performed a logical measurement. We compared logical measurements for $|1_L\rangle$ (red) and $|+_L\rangle$ (purple), for both 5-by-5 (5-sided markers) and 3-by-3 (3-sided markers) states. Stars indicate raw measurements. Open symbols indicate corrected measurements. Solid symbols indicate corrected measurements with dynamical decoupling during the wait time ($|+_L\rangle$ only). Each logical measurement used 10^4 repetitions.

gate errors, suggesting that continued efforts are needed to decrease the cycle duration and improve coherence.

Discussion

Our shallow quantum circuits for realizing toric code eigenstates and simulating braiding statistics can be extended to other topologically ordered states, including string-nets with non-Abelian anyons (*40*, *41*). The quasi-static protocol that uses controlled Pauli strings to simulate braiding can be generalized to dynamical adiabatic braiding (*42*) by using deeper circuits in future devices. Moreover, the tools we developed can be readily applied to a wide class of topologically ordered states generated on quantum processors. By encoding quantum information in the degenerate ground-state manifold of the toric code, we provide a method for studying coherence properties of logical qubit states. This method could be used to identify and mitigate noise correlations in the system, with critical implications for future error-correction experiments.

Note added in proof: A contemporary work by G. Semeghini *et al.* (*43*) studies topological spin liquid states on a Rydberg quantum simulator.

REFERENCES AND NOTES

1. D. C. Tsui, H. L. Stormer, A. C. Gossard, *Phys. Rev. Lett.* **48**, 1559–1562 (1982).
2. X. G. Wen, *Int. J. Mod. Phys. B* **4**, 239–271 (1990).
3. A. Y. Kitaev, *Ann. Phys.* **303**, 2–30 (2003).
4. M. H. Freedman, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 98–101 (1998).
5. S. B. Bravyi, A. Y. Kitaev, arXiv:quant-ph/9811052 (1998).
6. M. H. Freedman, D. A. Meyer, *Found. Comput. Math.* **1**, 325–332 (2001).
7. E. Dennis, A. Kitaev, A. Landahl, J. Preskill, *J. Math. Phys.* **43**, 4452–4505 (2002).
8. R. Raussendorf, J. Harrington, *Phys. Rev. Lett.* **98**, 190504 (2007).
9. A. G. Fowler, M. Mariantoni, J. M. Martinis, A. N. Cleland, *Phys. Rev. A* **86**, 032324 (2012).
10. C.-Y. Lu *et al.*, *Phys. Rev. Lett.* **102**, 030502 (2009).
11. J. K. Pachos *et al.*, *New J. Phys.* **11**, 083010 (2009).
12. G. Feng, G. Long, R. Laflamme, *Phys. Rev. A* **88**, 022305 (2013).
13. A. J. Park, E. McKay, D. Lu, R. Laflamme, *New J. Phys.* **18**, 043043 (2016).
14. Y. P. Zhong *et al.*, *Phys. Rev. Lett.* **117**, 110501 (2016).
15. H.-N. Dai *et al.*, *Nat. Phys.* **13**, 1195–1200 (2017).
16. Z. Luo *et al.*, *Nat. Phys.* **14**, 160–165 (2018).
17. C. Song *et al.*, *Phys. Rev. Lett.* **121**, 030502 (2018).
18. C. Liu *et al.*, *Optica* **6**, 264 (2019).
19. C. K. Andersen *et al.*, *Nat. Phys.* **16**, 875–880 (2020).
20. A. Rahmani *et al.*, *PRX Quantum* **1**, 020309 (2020).
21. A. Erhard *et al.*, *Nature* **589**, 220–224 (2021).
22. F. Arute *et al.*, *Nature* **574**, 505–510 (2019).
23. S. Bravyi, M. B. Hastings, F. Verstraete, *Phys. Rev. Lett.* **97**, 050401 (2006).
24. Materials and methods are available as supplementary materials.
25. A. Kitaev, J. Preskill, *Phys. Rev. Lett.* **96**, 110404 (2006).
26. M. Levin, X.-G. Wen, *Phys. Rev. Lett.* **96**, 110405 (2006).
27. A. Kitaev, *Ann. Phys.* **321**, 2–111 (2006).
28. A. Hamma, R. Ionicioiu, P. Zanardi, *Phys. Lett. A* **337**, 22–28 (2005).
29. S. J. van Enk, C. W. J. Beenakker, *Phys. Rev. Lett.* **108**, 110503 (2012).
30. A. Elben, B. Vermersch, M. Dalmonte, J. I. Cirac, P. Zoller, *Phys. Rev. Lett.* **120**, 050406 (2018).
31. T. Brydges *et al.*, *Science* **364**, 260–263 (2019).
32. B. Nachman, M. Urbanek, W. A. de Jong, C. W. Bauer, *npj Quantum Inf.* **6**, 84 (2020).
33. L. Jiang *et al.*, *Nat. Phys.* **4**, 482–488 (2008).
34. J. J. Wallman, J. Emerson, *Phys. Rev. A* **94**, 052325 (2016).
35. S. J. Beale *et al.*, True-Q Version 2. Zenodo (2020); doi: 10.5281/zenodo.3945250.
36. C. Horsman, A. G. Fowler, S. Devitt, R. Van Meter, *New J. Phys.* **14**, 123011 (2012).
37. S. Bravyi, A. Kitaev, *Phys. Rev. A* **71**, 022316 (2005).
38. D. Gottesman, *Quantum Information Processing: From Theory to Experiment* (IOS Press, 2006).
39. Z. Chen *et al.*, *Nature* **595**, 383–387 (2021).
40. M. A. Levin, X.-G. Wen, *Phys. Rev. B Condens. Matter Mater. Phys.* **71**, 045110 (2005).
41. Y.-J. Liu, K. Shtengel, A. Smith, F. Pollmann, arXiv:2110.02020 [quant-ph] (2021).
42. M. Levin, X.-G. Wen, *Phys. Rev. B Condens. Matter* **67**, 245316 (2003).
43. G. Semeghini *et al.*, *Science* **374**, 1242–1247 (2021).
44. K. Satzinger, Data for "Realizing topologically ordered states on a quantum processor". *Zenodo* (2021); doi: 10.5281/zenodo.5570481.

ACKNOWLEDGMENTS

We thank B. Bauer, A. Elben, B. Vermersch, and G. Vidal for useful discussions. **Funding:** F.P., Y.-J.L., A.S., and M.K. acknowledge support from the Technical University of Munich–Institute for Advanced Study, funded by the German Excellence Initiative and the European Union FP7 under grant agreement 291763; the Max Planck Gesellschaft (MPG) through the International Max Planck Research School for Quantum Science and Technology (IMPRS-QST); the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under Germany's Excellence Strategy–EXC–2111–390814868, TRR80, and DFG grant KN1254/2-1; and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreements 771537 and 851161). A.S. was supported by a Research Fellowship from the Royal Commission for the Exhibition of 1851. C.K. was supported by the Walter Burke Institute for Theoretical Physics at Caltech, and by the IQIM, an NSF Frontier center funded by the Gordon and Betty Moore Foundation, the Packard Foundation, and the Simons Foundation. **Author contributions:** A.S., M.K., F.P., K.J.S., Y.-J.L., C.K., and P.R. designed the experiment. K.J.S. and P.R. performed the experiment. K.J.S. and Y.-J.L. analyzed the data and wrote the supplement. Y.-J.L., A.S., C.K., M.K., F.P., and K.J.S. provided theoretical support and analysis. C.K., K.JS., Y.-J.L., A.S., M.K., F.P., and P.R. wrote the manuscript. All authors contributed to revising the manuscript and supplement. All authors contributed to the experimental and theoretical infrastructure to enable the experiment. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data and code used for analysis and simulation are available at (*44*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi8378
Supplementary Text
Figs. S1 to S25
References (*45*–*63*)

2 April 2021; accepted 28 October 2021
10.1126/science.abi8378

TOPOLOGICAL MATTER

Probing topological spin liquids on a programmable quantum simulator

G. Semeghini[1], H. Levine[1], A. Keesling[1,2], S. Ebadi[1], T. T. Wang[1], D. Bluvstein[1], R. Verresen[1], H. Pichler[3,4], M. Kalinowski[1], R. Samajdar[1], A. Omran[1,2], S. Sachdev[1,5], A. Vishwanath[1]*, M. Greiner[1]*, V. Vuletić[6]*, M. D. Lukin[1]*

Quantum spin liquids, exotic phases of matter with topological order, have been a major focus in physics for the past several decades. Such phases feature long-range quantum entanglement that can potentially be exploited to realize robust quantum computation. We used a 219-atom programmable quantum simulator to probe quantum spin liquid states. In our approach, arrays of atoms were placed on the links of a kagome lattice, and evolution under Rydberg blockade created frustrated quantum states with no local order. The onset of a quantum spin liquid phase of the paradigmatic toric code type was detected by using topological string operators that provide direct signatures of topological order and quantum correlations. Our observations enable the controlled experimental exploration of topological matter and protected quantum information processing.

Motivated by theoretical work carried out over the past five decades, a broad search has been underway to identify signatures of quantum spin liquids (QSLs) in correlated materials (*1*, *2*). Moreover, inspired by the intriguing predictions of quantum information theory (*3*), approaches to engineer such systems for topological protection of quantum information are being actively explored (*4*). Systems with frustration (*5*) caused by the lattice geometry or long-range interactions constitute a promising avenue in the search for QSLs. In particular, such systems can be used to implement a class of so-called dimer models (*6*–*10*), which are among the most promising candidates to host QSL states. However, realizing and probing such states is challenging because they are often surrounded by other competing phases. Moreover, in contrast to topological systems that involve time-reversal symmetry breaking, such as in the fractional quantum Hall effect (*11*), these states cannot be easily probed by means of, for example, quantized conductance or edge states. Instead, to diagnose spin liquid phases, it is essential to access nonlocal observables, such as topological string operators (*1*, *2*). Although some indications of QSL phases in correlated materials have been previously reported (*12*, *13*), thus far, these exotic states of matter have evaded direct experimental detection.

[1]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [2]QuEra Computing, Boston, MA 02135, USA. [3]Institute for Theoretical Physics, University of Innsbruck, Innsbruck A-6020, Austria. [4]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Innsbruck A-6020, Austria. [5]School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA. [6]Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: avishwanath@g.harvard.edu (A.V.); greiner@physics.harvard.edu (M.G.); vuletic@mit.edu (V.V.); lukin@physics.harvard.edu (M.D.L.)

Programmable quantum simulators are well suited for the controlled exploration of these strongly correlated quantum phases (*14*–*21*). In particular, recent work showed that various phases of quantum dimer models can be efficiently implemented by using Rydberg atom arrays (*22*) and that a dimer spin liquid state of the toric code type could be potentially created in a specific frustrated lattice (*23*). Toric code states have been dynamically created in small systems by using quantum circuits (*24*, *25*). However, some of the key properties, such as topological robustness, are challenging to realize in such systems. Spin liquids have also been explored by using quantum annealers, but the lack of coherence in these systems has precluded the observation of quantum features (*26*).

Dimer models in Rydberg atom arrays

The key idea of our approach is based on a correspondence (*23*) between Rydberg atoms placed on the links of a kagome lattice (or equivalently, the sites of a ruby lattice) (Fig. 1A) and dimer models on the kagome lattice (*8*, *10*). The Rydberg excitations can be viewed as "dimer bonds" that connect the two adjacent vertices of the lattice (Fig. 1B). Because of the Rydberg blockade (*27*), strong and properly tuned interactions constrain the density of excitations so that each vertex is touched by a maximum of one dimer. At 1/4 filling, each vertex is touched by exactly one dimer, resulting in a perfect dimer covering of the lattice. Smaller filling fractions result in a finite density of vertices with no proximal dimers, which are referred to as monomers. A QSL can emerge within this dimer-monomer model close to 1/4 filling (*23*) and can be viewed as a coherent superposition of exponentially many degenerate dimer coverings with a small admixture of monomers (Fig. 1C) (*10*). This corresponds to the resonating valence bond (RVB) state (*6*, *28*), which was predicted long ago but is so far still unobserved in any experimental system.

To create and study such states experimentally, we used two-dimensional arrays of 219 ^{87}Rb atoms individually trapped in optical tweezers (*29*, *30*) and positioned on the links of a kagome lattice (Fig. 1A). The atoms were initialized in an electronic ground state $|g\rangle$ and coupled to a Rydberg state $|r\rangle$ by means of a two-photon optical transition with Rabi frequency Ω. The atoms in the Rydberg state $|r\rangle$ interact with one another through a strong van der Waals potential $V = V_0/d^6$, where d is the interatomic distance. This strong interaction prevents the simultaneous excitation of two atoms within a blockade radius $R_b = (V_0/\Omega)^{1/6}$ (*27*). We adjusted the lattice spacing a and the Rabi frequency Ω so that for each atom in $|r\rangle$, its six nearest neighbors are all within the blockade radius (Fig. 1B), resulting in a maximum filling fraction of 1/4. The resulting dynamics correspond to unitary evolution $U(t)$ governed by the Hamiltonian

$$\frac{H}{\hbar} = \frac{\Omega(t)}{2}\sum_i \sigma_i^x - \Delta(t)\sum_i n_i + \sum_{i<j} V_{ij} n_i n_j \quad (1)$$

where $\hbar$ is Planck's constant h divided by 2π, $n_i = |r_i\rangle\langle r_i|$ is the Rydberg state occupation at site i, $\sigma_i^x = |g_i\rangle\langle r_i| + |r_i\rangle\langle g_i|$, and $\Delta(t)$ is the time-dependent two-photon detuning. After the evolution, the state was analyzed by means of projective readout of ground-state atoms (Fig. 1A, right) (*29*).

To explore many-body phases in this system, we used quasi-adiabatic evolution, in which we slowly turned on the Rydberg coupling Ω and subsequently changed the detuning Δ from negative to positive values by using a cubic frequency sweep over ~2 μs (Fig. 1D). We stopped the cubic sweep at different endpoints and first measured the density of Rydberg excitations $\langle n\rangle$. Away from the array boundaries (which result in edge effects permeating just two layers into the bulk), we observed that the average density of Rydberg atoms was uniform across the array (fig. S4) (*31*). Focusing on the bulk density, we found that for $\Delta/\Omega \gtrsim 3$, the system reaches the desired filling fraction $\langle n\rangle \sim 1/4$ (Fig. 1E, top). The resulting state does not have any obvious spatial order (Fig. 1A) and appears as a different configuration of Rydberg atoms in each experimental repetition (fig. S5) (*31*). From the single-shot images, we evaluated the probability for each vertex of the kagome lattice to be attached to one dimer (as in a perfect dimer covering), zero dimers (a monomer), or two dimers (representing weak blockade violations). Around $\Delta/\Omega \sim 4$, we observed an approximate plateau at which ~80% of the vertices were connected to a single dimer

(Fig. 1E), indicating an approximate dimer covering.

Measuring topological string operators

A defining property of a phase with topological order is that it cannot be probed locally. Hence, to investigate the possible presence of a QSL state, it is essential to measure nonlocal observables. In the case of dimer models, a particularly convenient set of nonlocal variables is defined in terms of topological string operators, which are analogous to those used in the toric code model (*3*). For the present model, there are two such string operators, the first of which characterizes the effective dimer description; the second probes quantum coherence between dimer states (*23*). We first focused on the diagonal operator $Z = \prod_{i \in s} \sigma_i^z$, with $\sigma_i^z = 1 - 2n_i$, which measures the parity of Rydberg atoms along a string S perpendicular to the bonds of the kagome lattice (Fig. 2A). For the smallest closed Z loop, which encloses a single vertex of the kagome lattice, $\langle Z \rangle = -1$ for any perfect dimer covering. Larger loops can be decomposed into a product of small loops around all the enclosed vertices, resulting in

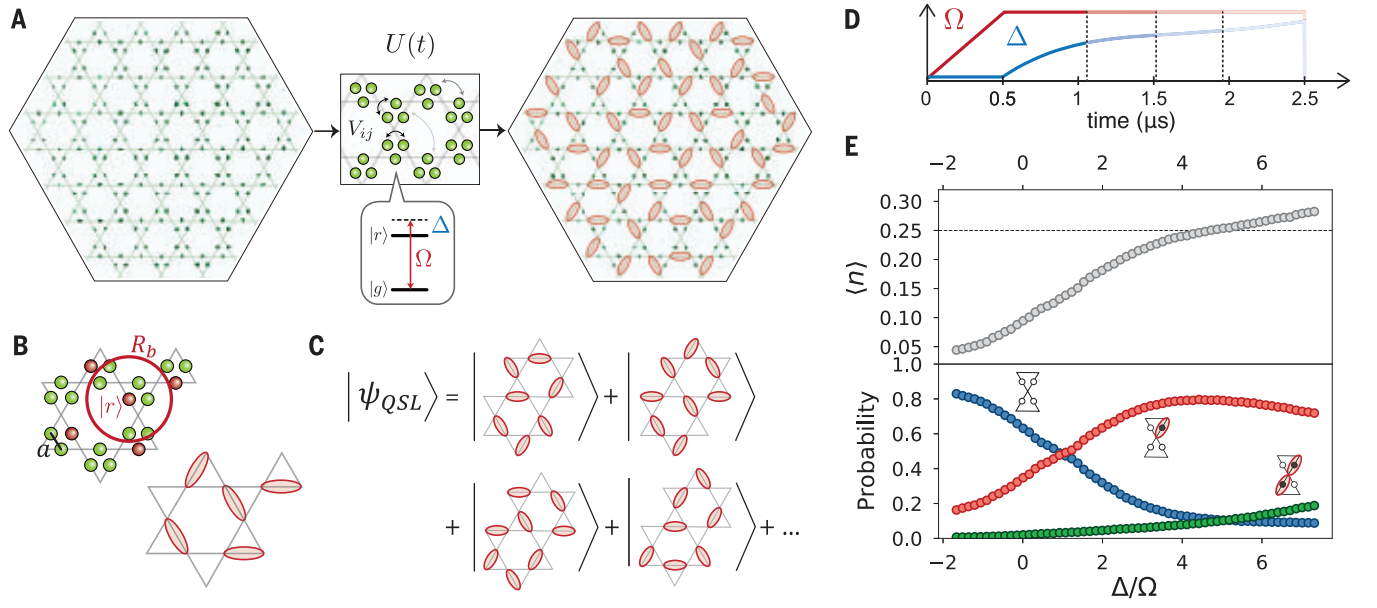


Fig. 1. Dimer model in Rydberg atoms arrays. (**A**) Fluorescence image of 219 atoms arranged on the links of a kagome lattice. The atoms, initially in the ground state $|g\rangle$, evolve according to the many-body dynamics $U(t)$. The final state of the atoms was determined by means of fluorescence imaging of ground-state atoms. Rydberg atoms are indicated with red dimers on the bonds of the kagome lattice. (**B**) We adjusted the blockade radius to $R_b/a = 2.4$ by choosing $\Omega = 2\pi \times 1.4$ MHz and $a = 3.9$ μm, so that all six nearest neighbors of an atom in $|r\rangle$ are within the blockade radius R_b. A state consistent with the Rydberg blockade at maximal filling can then be viewed as a dimer covering of the kagome lattice, where each vertex is touched by exactly one dimer. (**C**) In the idealized limit, the QSL state corresponds to a coherent superposition of exponentially many dimer coverings. (**D**) Detuning $\Delta(t)$ and Rabi frequency $\Omega(t)$ used for quasi-adiabatic state preparation. (**E**) (Top) Average density of Rydberg excitations $\langle n \rangle$ in the bulk of the system, excluding the outer three layers (*31*). (Bottom) Probabilities of empty vertices in the bulk (monomers; blue symbols), vertices attached to a single dimer (red symbols), or to double dimers (weakly violating blockade; green symbols). After $\Delta/\Omega \sim 3$, the system reaches ~1/4 filling, where most vertices are attached to a single dimer, which is consistent with an approximate dimer phase. The average density of defects per vertex in the approximate dimer phase is ~0.2.

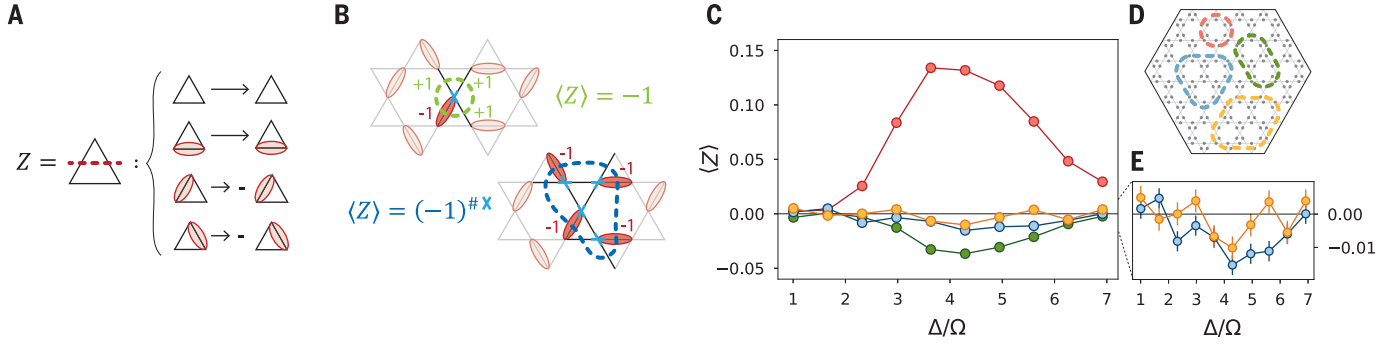


Fig. 2. Detecting a dimer phase by means of diagonal string operator. (**A**) The Z string operator measures the parity of dimers along a string. (**B**) A perfect dimer covering always has exactly one dimer touching each vertex of the array, so that $\langle Z \rangle = (-1)$ around a single vertex and $\langle Z \rangle = (-1)^{\#\text{enclosed vertices}}$ for larger loops. (**C**) Z parity measurements following the quasi-adiabatic sweep of Fig. 1D, with the addition of a 200-ns ramp-down of Ω at the end to optimize preparation. At different endpoints of the sweep and for (**D**) different loop sizes, we measured a finite $\langle Z \rangle$, which is consistent with an approximate dimer phase. The sign of $\langle Z \rangle$ properly matches the parity of the number of enclosed vertices: 6 (red), 11 (green), 15 (blue), and 19 (orange). (**E**) The measured $\langle Z \rangle$ for the two largest loops (fig. S9).

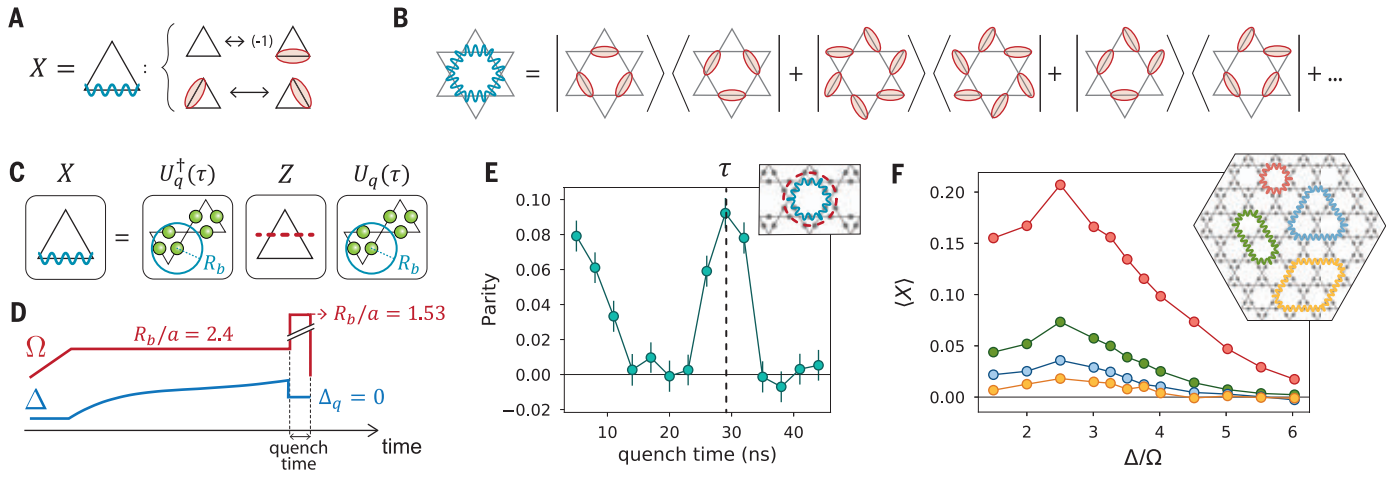


Fig. 3. Probing coherence between dimer states by means of off-diagonal string operator. (**A**) Definition of *X* string operator on a single triangle of the kagome lattice. (**B**) On any closed loop, the *X* operator maps any dimer covering into another valid dimer covering, so that $\langle X \rangle$ measures the coherence between pairs of dimer configurations. (**C**) The *X* operator is measured by evolving the initial state under the Hamiltonian (Eq. 1) with $\Delta = 0$ and reduced blockade radius to encompass only atoms within each individual triangle, implementing a basis rotation that maps *X* into *Z*. (**D**) In the experiment, after the state preparation, we set the laser detuning to $\Delta_q = 0$ and increased Ω to $2\pi \times 20$ MHz to reach $R_b/a = 1.53$. (**E**) By measuring the *Z* parity on the dual string (red) of a target *X* loop (blue) after a variable quench time, we identified the time τ for which the mapping in (C) is implemented. (**F**) We measured $\langle X \rangle$ for different final detunings of the cubic sweep and (inset) for different loop sizes and found that the prepared state has long-range coherence that extends over a large fraction of the array (*31*). The dual *Z* loops corresponding to the *X* loops shown in the inset are defined in fig. S3 and (*31*).

$\langle Z \rangle = (-1)^{\#\text{enclosed vertices}}$ (Fig. 2B). The presence of monomers or double-dimers reduces the effective contribution of each vertex, resulting in a reduced $\langle Z \rangle$.

To measure $\langle Z \rangle$ for different loop shapes (Figs. 2, C and D), we evaluated the string observables directly from single-shot images, averaging over many experimental repetitions and over all loops of the same shape in the bulk of the lattice (*31*). In the range of detunings where $\langle n \rangle \sim 1/4$, we clearly observed the emergence of a finite $\langle Z \rangle$ for all loop shapes, with the sign matching the parity of enclosed vertices, as expected for dimer states (Fig. 2B). The measured values were generally $|\langle Z \rangle| < 1$ and decreased with increasing loop size, suggesting the presence of a finite density of defects. Nevertheless, these observations indicate that the state we prepared was consistent with an approximate dimer phase.

We next explored quantum coherence properties of the prepared state. To this end, we considered the off-diagonal *X* operator, which acts on strings along the bonds of the kagome lattice. It is defined in Fig. 3A by its action on a single triangle (*23*). Applying *X* on any closed string maps a dimer covering to another valid dimer covering (for example, a loop around a single hexagon in Fig. 3B). A finite expectation value for *X* therefore implies that the state contains a coherent superposition of one or more pairs of dimer states coupled by that specific loop, which is a prerequisite for a QSL. The measurement of *X* can be implemented by performing a collective basis rotation, illustrated in Fig. 3C (*23*). This rotation was implemented through time evolution under the Rydberg Hamiltonian (Eq. 1) with $\Delta = 0$ and reduced blockade radius $R_b/a = 1.53$, so that only the atoms within the same triangle were subject to the Rydberg blockade constraint. Under these conditions, it was sufficient to consider the evolution of individual triangles separately, where each triangle can be described as a four-level system ($\triangle, \triangle, \triangle, \triangle$). Within this subspace, after a time $\tau = 4\pi/(3\Omega\sqrt{3})$, the collective three-atom dynamics realizes a unitary U_q that implements the basis rotation that transforms an *X* string into a dual *Z* string (*31*).

Experimentally, the basis rotation was implemented after the state preparation by quenching the laser detuning to $\Delta_q = 0$ and increasing the laser intensity by a factor of ~200 to reduce the blockade radius to $R_b/a = 1.53$ (Fig. 3D) (*31*). We calibrated τ by preparing the state at $\Delta/\Omega = 4$ and evolving under the quench Hamiltonian for a variable time. We measured the parity of a *Z* string that was dual to a target *X* loop and observed a sharp revival of the parity signal at $\tau \sim 30$ ns (Fig. 3E) (*23*). Fixing the quench time τ, we measured $\langle X \rangle$ for different values of the detuning Δ at the end of the cubic sweep (Fig. 3F) and observed a finite *X* parity signal for loops that extend over a large fraction of the array. These observations clearly indicate the presence of long-range coherence in the prepared state.

Probing spin liquid properties

The study of closed string operators showed that we prepared an approximate dimer phase with quantum coherence between dimer coverings. Although these closed loops are indicative of topological order, we needed to compare their properties with those of open strings to distinguish topological effects from trivial ordering—the former being sensitive to the topology of the loop (*32*–*34*). This comparison is shown in Fig. 4, D and E, and indicates several distinct regimes. For small Δ, we found that both *Z* and *X* loop parities factorize into the product of the parities on the half-loop open strings; in particular, the finite $\langle Z \rangle$ is a trivial result of the low density of Rydberg excitations. By contrast, loop parities no longer factorize in the dimer phase ($3 \lesssim \Delta/\Omega \lesssim 5$). Instead, the expectation values for both open string operators vanish in the dimer phase, indicating the nontrivial nature of the correlations measured with the closed loops (*31*). More specifically, topological ordering in the dimer-monomer model can break down either because of a high density of monomers, corresponding to the trivial disordered phase at small Δ/Ω, or owing to the lack of long-range resonances, corresponding to a valence bond solid (VBS) (*23*). Open *Z* and *X* strings distinguish the target QSL phase from these proximal phases: When normalized according to the definition from Fredenhagen and Marcu (FM) (Fig. 4, B and C) (*32*, *33*), vanishing expectation values for these open strings can be considered to be key signatures for the QSL. In particular, open *Z* strings have a finite expectation value when the dimers form an ordered spatial arrangement, as in the VBS phase. At the same time, open *X* strings create pairs

of monomers at their endpoints (Fig. 4A), so a finite $\langle X \rangle$ can be achieved in the trivial phase, where there is a high density of monomers. Therefore, the QSL can be identified as the only phase where both FM string order parameters vanish for long strings (*23*).

The measured values of the FM order parameters are shown in Fig. 4, F and G. We found that $\langle Z \rangle_{\mathrm{FM}}$ is compatible with zero over the entire range of Δ/Ω, where we observed a finite Z parity on closed loops, indicating the absence of a VBS phase (Fig. 4F), which is consistent with our analysis of density-density correlations (fig. S6) (*31*). At the same time, $\langle X \rangle_{\mathrm{FM}}$ converges toward zero on the longest strings for $\Delta/\Omega \gtrsim 3.3$ (Fig. 4G), indicating a transition out of the disordered phase. By combining these two measurements with the regions of nonvanishing parity for the closed Z and X loops (Figs. 2 and 3), we conclude that for $3.3 \lesssim \Delta/\Omega \lesssim 4.5$, our results constitute a direct detection of the onset of a QSL phase (Fig. 4, F and G, shaded area).

The measurements of the closed-loop operators in Figs. 2 and 3 show that $|\langle Z \rangle|, |\langle X \rangle| < 1$ and that the amplitude of the signal decreases with increasing loop size, which results from a finite density of quasiparticle excitations. Specifically, defects in the dimer covering such as monomers and double-dimers can be interpreted as electric (e) anyons in the language of lattice gauge theory (*23*). Because the presence of a defect inside a closed loop changes the sign of Z, the parity on the loop is reduced according to the number of enclosed e-anyons as $|\langle Z \rangle| = \left| \left\langle (-1)^{\#\text{enclosed } e\text{–anyons}} \right\rangle \right|$. The average number of defects inside a loop is expected to scale with the number of enclosed vertices—with the area of the loop—and we observed an approximate area-law scaling of $|\langle Z \rangle|$ for small loop sizes (Fig. 4H). However, for larger loops we observed a deviation from area-law scaling, closer to a perimeter law. This can emerge if pairs of anyons are correlated over a characteristic length scale smaller than the loop size [a discussion of the expected scaling is provided in (*31*)]. Pairs of correlated anyons that are both inside the loop do not change its parity because their contributions cancel out; they only affect $\langle Z \rangle$ when they sit across the loop, leading to a scaling with the length of the perimeter. These pairs can be viewed as resulting from the application of X string operators to a dimer covering (Fig. 4A), originating, for example, from virtual excitations in the dimer-monomer model (*31*) or from errors caused by state preparation and detection. State preparation with larger Rabi frequency (improved adiabaticity) results in larger Z parity signals and reduced e-anyon density (fig. S9).

A second type of quasiparticle excitation that could arise in this model is the so-called magnetic (m) anyon. Analogous to e-anyons, which live at the endpoints of open X strings (Fig. 4A), m-anyons are created by open Z strings and correspond to phase errors between dimer coverings (fig. S11) (*31*). These excitations cannot be directly identified from individual snapshots but are detected with the measurement of closed X loop operators. The perimeter law scaling observed in Fig. 4I indicates that m-anyons only appear in pairs with short correlation lengths (*31*). These observations highlight the prospects for using topological string operators to detect and probe quasiparticle excitations in the system.

Toward a topological qubit

To further explore the topological properties of the spin liquid state, we created an atom array with a small hole by removing three atoms on a central triangle (Fig. 5), which creates

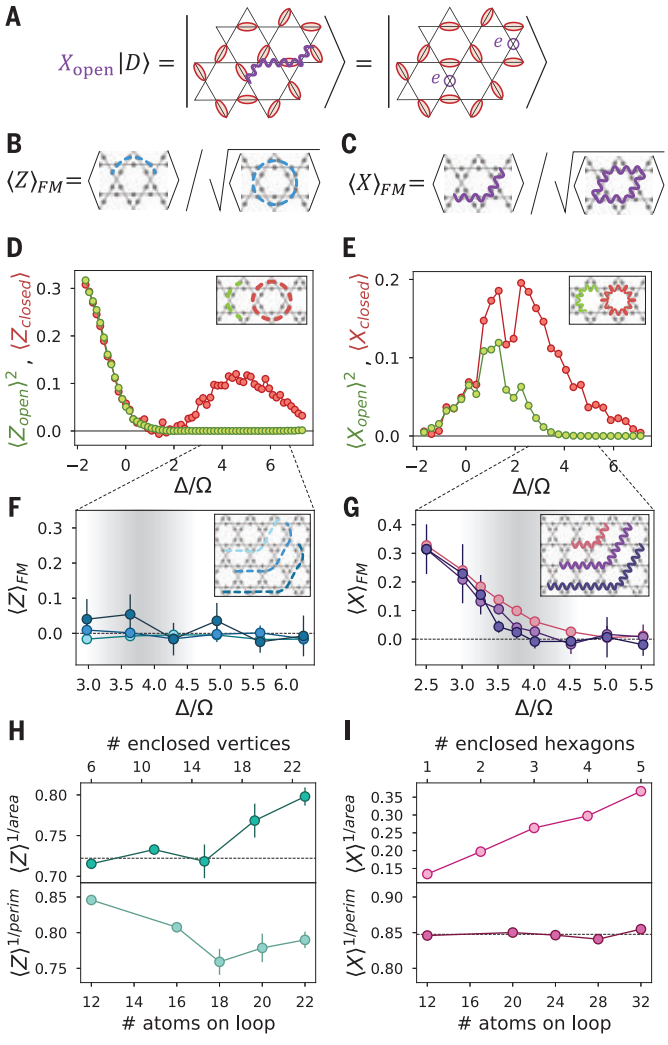


Fig. 4. String order parameters and quasi-particle excitations. (**A**) An open string operator X_{open} acting on a dimer state $|D\rangle$ creates two monomers (e-anyons) at its endpoints (m-anyons are shown in fig. S11). (**B** and **C**) Definition of the string order parameters $\langle Z \rangle_{\mathrm{FM}}$ and $\langle X \rangle_{\mathrm{FM}}$. (**D**) Comparison between $\langle Z_{\mathrm{closed}} \rangle$ and $\langle Z_{\mathrm{open}} \rangle^2$ measured on the strings shown in the inset. The expectation value shown for the open string is squared to account for a factor of two in the string lengths. (**E**) Analogous comparison for X. (**F** and **G**) Zooming in on the range with finite closed loop parities, we measured the FM order parameters for different open strings (insets). We found that $\langle Z \rangle_{\mathrm{FM}}$ is consistent with zero over the entire range of Δ, whereas $\langle X \rangle_{\mathrm{FM}}$ vanishes for $\Delta/\Omega \gtrsim 3.3$, which allowed us to identify a range of detunings consistent with the onset of a QSL phase (shaded area). (**H**) Rescaled parities $\langle Z \rangle^{1/\mathrm{area}}$ and $\langle Z \rangle^{1/\mathrm{perim}}$ evaluated for $\Delta/\Omega = 3.6$, where area and perimeter are defined as the number of vertices enclosed by the loop and the number of atoms on the loop, respectively. For small loops, Z scales with an area law but deviates from this behavior for larger loops, converging toward a perimeter law. (**I**) $\langle X \rangle^{1/\mathrm{area}}$ (the area, in this case, is the number of enclosed hexagons) and $\langle X \rangle^{1/\mathrm{perim}}$ evaluated for $\Delta/\Omega = 3.5$, indicating an excellent agreement with a perimeter-law scaling.

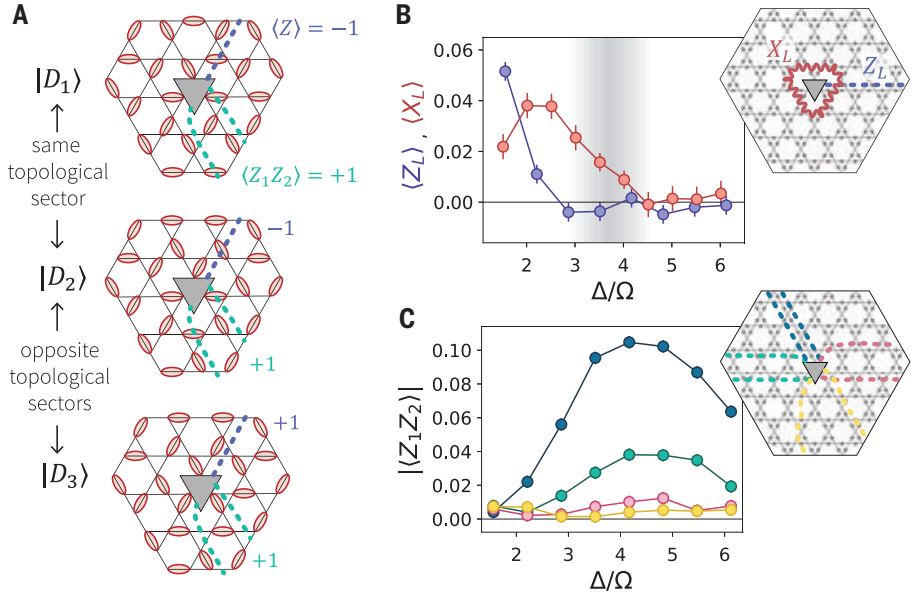


Fig. 5. Topological properties in array with a hole. (**A**) A lattice with nontrivial topology is obtained by removing three atoms at the center to create a small hole. The dimer states can be divided into two distinct topological sectors 0 and 1. Z strings connecting the hole to the boundary always have a well-defined expectation value within each sector and opposite sign between the two sectors; the correlations between two such strings Z_1Z_2 are identical for both sectors. (**B**) Measured expectation values for the operators Z_L and X_L, defined in the inset, indicate that in the QSL region (shaded area), we prepared a superposition state of the two topological sectors ($\langle Z_L\rangle = 0$) with a finite overlap with the $|+\rangle$ state ($\langle X_L\rangle > 0$). (**C**) Finite expectation values for the correlations between pairs of hole-to-boundary Z strings (inset), which is consistent with (A).

an effective inner boundary [both inner and outer boundaries here correspond to the so-called m-type boundaries (*31*)]. This resulted in two distinct topological sectors for the dimer coverings, where states belonging to different sectors can be transformed into each other only through large X loops that enclose the hole, constituting a highly nonlocal process (involving at least a 16-atom resonance) (fig. S13). We define the logical states $|0_L\rangle$ and $|1_L\rangle$ as the superpositions of all dimer coverings from the topological sectors 0 and 1, respectively. One can define (*23*) the logical operator σ_L^z as being proportional to any Z_L string operator that connects the hole with the outer boundary, given that these have a well-defined eigenvalue ±1 for all dimer states in the same sector but opposite for the two sectors. The logical σ_L^x is instead proportional to X_L, which is any X loop around the hole. This operator anticommutes with Z_L and has eigenstates $|+\rangle \sim (|0_L\rangle + |1_L\rangle)/\sqrt{2}$ and $|-\rangle \sim (|0_L\rangle - |1_L\rangle)/\sqrt{2}$.

We measured Z_L and X_L on the strings defined in Fig. 5B, inset, following the same quasi-adiabatic preparation as in Fig. 1D. We found that in the range of Δ/Ω associated with the onset of a QSL phase, $\langle Z_L\rangle = 0$, and $\langle X_L\rangle > 0$, indicating that the system is in a superposition of the two topological sectors, with a finite overlap with the $|+\rangle$ state (Fig. 5B), which is consistent with the symmetric initial state and the quasi-adiabatic preparation procedure (*31*). To further support this conclusion, we evaluated correlations $\langle Z_1Z_2\rangle$ between hole-to-boundary strings, which are expected to have the same expectation values for both topological sectors (Fig. 5A). In agreement with this prediction, we found that the correlations between different pairs of strings have finite expectation values, with amplitudes decreasing with the distance between the strings (Fig. 5C) owing to imperfect state preparation. These measurements represent the first steps toward initialization and measurement of a topological qubit in our system.

Discussion and outlook

It is not possible to classically simulate quantum dynamics for the full experimental system, so we compare our results with several theoretical approaches. First, our observations qualitatively disagree with the ground-state phase diagram obtained from density-matrix-renormalization-group (DMRG) (*35*, *36*) simulations on infinitely long cylinders. For the largest accessible system sizes, including van der Waals interactions only up to intermediate distances (~4a), we found a $\mathbb{Z}_2$ spin liquid in the ground state (fig. S15). However, unlike in deformed lattices (*23*), longer-range couplings destabilize the spin liquid in the ground state of the Hamiltonian (Eq. 1) on the specific ruby lattice used in the experiment, leading to a direct first-order transition from the disordered phase to the VBS phase (figs. S15 and S16). By contrast, we experimentally observed the onset of the QSL phase in a relatively large parameter range, and no signatures of a VBS phase were detected.

To develop additional insight, we performed time-dependent DMRG calculations (*35*–*37*) that simulated the same state preparation protocol as in the experiment on an infinitely long cylinder with a seven-atom-long circumference (*31*). The results of these simulations are in good qualitative agreement with our experimental observations (fig. S19). Specifically, similar to the results in Fig. 4, in the region $\frac{\Delta}{\Omega} \sim 3.5$ to 4.5 we found nonzero signals for closed Z and X loops, which cannot be factorized into open strings (fig. S19). Consistent with experimental observations, these indicate the onset of spin liquid correlations. In addition, exact diagonalization studies of a simplified blockade model reveal how the dynamical state preparation creates an approximate equal-weight and equal-phase superposition of many dimer states, instead of the VBS ground state (*31*). We conclude that quasi-adiabatic state preparation occurring over a few microseconds is insensitive to longer-range couplings and generates states that retain the QSL character (*31*). Although this phenomenon deserves further theoretical studies, these considerations indicate the creation of a metastable state with key characteristic properties of a QSL.

Our experiments offer detailed insights into elusive topological quantum matter. These studies can be extended along a number of directions, including improvement of the robustness of the QSL by using modified lattice geometries and boundaries (*22*, *23*) as well as optimization of the state preparation to minimize quasiparticle excitations; understanding and mitigation of environmental effects associated, for example, with dephasing and spontaneous emission (*31*); and optimization of string operator measurements by using quasi-local transformations (*38*), potentially with the help of quantum algorithms (*39*). At the same time, hardware-efficient techniques for robust manipulation and braiding of topological qubits can be explored. Furthermore, methods for anyon trapping and annealing can be investigated, with eventual applications toward fault-tolerant quantum information processing (*40*). With improved programmability and control, a broader class of topological quantum matter and lattice gauge theories can be efficiently implemented (*41*, *42*), opening the door to their detailed exploration under controlled experimental conditions and providing a route for the design of quantum materials that can supplement exactly

solvable models (*3*, *43*) and classical numerical methods (*35*, *36*).

Note added in proof: During the completion of this manuscript, we became aware of related work demonstrating the preparation of toric code states by using quantum circuits on a superconducting processor (*44*).

REFERENCES AND NOTES

1. X.-G. Wen, *Rev. Mod. Phys.* **89**, 041004 (2017).
2. S. Sachdev, *Rep. Prog. Phys.* **82**, 014001 (2019).
3. A. Kitaev, *Ann. Phys.* **303**, 2–30 (2003).
4. C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
5. L. Savary, L. Balents, *Rep. Prog. Phys.* **80**, 016502 (2017).
6. D. S. Rokhsar, S. A. Kivelson, *Phys. Rev. Lett.* **61**, 2376–2379 (1988).
7. N. Read, S. Sachdev, *Phys. Rev. Lett.* **66**, 1773–1776 (1991).
8. S. Sachdev, *Phys. Rev. B Condens. Matter* **45**, 12377–12396 (1992).
9. R. Moessner, S. L. Sondhi, *Phys. Rev. Lett.* **86**, 1881–1884 (2001).
10. G. Misguich, D. Serban, V. Pasquier, *Phys. Rev. Lett.* **89**, 137202 (2002).
11. B. I. Halperin, J. K. Jain, *Fractional Quantum Hall Effects* (World Scientific, 2020).
12. T.-H. Han *et al*., *Nature* **492**, 406–410 (2012).
13. M. Fu, T. Imai, T.-H. Han, Y. S. Lee, *Science* **350**, 655–658 (2015).
14. C. Gross, I. Bloch, *Science* **357**, 995–1001 (2017).
15. H. Weimer, M. Müller, I. Lesanovsky, P. Zoller, H. P. Büchler, *Nat. Phys.* **6**, 382–388 (2010).
16. M. Hermele, V. Gurarie, A. M. Rey, *Phys. Rev. Lett.* **103**, 135301 (2009).
17. N. Y. Yao *et al*., *Phys. Rev. Lett.* **110**, 185302 (2013).
18. A. W. Glaetzle *et al*., *Phys. Rev. X* **4**, 041037 (2014).
19. A. Celi *et al*., *Phys. Rev. X* **10**, 021057 (2020).
20. S. de Léséleuc *et al*., *Science* **365**, 775–780 (2019).
21. T. Ozawa *et al*., *Rev. Mod. Phys.* **91**, 015006 (2019).
22. R. Samajdar, W. W. Ho, H. Pichler, M. D. Lukin, S. Sachdev, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015785118 (2021).
23. R. Verresen, M. D. Lukin, A. Vishwanath, *Phys. Rev. X* **11**, 031005 (2021).
24. C. Song *et al*., *Phys. Rev. Lett.* **121**, 030502 (2018).
25. C. K. Andersen *et al*., *Nat. Phys.* **16**, 875–880 (2020).
26. S. Zhou, D. Green, E. D. Dahl, C. Chamon, *Phys. Rev. B* **104**, L081107 (2021).
27. M. Saffman, T. G. Walker, K. Mølmer, *Rev. Mod. Phys.* **82**, 2313–2363 (2010).
28. P. Anderson, *Mater. Res. Bull.* **8**, 153–160 (1973).
29. S. Ebadi *et al*., *Nature* **595**, 227–232 (2021).
30. P. Scholl *et al*., *Nature* **595**, 233–238 (2021).
31. Materials and methods are available as supplementary materials.
32. K. Fredenhagen, M. Marcu, *Commun. Math. Phys.* **92**, 81–119 (1983).
33. J. Bricmont, J. Frölich, *Phys. Lett. B* **122**, 73–77 (1983).
34. K. Gregor, D. A. Huse, R. Moessner, S. L. Sondhi, *New J. Phys.* **13**, 025009 (2011).
35. S. R. White, *Phys. Rev. Lett.* **69**, 2863–2866 (1992).
36. J. Hauschild, F. Pollmann, *SciPost Phys. Lect. Notes* p. 5 (2018).
37. M. P. Zaletel, R. S. K. Mong, C. Karrasch, J. E. Moore, F. Pollmann, *Phys. Rev. B Condens. Matter Mater. Phys.* **91**, 165112 (2015).
38. M. B. Hastings, X.-G. Wen, *Phys. Rev. B Condens. Matter Mater. Phys.* **72**, 045141 (2005).
39. I. Cong, S. Choi, M. D. Lukin, *Nat. Phys.* **15**, 1273–1278 (2019).
40. E. Dennis, A. Kitaev, A. Landahl, J. Preskill, *J. Math. Phys.* **43**, 4452–4505 (2002).
41. M. C. Bañuls *et al*., *Eur. Phys. J. D* **74**, 165 (2020).
42. L. Tagliacozzo, A. Celi, A. Zamora, M. Lewenstein, *Ann. Phys.* **330**, 160–191 (2013).
43. A. Kitaev, *Ann. Phys.* **321**, 2–111 (2006).
44. K. J. Satzinger *et al*., *Science* **374**, 1237–1241 (2021).
45. G. Semeghini *et al*., Probing topological spin liquids on a programmable quantum simulator. *Harvard Dataverse* (2021); doi:10.7910/DVN/BDCTRX.
46. R. Verresen, rubenverr/DMRG-QSL-Rydberg-experiment. Zenodo (2021); doi:10.5281/zenodo.5553679.

ACKNOWLEDGMENTS

We thank many members of the Harvard AMO community, particularly E. Urbach, S. Dakoulas, and J. Doyle for their efforts enabling operation of our laboratories during 2020-2021. We thank S. Choi, I. Cong, E. Demler, G. Giudici, W. W. Ho, N. Maskara, K. Najafi, N. Yao, and S. Yelin for stimulating discussions. **Funding:** We acknowledge financial support from the Center for Ultracold Atoms, the National Science Foundation, the U.S. Department of Energy (DE-SC0021013 and LBNL QSA Center), the Army Research Office MURI, the DARPA ONISQ program, QuEra Computing, and Amazon Web Services. We further acknowledge support from the Max Planck/Harvard Research Center for Quantum Optics fellowship (to G.S.), the National Defense Science and Engineering Graduate (NDSEG) fellowship (to H.L.), Gordon College (to T.T.W.), the NSF Graduate Research Fellowship Program (grant DGE1745303) and The Fannie and John Hertz Foundation (to D.B.), the Harvard Quantum Initiative Postdoctoral Fellowship in Science and Engineering (to R.V.), the Simons Collaboration on Ultra-Quantum Matter (Simons Foundation grant 651440 to R.V., A.V., and S.S.). R.S. and S.S. were supported by the U.S. Department of Energy under grant DE-SC0019030. The DMRG simulations were performed by using the Tensor Network Python (TeNPy) package developed by J. Hauschild and F. Pollmann (*36*) and were run on the FASRC Cannon and Odyssey clusters supported by the FAS Division of Science Research Computing Group at Harvard University. **Author contributions:** G.S., H.L., A.K., S.E., T.T.W., D.B., and A.O. contributed to building the experimental setup, performed the measurements, and analyzed the data. R.V., H.P., and A.V. contributed to developing methods for detecting QSL correlations, performed numerical simulations, and contributed to the theoretical interpretation of the results. M.K. and R.S. contributed to the theoretical interpretation of the results. All work was supervised by S.S., M.G., V.V., and M.D.L. All authors discussed the results and contributed to the manuscript. **Competing interests:** M.G., V.V., and M.D.L. are cofounders and shareholders of QuEra Computing. A.K. is CEO and shareholder of QuEra Computing. A.O. is shareholder of QuEra Computing. Some of the techniques and methods used in this work are included in provisional and pending patent applications filed by Harvard University (US patent application nos. 16/630719, 63/116,321, and 63/166,165). **Data and materials availability:** The data and code are available on Harvard Dataverse (*45*) and Zenodo (*46*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi8794
Materials and Methods
Figs. S1 to S19
References (*47*–*56*)

8 April 2021; accepted 28 October 2021
10.1126/science.abi8794

PLANT SCIENCE

Ground tissue circuitry regulates organ complexity in maize and *Setaria*

Carlos Ortiz-Ramírez[1,2], Bruno Guillotin[1]†, Xiaosa Xu[3]†, Ramin Rahni[1]†, Sanqiang Zhang[1], Zhe Yan[4]‡, Poliana Coqueiro Dias Araujo[1]§, Edgar Demesa-Arevalo[3], Laura Lee[1], Joyce Van Eck[5,6], Thomas R. Gingeras[3], David Jackson[3], Kimberly L. Gallagher[4], Kenneth D. Birnbaum[1]*

Most plant roots have multiple cortex layers that make up the bulk of the organ and play key roles in physiology, such as flood tolerance and symbiosis. However, little is known about the formation of cortical layers outside of the highly reduced anatomy of *Arabidopsis*. Here, we used single-cell RNA sequencing to rapidly generate a cell-resolution map of the maize root, revealing an alternative configuration of the tissue formative transcription factor SHORT-ROOT (SHR) adjacent to an expanded cortex. We show that maize SHR protein is hypermobile, moving at least eight cell layers into the cortex. Higher-order *SHR* mutants in both maize and *Setaria* have reduced numbers of cortical layers, showing that the *SHR* pathway controls expansion of cortical tissue to elaborate anatomical complexity.

Roots are radially symmetrical organs composed of three fundamental tissue types: the epidermis on the outside, the ground tissue at the middle, and a core of vascular elements plus pericycle that lie in a central cylinder known as the stele (*1*). The ground tissue is further divided into two different cell types, the endodermis and cortex, which are arranged as concentric layers around the stele. Variations in ground tissue patterning, particularly the number of cortex cell layers, are common across species and represent one of the defining features giving rise to interspecies root morphological diversity (*1*). This diversity allows plants to cope with biotic and abiotic stress and adapt to challenging environments. For example, maize cortex plays a role in drought and flood tolerance and hosts colonization of beneficial mycorrhizal associations that reduce stress and improve nutrient uptake (*2*–*5*). Therefore, an ongoing question in root biology is how tissue patterning is adjusted to produce divergent root morphologies, and what alterations in the genetic networks control developmental differences among species.

A current limitation is that knowledge of radial patterning mechanisms in roots comes largely from the study of *Arabidopsis*, which

[1]Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA. [2]UGA Laboratorio Nacional de Genómica para la Biodiversidad, CINVESTAV Irapuato, Guanajuato 36821, México. [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [4]School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 1904, USA. [5]Boyce Thompson Institute, Ithaca, NY 14853, USA. [6]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA.
***Corresponding author. Email: ken.birnbaum@nyu.edu**
†These authors contributed equally to this work.
‡Present address: The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI)/Key Lab of Germplasm Utilization (MOA), Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.
§Present address: Department of Agronomic and Forest Sciences, Universidade Federal Rural do Semi-Árido, RN, Brazil.

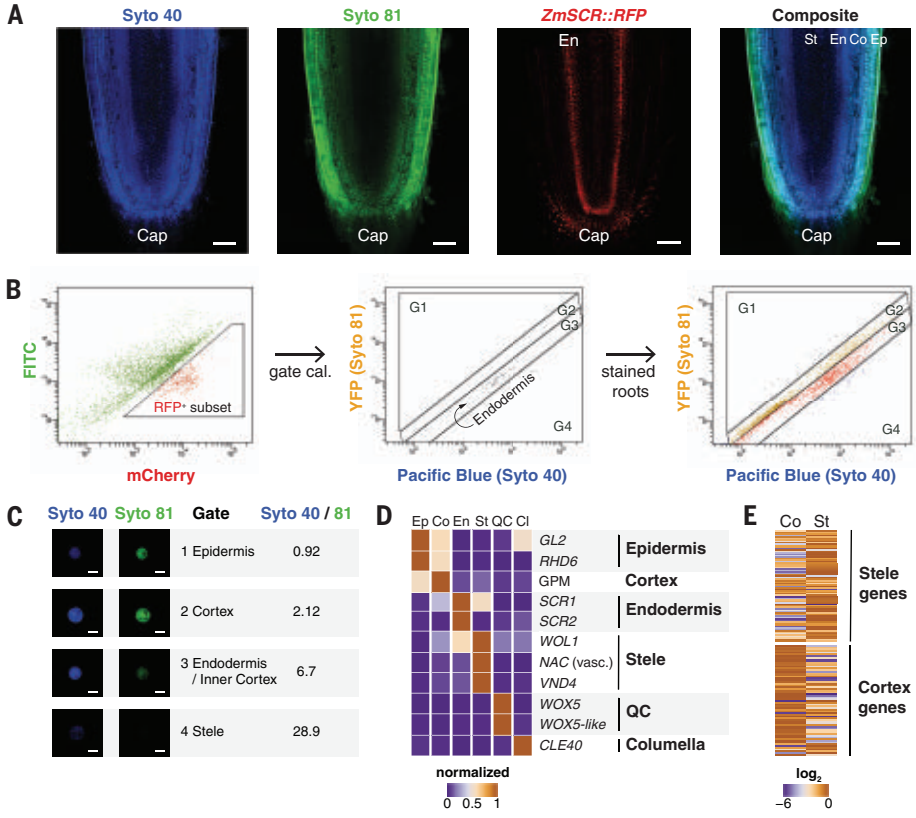


Fig. 1. Dye penetrance labeling (DPL) and tissue transcriptome analysis in maize. (**A**) Representative images of a deeply penetrating dye (Syto 40), a superficially penetrating dye (Syto 81), the *ZmSCR*::RFP marking endodermis, and a composite image of Syto 40 and Syto 81 staining. (**B**) Cell sorting gating strategy, showing the *ZmSCR*::RFP population (left), backgated onto a YFP versus Pacific blue plot with RFP positive (middle), and (right) the gated boundaries for endodermal, outside of endodermis (G1, G2), and inside of endodermis (G4). FITC, fluorescein-5-isothiocyanate. (**C**) Validation of ratiometric cell sorting strategy by collecting sorted cells from gates and quantifying fluorescence from microscopy images. (**D**) Validation of sorted cell RNA-seq profiles by analysis of known markers. (**E**) Global validation comparing sorted cells versus mechanically dissected stele and cortex tissues, with heat map showing expression in sorted cortex versus stele gates, categorized by previously determined stele and cortex markers. Scale bars, 100 μm (A) and 15 μm (C).

possesses a simple root anatomy. In *Arabidopsis*, only two ground tissue layers develop in primary development—one endodermal and one cortical—that originate from an asymmetric cell division at or near the initials or stem cells (*6*). This division is controlled by the *SHORT-ROOT* (*SHR*)–*SCARECROW* (*SCR*) genetic pathway (*7*, *8*). Mutants in either transcription factor develop a monolayered ground tissue. In addition, SHR mutants lack an endodermis, giving *SHR* a role in both tissue formation and cell identity (*9*).

SHR functions as a mobile signal whose protein travels from the stele, where the gene is transcribed, into the surrounding endodermis, where it induces the expression of the downstream transcription factor *SCR* (*8*). The pathway then triggers the division that generates the cortex and endodermis layers (*9*). In maize, the *SHR-SCR* pathway has been implicated with a role in leaf anatomy (*10*–*13*). Movement of SHR from the stele further out into the cortex could potentially cause extra cell divisions, giving rise to multiple cortex layers (*12*). Supporting this idea, the SHR-SCR pathway was recently implicated in cortical cell division during nodule formation in the dicot *Medicago* (*14*). However, the role of *SHR* in the expansion of cortical layer number in maize, a monocot, is not known.

Dye penetrance labeling for rapid tissue profiling

We sought to produce a high-resolution spatial and temporal map of gene expression in a complex root that could provide clues to the genetic networks controlling morphological diversity in patterning. Therefore, we generated cell type–specific gene expression profiles using high-throughput single-cell RNA sequencing (scRNAseq) to profile maize roots. Maize is a valuable model for comparative studies because its roots develop multilayered cortical tissues (8 to 9 cortex cell layers) within the root meristem and it is amenable to protoplast generation, an essential step in plant scRNAseq (*15*). However, a challenge of scRNAseq studies in species for which genomic resources are limited is the correct identification of cell types. The use of homologs of *Arabidopsis* markers obtained by high-throughput cell sorting did not provide a clear identification of morphologically homologous cell types in maize. This is likely because gene orthology is not always apparent and localization over such broad evolutionary distance is not well conserved.

To overcome this problem, we first took advantage of the concentric arrangement of tissues in roots to develop a technique to fluorescently mark cell layers by dye penetrance labeling (DPL). In brief, a highly penetrant dye (Syto 40 blue) stains the entire root with low but detectable staining in stele, whereas a weakly penetrant dye (Syto 81 green) stains the outer tissue layers strongly, with a gradual drop in signal intensity toward the inner tissues (Fig. 1A). This dual labeling was reproducible across roots and batches and had a negligible effect on transcription (fig. S1). The approach allowed us to enrich for different concentric tissue layers using blue/green ratios in fluorescence-activated cell sorting (FACS). We calibrated dye ratio to radial position by using DPL on a line expressing a fluorescent protein driven by the *SCR* promoter (*ZmSCR1*::*RFP*; Fig. 1, A and B), which expresses in the endodermis. Red fluorescent protein (RFP)–positive cells were used to calibrate a reference dye ratio for this middle layer, allowing demarcation of inner and outer tissues (Fig. 1, B and C, and fig. S2A). We dissected seminal root tips (5 mm and, in one subset, 5 to 8 mm from the root tip) and then rapidly enzymatically digested their cell walls, sorting cells belonging to different tissue layers using their specific dye ratio. We also generated a set of whole meristem protoplasts versus intact root controls to filter out any effects of cell wall digestion. Digested and undigested controls clustered closely together, and replicate samples yielded consistent profiles (fig. S2, B to D). In addition, we obtained expression profiles of the root cap by dissection and quiescent center (QC), using FACS on a stable QC marker line, *ZmWOX5*::RFP (fig. S2E). To validate the entire dataset, we compared the six tissue expression profiles to known markers and to a previous study that used mechanical separation of inner versus outer layers (*16*), finding 80% agreement or higher (Fig. 1, D and E). We also used a panel of conserved and well-characterized markers to validate sorted-cell profiles, showing close agreement (fig. S3 and table S1). In this manner, we developed a set of at least 170 markers for each radially arranged tissue (table S2).

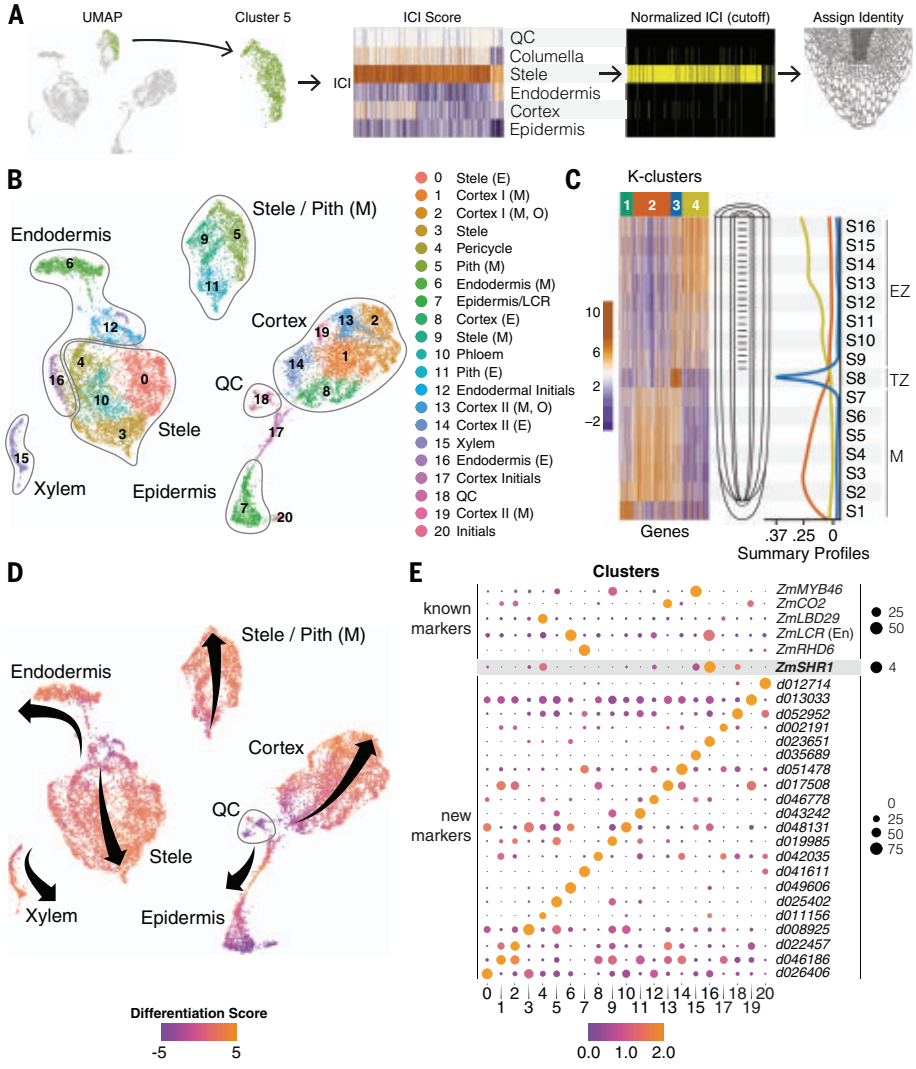


Fig. 2. Single-cell RNA-seq spatial and temporal transcriptome maps of the maize meristem. (**A**) The ICI method of diagnosing cell identity of UMAP clusters using known markers and randomization testing (e.g., cluster 5). (**B**) Cluster identities as determined by ICI and cell-type specific markers (E, early; M, mature; O, outer; Pith, pith parenchyma). (**C**) Heat map of highly variant genes along a longitudinal axis of the root meristem. Developmental patterns show transcripts and markers that peak in the early meristem (M), transition zone (TZ), and elongation zone (EZ). (**D**) Trajectories of developmental "pseudo-time" in each cell cluster mapped onto the same UMAP depicted in (B), where a differentiation score is calculated as a $\log_2$ ratio of all EZ/M markers identified in (C). Arrow origins for each cluster represent cells in the meristem near the stem cells progressing to more proximal cells near the arrowheads. (**E**) Select known (top) and new markers (bottom) for each cluster. Size of spot represents percent of cells in cluster expressing the marker, and color represents their relative expression level in those cells.

A single-cell map of the maize root

To generate a single-cell resolution map of the maize root meristem, we then dissected seminal root tips from 7-day-old wild-type B73 maize seedlings and enzymatically digested their cell walls, as above. We then used the cells to prepare single-cell cDNA libraries using the 10x Genomics Chromium platform. A total of 14,755 high-quality cells were sequenced in three different batches with a mean of 31,105 unique molecular indices per cell and 5683 detected genes per cell (fig. S4). A total of 21 cell clusters were defined and visualized in two dimensions in Seurat using the uniform manifold approximation and projection (UMAP) method (*17*). To quantify cell identity and classify clusters using the DPL markers, we applied the Index of Cell Identity (ICI) algorithm (*18*), which generates a cell identity score based on the mean expression of a predefined marker gene set, in this case, from FACS-isolated tissues (Fig. 2A and figs. S5 and S6). Overall, the technique allowed us to identify all the UMAP clusters, providing a detailed spatial map of transcripts in specific cell types of the maize root (Fig. 2B).

The high-resolution cellular map of the meristem showed multiple cell type subclusters within the stele and cortex, suggesting cellular specialization within the latter tissue's multiple layers. However, because root cells differentiate as they transition away from the root tip, the possibility remained that some subclusters represented different maturation states of an individual cell type. To distinguish subclusters formed by distinct identity rather than differentiation state, we further generated a set of cell maturation marker transcripts by dissecting 16 longitudinal root slices that together comprised the meristematic, transition, and elongation zones, and we subjected the samples to RNAseq analysis (Fig. 2C). Using K-means clustering, we identified three main expression programs: early meristematic (high expression in the meristem and gradual decrease toward the transition zone), transition zone (specific to the mid-maturation point), and posttransition (low expression in the meristematic zone and gradual increase toward the elongation zone). We then generated a cellular differentiation score to label the maturation status of each cell, resolving developmental trajectories of cells in our high-resolution map of cell identities (Fig. 2D).

High-resolution profiling reveals cortical complexity

In a few cases, the state of cell maturation is indeed the main factor influencing grouping of cells into subclusters. For example, five clusters (8, 11, 12, 14, and 16) had the same identity but a different maturation state compared with adjacent clusters. However, the majority of subclusters were composed of cells with a wide range of differentiation states, showing that the grouping in most cases represented distinct cell identities. Although two cortex subclusters appeared to be a precursor state of mature cortex (clusters 8 and 14), our analysis confirmed the existence of at least four distinct cortex subtypes (clusters 1, 2, 13, and 19; Fig. 2B). Furthermore, using the receiver operating characteristic analysis in Seurat, which identified 2436 differentially expressed genes (DEGs) across all clusters, we found 471 transcripts that mark some subset of the four different cortical subtypes (Fig. 2E and table S3). Thus, we provide quantitative evidence for the subspecialization of cortex that underlies expansion of root complexity, yielding strong evidence for cortical cell diversification.

One question that follows is what signaling mechanisms allow maize to form the extra layers that permit cortex subspecialization. We observed that a short list of functional markers with a role in patterning or cell identity in *Arabidopsis* had conserved localization in homologous tissues in maize (e.g., *CO2*, cortex; *MYB46*, xylem; *RHD6*, epidermis;

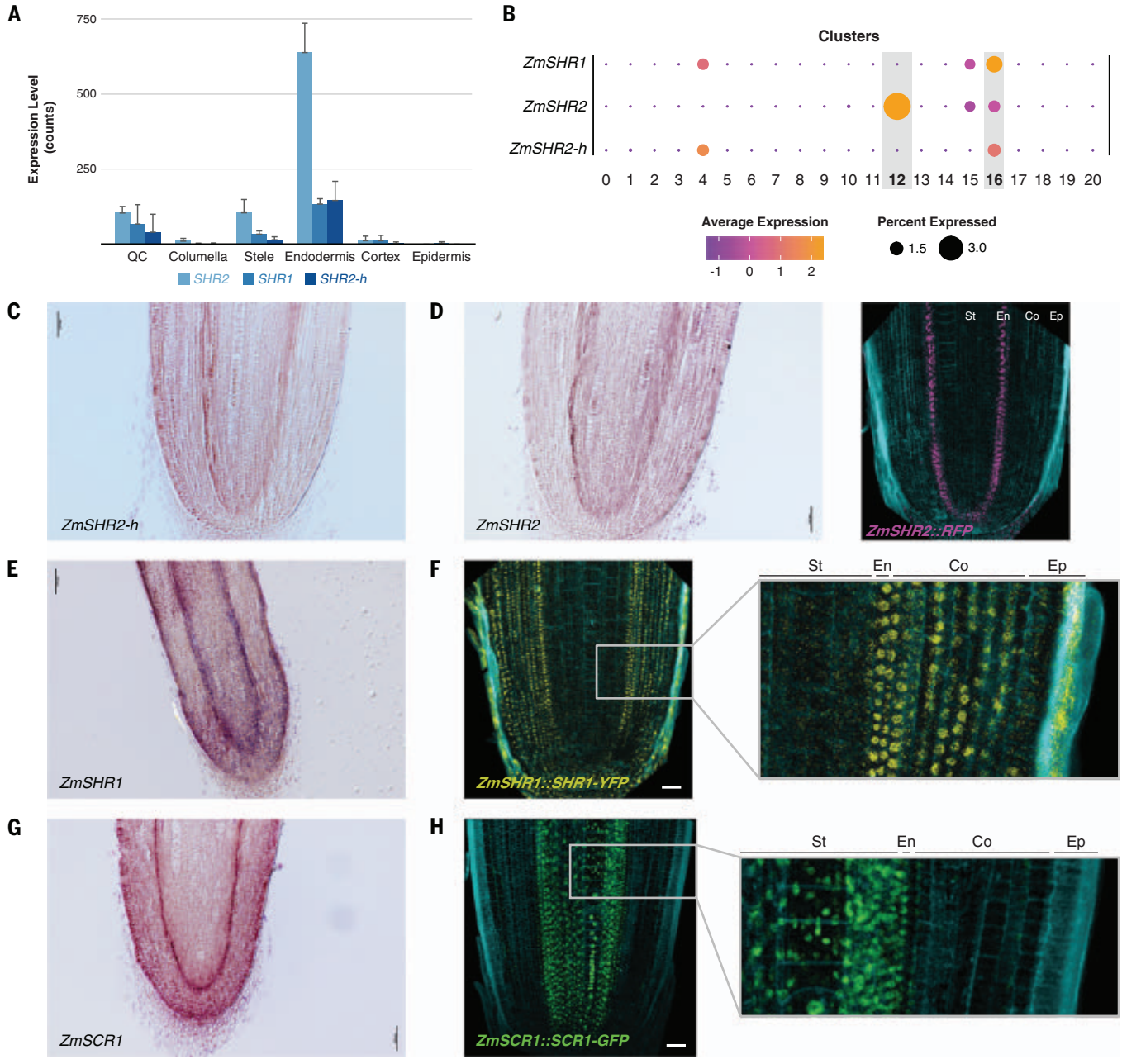


Fig. 3. SHR and SCR expression in maize endodermis and differences between transcriptional and translational reporters. (**A**) Expression of the three *ZmSHR* paralogs (*SHR1,2,2-h*) from sorted cells. Error bars are SDs. (**B**) Dot plots representing scRNAseq analysis showing expression of *ZmSHR* paralogs in endodermis and, at a low level, stele. (**C**) *ZmSHR2-h* in situ hybridization showing endodermal expression. (**D**) *ZmSHR2* in situ hybridization and transcriptional RFP reporter. (**E**) In situ hybridization of *ZmSHR1*. (**F**) Translational YFP reporter for *Zm*SHR1 with inset showing protein in cortex layers. (**G**) In situ hybridization of *ZmSCR1*. (**H**) Translational reporter SCR1-GFP reporter showing protein in the stele. Scale bars, 50 μm [(F) and (H)] and 100 μm [(C), (D), (E), and (G)]. Cyan in confocal images [(D), (F), (H)] is autofluorescence at an excitation wavelength of 405 nm.

and *LBD29*, stele) (Fig. 2E). However, our scRNAseq data showed that expression of the core patterning gene *SHR* was specific to the endodermis and not to the stele (Fig. 2E), where the *Arabidopsis* ortholog is expressed (*9*). All three maize paralogs of *SHR* (designated *ZmSHR1*, *ZmSHR2*, and *ZmSHR2-h*) showed the same endodermal enrichment in the profiles generated by DPL and single-cell analysis (Fig. 3, A and B). We speculated that the expression of this mobile, division-inducing transcription factor adjacent to the cortex could be related to a role in the expansion of that tissue.

Mobile SHR regulates cortical complexity in maize and *Setaria*

To confirm SHR transcript localization, we performed in situ hybridization on all three *SHR* mRNAs, confirming their localization in the endodermis (Fig. 3, C to E, and fig. S7, A to I). A 1.4-kb upstream and downstream transcriptional reporter for *ZmSHR2* also showed signal in the root endodermis, in agreement with our dye-sorted and single-cell profiles (Fig. 3D). No signal was found in the stele, confirming that *SHR2* transcript localizes to different cell types in maize compared with *Arabidopsis*. Consistently,

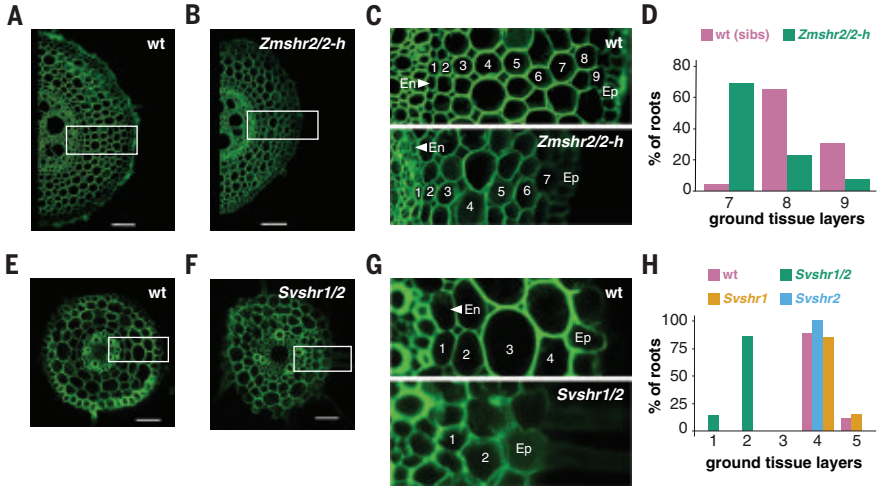


Fig. 4. Cortical cell layer analysis in wild-type and *shr* mutants in monocots. (**A** and **B**) Representative maize root cross sections of wild-type (sibs) (A) versus *Zmshr2/2-h* double homozygous mutant (B). (**C**) Enlarged regions from boxes in (A) and (B) showing endodermal (Ed), cortical (numbered), and epidermal (Ep) layers of wild-type (top) and *Zmshr2/2-h* mutant (bottom). (**D**) Quantification of the cortical cell layers in wild-type and heterozygous sibs (sample size (*n*) = 23) versus *Zmshr2/2-h* mutant (*n* = 13, $p < 0.001$, Mann-Whitney rank test). (**E** and **F**) Representative cross sections of *Setaria* wild type (E) and *Svshr1/2* mutants (F). (**G**) Enlarged regions from dashed boxes in (E) and (F) showing endodermal (Ed), cortical (numbered), and epidermal (Ep) layers of wild type (top) and *Svshr1/2* mutant (bottom). (**H**) Quantification of cortical layers in *Setaria* wild type (*n* = 9), *Svshr1* single (*n* = 7), *Svshr2* single (*n* = 6), and *Svshr1/2* double mutants (*n* =7, $p < 0.001$, Tukey test after one-way analysis of variance on all genotypes,). Scale bars, 100 μm [(A) and (C)] and 50 μm [(E) and (F)]. Green is autofluorescence at an excitation wavelength of 405 nm and emission wavelengths of 510 to 535 nm.

the *ZmSHR* endodermal domain in the root is reminiscent of *ZmSHR* expression in the shoot bundle sheath (*19*, *20*), which is in an analogous position to root endodermis. In addition, expression of *SHR* orthologs was shown to localize to the endodermis in date palm (*21*).

Given evidence that rice SHR proteins are hypermobile when expressed heterologously in *Arabidopsis* (*8*), we assessed whether maize SHR protein could move from the endodermis, where its mRNA is expressed, into the adjacent cortex. For this purpose, we made a natively expressed protein reporter of ZmSHR1 fused to yellow fluorescent protein (*ZmSHR1*::*SHR1*-YFP). Indeed, compared to endodermal localization of the *ZmSHR1* transcript, the maize SHR1 protein reporter was present in the cortex (Fig. 3F). Moreover, SHR1-YFP signal was not restricted to the adjacent tissue layer as in *Arabidopsis* but was observed in all cortex layers, suggesting that the endogenous maize SHR1 protein moves through at least eight cortex layers (Fig. 3F and fig. S7E). In addition, the ZmSHR1 protein also appeared to be hypermobile when expressed in the endodermis of *Arabidopsis* (fig. S8, A and B). Whether additional SHR paralogs are mobile is an open question.

SHR's role in promoting division in *Arabidopsis* works through direct interaction with SCR (*22*). Therefore, we also generated maize *SCR* reporter lines to determine colocalization with *SHR*. Both in situ localization and a promoter RFP reporter revealed a strong signal in the root endodermis, as shown previously (*23*) and similar to its localization in *Arabidopsis* (Fig. 3G and fig. S7, F to I). In addition, we observed low-level cortical expression of *SCR1* in both the scRNAseq data and high-sensitivity confocal imaging (e.g., fig. S7G). However, natively expressed ZmSCR-GFP protein fusions showed a signal in the stele, suggesting that SCR protein in maize moves from cell to cell in the opposite direction from that of SHR (Fig. 3H and fig. S8C). Our results show that SHR and SCR colocalize in the endodermis and possibly in the cortex. A SCR translational reporter in a second monocot, *Setaria viridis* (*Setaria*), showed the same localization in the stele, further corroborating the divergent localization of SCR protein in monocots (fig. S9, A and B).

The model that implicates SHR in cortical expansion posits that increased outward movement of the protein could trigger periclinal cell divisions giving rise to extra ground tissue layers (*12*). To test the model, we targeted the three different maize *SHR* paralogs to generate loss-of-function mutants in maize using CRISPR-Cas9 (fig. S10, A and B). We recovered mutants in two of the genes (*ZmSHR2* and *ZmSHR2-h*), including the most highly expressed paralog, *ZmSHR2*. Single mutants in *Zmshr2* or *Zmshr2-h* had no difference in their root anatomy compared to wild-type siblings. However, *Zmshr2/2-h* double mutants had a significant reduction in the number of cortical layers, with most roots having seven layers compared with eight or nine layers in wild type (Fig. 4, A to D). Mutants in the single *SHR* gene in *Arabidopsis* lack an endodermal layer. However, staining for suberin and morphological analysis showed that *Zmshr2/2-h* roots still developed an endodermis (fig. S11). We posit that the remaining functional *ZmSHR1* gene in the *Zmshr2/2-h* background may still enable specification of endodermal identity. Alternatively, *ZmSCR1* (and *ZmSCR2*) may be the primary factors in the specification of the maize endodermis (*13*). Overall, our results suggest that *SHR* function in maize is necessary for the full expansion of cortical identity.

We sought to validate the monocot *SHR* mutant phenotype with a more severe loss of function by testing its role in *S. viridis*, a close relative of maize. In *Setaria*, we were able to generate loss-of-function mutants in the two *SHR* orthologs using CRISPR-Cas9 (fig. S12, A and B). One *Setaria SHR* mutant, *Svshr2*, showed a slight reduction in cortical layers, and a single mutant in the second, *Svshr1*, showed no phenotype. However, double mutants showed a marked reduction in the number of ground tissue layers, resulting in one to two layers compared to four to five layers in wild-type siblings (Fig. 4, E to H). These results corroborate the role of *SHR* in controlling the expansion of cortical layers in two monocots.

The extra cortical divisions mediated by SHR could function through direct interaction with SCR by mediating successive divisions of the endodermis near the stem cell niche. Alternatively, SHR hypermobility could lead to divisions directly in the cortical layers possibly interacting with low levels of SCR or another protein. At present, we cannot distinguish between these two models.

The results show that SHR has a role in monocots in controlling the expansion of cortex, which sets up many traits for environmental acclimation. This illustrates how subtle divergence of a conserved developmental regulator can mediate anatomical complexity that has given rise to specialized functions. Related to the complexity of the root, we identify four distinct cortical cellular subtypes in our bioinformatic analysis, although further work is needed to verify their spatial relationship. Finally, the results show that rapid transcriptome mapping using single-cell dissection can provide insights into the mechanisms that mediate anatomical diversity. The use of dye labeling to generate a scaffold locational map together with scRNAseq provides a maize root tissue map that can be used as a reference in maize and related plants.

REFERENCES AND NOTES

1. K. Esau, *Anatomy of Seed Plants* (Wiley, ed. 2, 1977).
2. M. Kaldorf, A. J. Kuhn, W. H. Schröder, U. Hildebrandt, H. Bothe, *J. Plant Physiol.* **154**, 718–728 (1999).

3. F. Liu *et al*., *Plant Cell Physiol.* **59**, 1683–1694 (2018).
4. J. Zhu, K. M. Brown, J. P. Lynch, *Plant Cell Environ.* **33**, 740–749 (2010).
5. A. H. Gunawardena, D. M. Pearce, M. B. Jackson, C. R. Hawes, D. E. Evans, *Planta* **212**, 205–214 (2001).
6. L. Dolan *et al*., *Development* **119**, 71–84 (1993).
7. P. N. Benfey *et al*., *Development* **119**, 57–70 (1993).
8. K. Nakajima, G. Sena, T. Nawy, P. N. Benfey, *Nature* **413**, 307–311 (2001).
9. Y. Helariutta *et al*., *Cell* **101**, 555–567 (2000).
10. T. L. Slewinski *et al*., *Mol. Plant* **7**, 1388–1392 (2014).
11. T. L. Slewinski, A. A. Anderson, C. Zhang, R. Turgeon, *Plant Cell Physiol.* **53**, 2030–2037 (2012).
12. S. Wu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16184–16189 (2014).
13. T. E. Hughes, O. V. Sedelnikova, H. Wu, P. W. Becraft, J. A. Langdale, *Development* **146**, dev.177543 (2019).
14. W. Dong *et al*., *Nature* **589**, 586–590 (2021).
15. A. Senn, P.-E. Pilet, *Z. Pflanzenphysiol.* **100**, 299–310 (1980).
16. N. Opitz *et al*., *J. Exp. Bot.* **67**, 1095–1107 (2016).
17. T. Stuart *et al*., *Cell* **177**, 1888-1902 e1821 (2019).
18. I. Efroni, P. L. Ip, T. Nawy, A. Mello, K. D. Birnbaum, *Genome Biol.* **16**, 9 (2015).
19. Y. M. Chang *et al*., *Plant Physiol.* **160**, 165–177 (2012).
20. X. Xu *et al*., *Dev. Cell* **56**, 557–568.e6 (2021).
21. T. T. Xiao *et al*., *Plant Cell* **31**, 1751–1766 (2019).
22. H. Cui *et al*., *Science* **316**, 421–425 (2007).
23. J. Lim *et al*., *Plant Cell* **12**, 1307–1318 (2000).

ACKNOWLEDGMENTS

Funding: K.D.B and D.J. are supported by NSF grant IOS-1934388. K.D.B, D.J., and T.R.G. were supported by NSF grant IOS 1445025. K.D.B. is funded by NIH grant R35GM136362. D.J. is funded by NSF IOS-1930101. K.L.G. is funded by NSF PGRP-23020. B.G. is funded by the Human Frontier Science Program Organization grant: LT - 000972/2018. **Author contributions:** C.O.-R., B.G., S.Z., and L.L. performed all transcriptomic experiments and expression analysis. C.O.R. performed mutant analysis. C.O.R. and K.D.B. designed the experiments and wrote the manuscript. C.O.R., K.D.B., D.J., T.R.G., and K.L.G. conceived the project and guided the experiments. X.X. performed in situ hybridizations. P.C.D.A. and S.Z. assisted in mutant analysis and marker analysis. R.R. designed and carried out microscopy protocols and designed graphic layouts. E.D.-A. and X.X. generated transcriptional reporters in maize. C.O.R., Z.Y., and J.V.E generated CRISPR-Cas9 knockouts and translational reporters in maize and *Setaria* and generated the *Setaria* mutants. **Competing interests:** The authors declare no competing interests. **Data materials and availability:** All data are included in the main paper or the supplementary materials or are deposited at the Gene Expression Omnibus website (https://www.ncbi.nlm.nih.gov/geo/) under the SuperSeries accession number GSE172302. *pZmWOX5::RFP*, *pZmSCR1::RFP*, and *pZmSHR2::RFP* reporter lines are available from D.J. under a material transfer agreement with Cold Harbor Spring Laboratory.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj2327
Materials and Methods
Figs. S1 to S13
Tables S1 to S3
References (*24*–*30*)
MDAR Reproducibility Checklist

28 April 2021; accepted 13 October 2021
10.1126/science.abj2327

REPORTS

MOLECULAR BIOLOGY

Error-prone, stress-induced 3′ flap–based Okazaki fragment maturation supports cell survival

Haitao Sun[1], Zhaoning Lu[1], Amanpreet Singh[1], Yajing Zhou[1], Eric Zheng[1,2], Mian Zhou[1], Jinhui Wang[3], Xiwei Wu[3], Zunsong Hu[4], Zhaohui Gu[4], Judith L. Campbell[5], Li Zheng[1]*, Binghui Shen[1]*

How cells with DNA replication defects acquire mutations that allow them to escape apoptosis under environmental stress is a long-standing question. Here, we report that an error-prone Okazaki fragment maturation (OFM) pathway is activated at restrictive temperatures in *rad27*Δ yeast cells. Restrictive temperature stress activated Dun1, facilitating transformation of unprocessed 5′ flaps into 3′ flaps, which were removed by 3′ nucleases, including DNA polymerase δ (Polδ). However, at certain regions, 3′ flaps formed secondary structures that facilitated 3′ end extension rather than degradation, producing alternative duplications with short spacer sequences, such as *pol3* internal tandem duplications. Consequently, little 5′ flap was formed, suppressing *rad27*Δ-induced lethality at restrictive temperatures. We define a stress-induced, error-prone OFM pathway that generates mutations that counteract replication defects and drive cellular evolution and survival.

Understanding the mutagenesis mechanisms that are active in cells under stress conditions is crucial for developing strategies to intervene in microbial pathogenesis, tumorigenesis, and drug resistance (*1*, *2*). Lagging-strand DNA synthesis is particularly vulnerable to stress and environmental factors. During replication, lagging-strand DNA is synthesized as discrete Okazaki fragments (*3*), which contain short primase- and DNA polymerase α (Polα)–synthesized RNA-DNA primers at their 5′ ends as well as the DNA fragment that is extended by DNA polymerase δ (Polδ) (*4*–*6*). During Okazaki fragment maturation (OFM), the RNA portion and any Polα-synthesized DNA with a high number of incorporation errors are displaced through Polδ-mediated DNA synthesis, which produces a 5′ RNA-DNA flap (*4*–*6*). The 5′ flap structure is removed by flap endonuclease 1 (FEN1) or through the sequential actions of DNA2 nuclease-helicase and FEN1 (*7*–*9*). FEN1 deficiency leads to the accumulation of unprocessed 5′ flap structures, which may prevent ligation of Okazaki fragments, leaving DNA nicks or gaps that lead to the collapse of replication forks and DNA double-strand breaks. In yeast, deletion of the FEN1 homolog *RAD27* (*rad27*Δ) results in slow growth at permissive growth temperatures (30°C) and death at restrictive growth temperatures (37°C) (*10*).

Nevertheless, we discovered that a small population of *rad27*Δ yeast cells, which we called revertants, could grow at a similar rate as wild-type (WT) cells at 37°C (Fig. 1A). To determine if the revertants acquired somatic mutation(s) that permitted growth and to identify any such mutation(s), we conducted whole-genome sequencing (WGS) of WT, parental *rad27*Δ, and a revertant strain of yeast cells. We identified 21 somatic DNA mutations specific to one revertant colony (table S1). A mutation in *POL3*, the Polδ catalytic subunit (*11*), was the only nonsynonymous mutation that had 100% allele frequency in the revertant. Subsequent polymerase chain reaction (PCR)–based DNA sequencing analysis of the *POL3* gene in independent *rad27*Δ revertant colonies (n = 31) revealed that each revertant colony harbored a *pol3* mutation (Fig. 1B). This suggested that these *pol3* mutations, which mapped onto Pol3 functional motifs (Fig. 1B and supplementary text S1) and possibly affected its biochemical activities, might provide a survival advantage for *rad27*Δ cells grown under restrictive temperature stress. Furthermore, knock in of the 458–477 internal tandem duplication (ITD) mutation, which occurred in 19 of the 31 independent colonies, or any of the four representative point mutations [Arg^{470}→Gly (R470G), Arg^{475}→Ile (R475I), Ala^{484}→Val (A484V), and Ser^{847}→Tyr (S847Y)] successfully reversed the restrictive temperature–induced lethality phenotype of *rad27*Δ cells (Fig. 1C and fig. S1). *rad27*Δ cells are sensitive to methyl methanesulfonate (MMS) (*10*). Although *rad27*Δ revertant cells and *rad27*Δ *pol3* ITD knock-in mutant cells were resistant to

[1]Department of Cancer Genetics and Epigenetics, Beckman Research Institute, City of Hope, 1500 East Duarte Road, Duarte, CA 91010, USA. [2]Department of Molecular, Cellular, and Developmental Biology, University of California at Santa Barbara, Santa Barbara, CA 93106, USA. [3]Department of Molecular and Cellular Biology, Beckman Research Institute, City of Hope, 1500 East Duarte Road, Duarte, CA 91010, USA. [4]Department of Computational and Quantitative Medicine, Beckman Research Institute, City of Hope, 1500 East Duarte Road, Duarte, CA 91010, USA. [5]Braun Laboratories, California Institute of Technology, Pasadena, CA 91125, USA.
***Corresponding author. Email: lzheng@coh.org (L.Z.); bshen@coh.org (B.S.)**

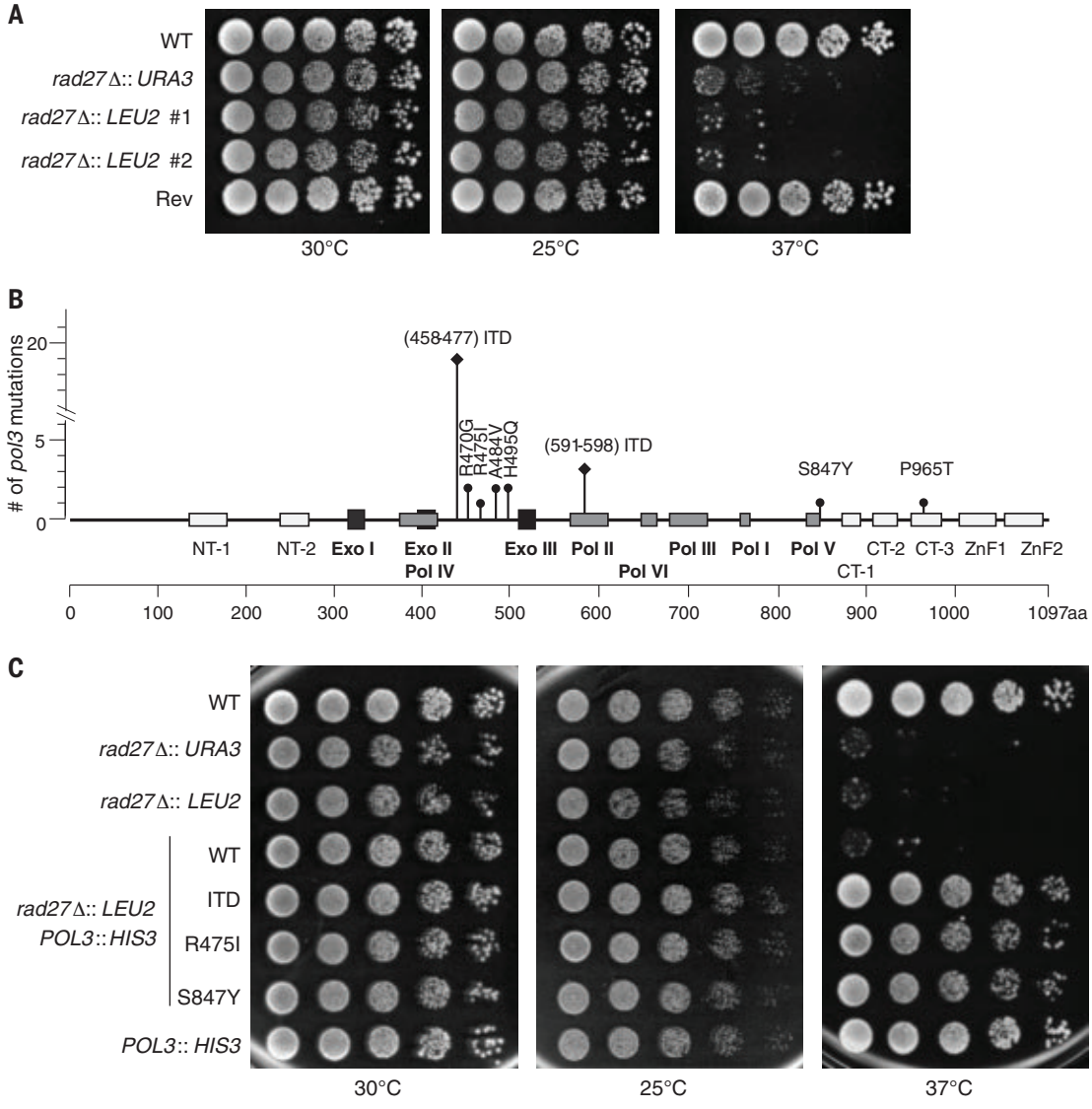


Fig. 1. *pol3* ITD and missense mutations drive resistance to *rad27*Δ-induced conditional lethality. (**A**) Spot assays of WT, *rad27*Δ, or *rad27*Δ revertant (Rev) yeast cells grown at 30°C (optimal temperature), 25°C (suboptimal temperature), or 37°C (restrictive temperature) for 48 hours. *rad27*Δ::*URA3* and *rad27*Δ::*LEU2* represent the *rad27*Δ allele with a linked *URA3* or *LEU2* selection marker gene, respectively. (**B**) *pol3* mutations detected in independent revertant strains (n = 31). Circles and diamonds represent base substitution and ITD mutations, respectively. The domain structures were defined as previously described (*23*). aa, amino acid; H495Q, His^{495}→Gln; P965T, Pro^{965}→Thr. (**C**) Spot assays of WT, *rad27*Δ, or *rad27*Δ yeast cells with indicated *pol3* knock-in mutations grown at 30°, 25°, or 37°C for 48 hours. *POL3*::*HIS3* represents the *POL3* (WT or mutant) alleles with a linked *HIS3* selection marker gene.

a low concentration (0.005%) of MMS, they were sensitive to higher concentrations (≥0.01%) of MMS (fig. S2). We observed that *pol3* ITD cells in a WT RAD27 background were also sensitive to high concentrations of MMS (fig. S2). This at least partially explains why the *pol3* ITD could not suppress MMS-induced lethality of *rad27*Δ cells at high MMS concentrations. In addition, *pol3* ITD did not rescue the synthetic lethality that occurred in the context of *rad27*Δ coupled with deficiency of the 5′ nucleases *EXO1* or *DNA2* nuclease-helicase (tables S2 and S3 and supplementary text S2).

Two types of duplications were present in the revertants: *pol3* 591–598 ITD (fig. S3, A and B), a previously reported classic duplication resulting from the realignment and ligation of unprocessed 5′ flaps (*12*), and *pol3* 458–477 ITD, which contained a 55–base pair (bp) duplication with a 5-bp spacer between the duplicated units (fig. S3, C and D). We named the duplication with an intervening spacer an "alternative duplication." Both *pol3* 591–598 ITD and *pol3* 458–477 ITD resembled ITDs detected in human cancer (*13*–*15*). To determine how the alternative duplication *pol3* 458–477 ITD was created, we conducted WGS of WT and *rad27*Δ cells grown at 37° or 30°C for 4 hours. The mutation frequency of WT cells was the same at both temperatures (Fig. 2A). By contrast, the mutation frequency of *rad27*Δ cells increased by twofold at restrictive temperature; in particular, the frequency of duplications and base substitutions was increased (Fig. 2A). In addition, duplication insertions in *rad27*Δ cells grown at 37°C were considerably longer than those in *rad27*Δ cells grown at 30°C (Fig. 2B). We found that *rad27*Δ cells grown at 37°C exhibited alternative duplications similar to the *pol3* 458–477 ITD. The alternative duplications were not detected in WT cells (30° or 37°C) or in *rad27*Δ cells grown at 30°C (Fig. 2C), suggesting that alternative duplications were induced by restrictive temperature stress.

We further noted that the sequences of these alternative duplications suggested the formation of three different types of hairpin structures (Fig. 2D; fig. S4, A to D; and supplementary text S3). This supported a model of sequential actions, including the conversion of a 5′ Okazaki fragment flap to a 3′ flap, annealing of the flap to a complementary sequence, extension of the 3′ flap, realignment, and ligation of the extended 3′ flap to produce an alternative duplication, like *pol3* 458–477 ITD. Consistent with this model, our WGS data indicated that 40% of the alternative duplications also carried base substitutions at the duplication unit. These substitutions most likely resulted from failure to remove Polα-generated errors on the 5′ flap. To determine if the restrictive temperature induced 3′ flap formation

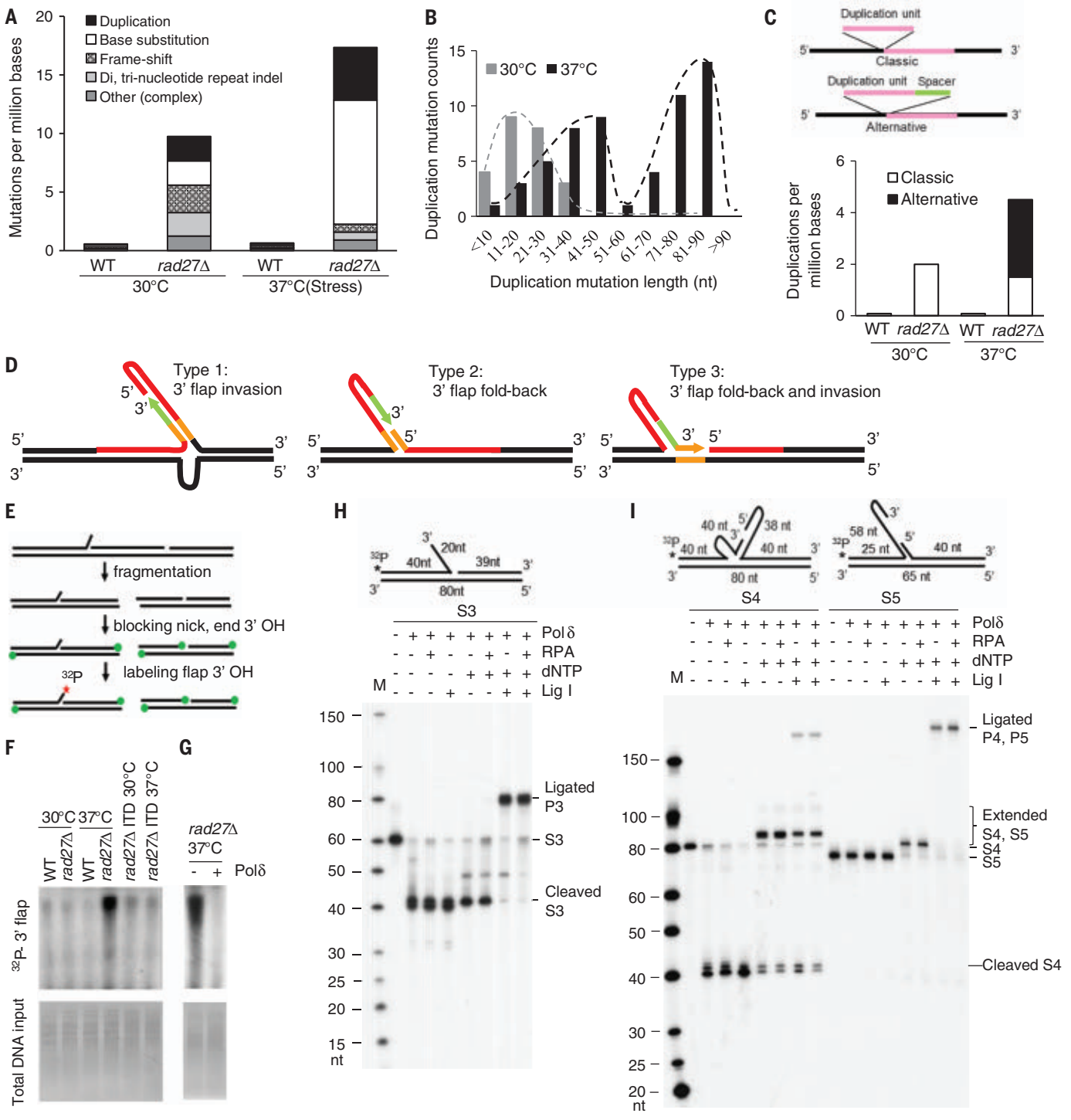


Fig. 2. Restrictive temperature stress induces 3′ flap–based OFM and results in alternative duplications. (**A**) Somatic mutation frequencies and types, as determined by WGS, in WT and *rad27*Δ cells grown at 30° or 37°C for 4 hours. (**B**) Lengths of inserted DNA sequences in duplications in *rad27*Δ cells grown at 30° or 37°C for 4 hours. The dashed lines indicate the trend lines of corresponding bar graphs. nt, nucleotide. (**C**) A diagram of classic and alternative duplications is shown at the top. Frequencies of classic and alternative duplications are shown at the bottom. (**D**) Predicted structures leading to three types of alternative duplications. Red and green lines represent the DNA sequences in red and green in fig. S4A; orange lines represent the orange-highlighted DNA sequences in fig. S4, A to D. (**E**) Schematic for specific labeling of 3′ flaps in genomic DNA. Green dots indicate dideoxyribonucleotide; the red star indicates ^{32}P-deoxyribonucleotide. (**F**) Levels of 3′ flaps in genomic DNA from WT, *rad27*Δ, or *rad27*Δ *pol3* ITD (*rad27*Δ-ITD) cells grown at 30° or 37°C for 4 hours. (**G**) Levels of 3′ flaps in genomic DNA from *rad27*Δ cells grown at 37°C with or without pretreatment with Polδ. (**H** and **I**) Reconstitution assays using ^{32}P-labeled 3′ flap substrate S3 (H) or ^{32}P-labeled secondary structure–forming 3′ flap substrate S4 or S5 (I). Substrate structures are shown above a representative image of 8% denaturing polyacrylamide gel electrophoresis (PAGE). DNA substrates (S3, S4, S5), cleavage products (cleaved S3, S4), unligated extended 3′ flap intermediates (extended S4, S5), and ligated extended products (ligated P3, P4, P5) are indicated. dNTP, deoxyribonucleotides; M, marker.

in *rad27*Δ cells, we developed an approach to specifically label the OH group on the 3′ flap on genomic DNA, in which 3′ OH at the nick or at the DNA end was preblocked with dideoxyribonucleotides (Fig. 2E). We detected a considerable number of 3′ flaps in *rad27*Δ cells grown at 37°C; by contrast, we detected few flaps in *rad27*Δ cells grown at 30°C, in WT cells grown at either temperature, or in *rad27*Δ cells carrying *pol3* 458–477 ITD grown at either temperature (Fig. 2F). Furthermore, preincubation of Polδ with genomic DNA from *rad27*Δ cells grown at 37°C could effectively remove the 3′ flaps (Fig. 2G), suggesting that Polδ may process 3′ flaps during OFM.

To define the proposed 3′ flap–based OFM mechanism, we reconstituted the sequential reactions of 3′ flap cleavage, DNA synthesis,

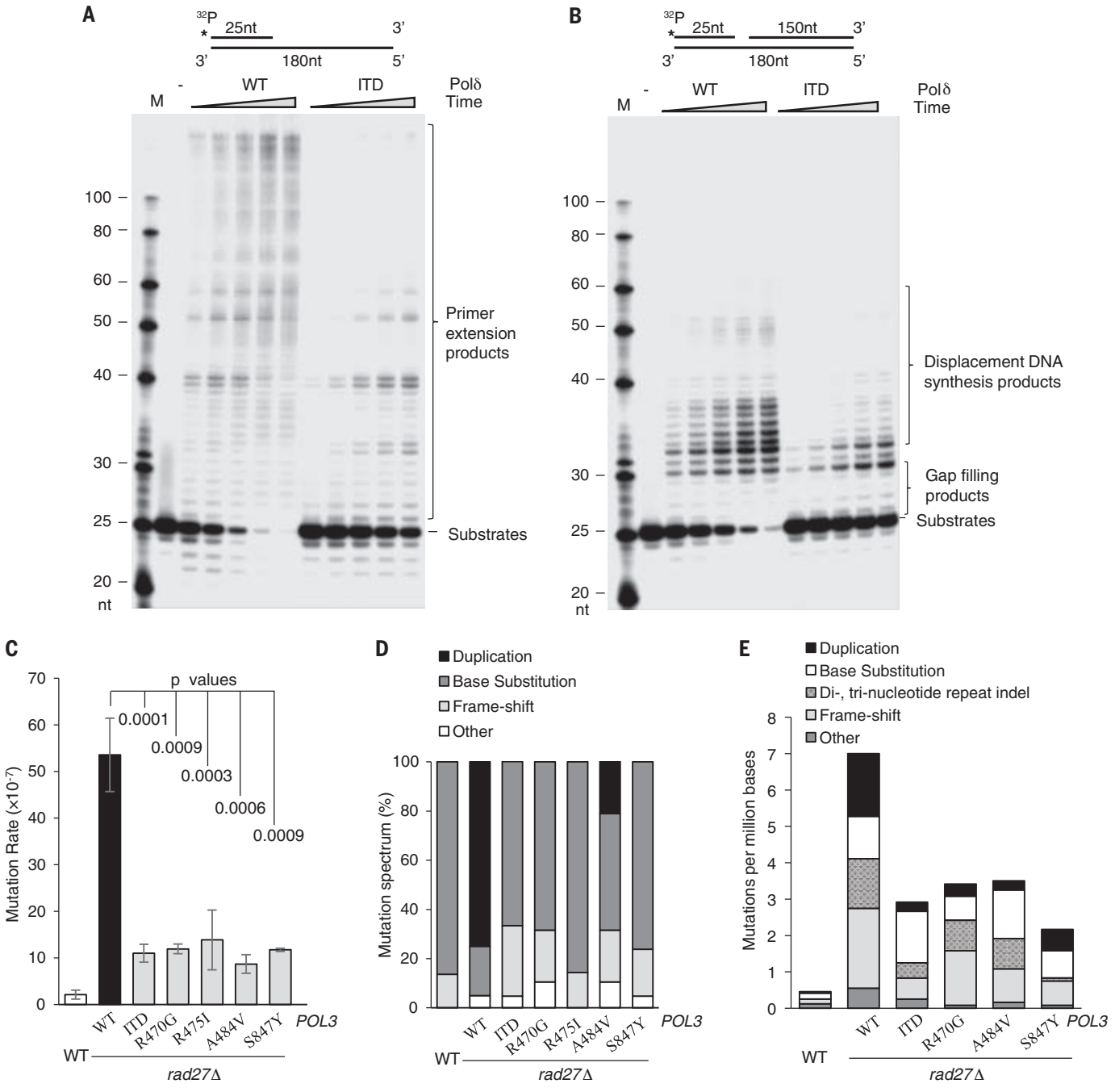


Fig. 3. Polδ-ITD suppresses 5′ flap formation. (**A** and **B**) In vitro assays of primer extension (A) and displacement DNA synthesis (B) by WT Polδ or Polδ-ITD. DNA substrates and primer extension products in (A) and DNA substrates, gap-filling products, or displacement DNA synthesis products in (B) are indicated. (**C**) Mean Can^r mutation rates of WT (n = 5), *rad27*Δ (n = 5), or *rad27*Δ yeast cells with knock-in of *pol3* 458–477 ITD (n = 3), *pol3* R470G (n = 2), *pol3* R475I (n = 3), *pol3* A484V (n = 2), and *pol3* S847Y (n = 2). Error bars indicate SD; p values were calculated using Student's t test. (**D**) Can^r mutation spectra of the indicated yeast strains. Values shown are percentages of the specific type of Can^r mutation in WT (n = 22), *rad27*Δ (n = 20), or *rad27*Δ with knock-in of *pol3* 458–477 ITD (n = 21), *pol3* R470G (n = 10), *pol3* R475I (n = 21), *pol3* A484V (n = 19), or *pol3* S847Y (n = 21). (**E**) Mutation frequencies and types present across the genome, as determined by WGS, in WT, *rad27*Δ, or *rad27*Δ cells with indicated *pol3* knock-in mutations grown at 30°C (n = 1).

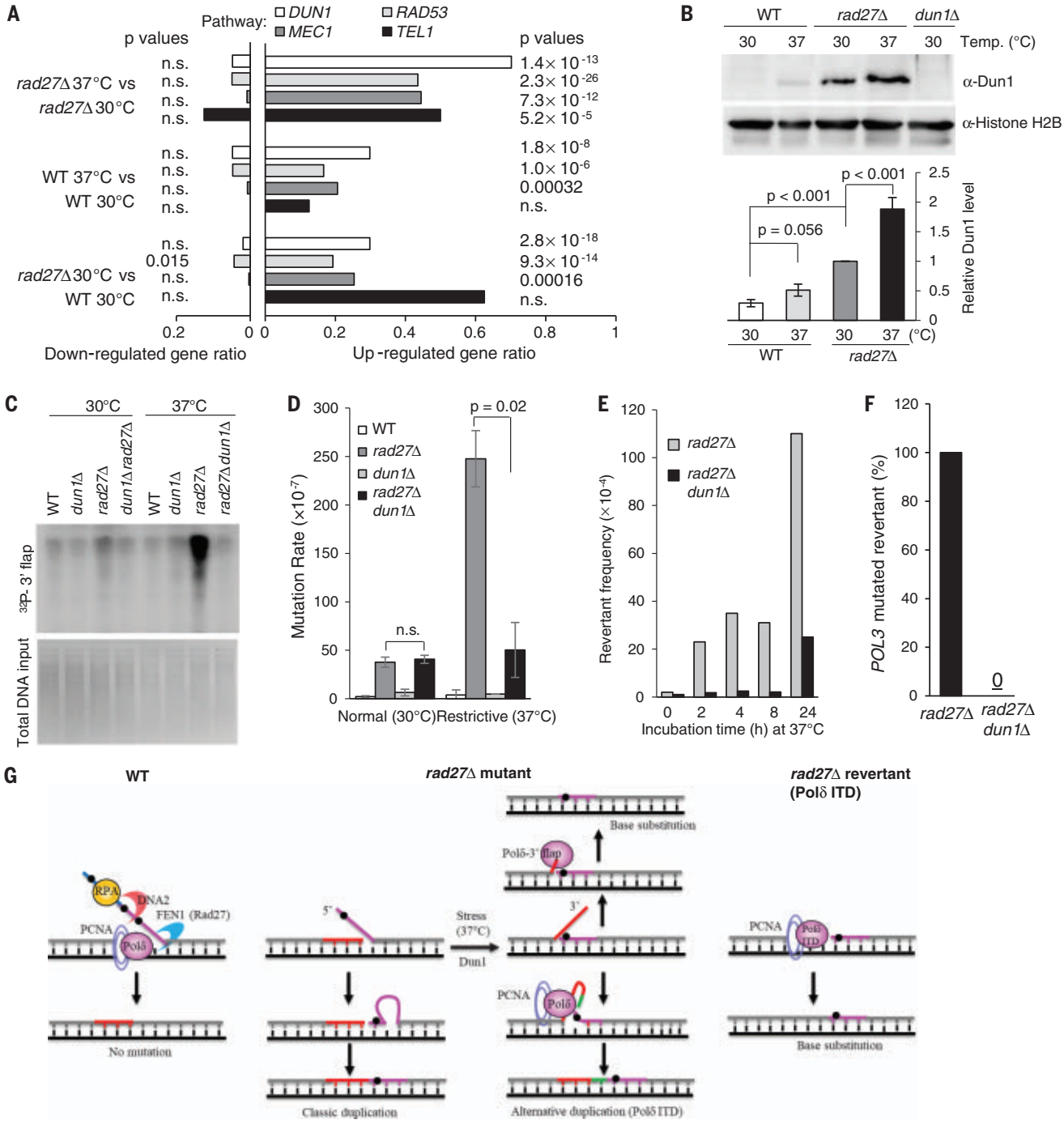


Fig. 4. Restrictive temperature stress activates signaling that facilitates error-prone OFM and generation of *rad27*Δ revertants. (**A**) Up-regulated or down-regulated gene ratios in Dun1-, Rad53-, Mec1-, or Tel1-controlled pathways in WT or *rad27*Δ yeast cells exposed to 30°C (4 hours) or 37°C (4 hours). *p* values were calculated using the hypergeometric test; n.s. indicates not significant. (**B**) A Western blot of chromatin-associated Dun1 protein in WT or *rad27*Δ cells exposed to 30°C (4 hours) or 37°C (4 hours) is shown at the top. Histone H2B was used as a loading control for the chromatin fraction in each sample. *dun1*Δ is a negative control. Quantification of chromatin-associated Dun1 relative to the loading control is shown at the bottom. The Dun1 level in *rad27*Δ cells grown at 30°C was arbitrarily set as 1, and the relative Dun1 levels in other samples were calculated by dividing their Dun1 levels by that in *rad27*Δ cells grown at 30°C. Error bars indicate SEM (*n* = 4 biological replicates). (**C**) Levels of 3′ flaps in genomic DNA from WT, *dun1*Δ, *rad27*Δ, or *rad27*Δ *dun1*Δ double-mutant cells grown at 30° or 37°C for 4 hours. (**D**) Mean Can^r mutation rates of WT, *rad27*Δ, *dun1*Δ, and *rad27*Δ *dun1*Δ cells exposed to 30°C (4 hours) or 37°C (4 hours). Error bars indicate SD (*n* = 3 independent assays). (**E**) Median revertant frequencies of *rad27*Δ or *rad27*Δ *dun1*Δ cells (*n* = 3 independent assays). (**F**) Percentage of *rad27*Δ or *rad27*Δ *dun1*Δ revertants that carry a *pol3* mutation (*n* = 19 for each strain). (**G**) Schematic illustrating error-free 5′ flap–mediated OFM in WT cells, error-prone 3′ flap–mediated OFM and the corresponding consequences in *rad27*Δ cells under restrictive temperature stress (37°C), and the impact of Polδ-ITD on OFM in the revertant. Blue and pink segments represent primase- and Polα-synthesized RNA primer and DNA. The red segment represents Polδ-synthesized DNA, which replaces Polα-synthesized DNA. The green segment represents Polδ-synthesized spacer DNA as part of alternative duplication. Black dots represent Polα incorporation errors.

and ligation of oligonucleotide-based DNA substrates (S) with a simple 3′ flap (S2 or S3; Fig. 2H and fig. S5B) or a secondary structure–forming 3′ flap (S4 or S5; Fig. 2I) for formation of type I or type II alternative duplications. In the presence of deoxyribonucleotide, Polδ could effectively cleave 3′ flap substrates S2 and S3 and stop at the junction of the 3′ flap and DNA duplex, generating ligatable DNA nicks for DNA ligase I (Lig I) (Fig. 2H, fig. S5, and supplementary text S4). However, deoxyribonucleotides inhibited the cleavage of hairpin-forming 3′ flaps and promoted extension of the annealed 3′ flap, producing ligated extended products (Fig. 2I and supplementary text S4); this process resembled the formation of alternative duplications. When extension of the annealed 3′ flap could not generate ligatable nicks, only unligatable extended products were produced (fig. S6, A to D), leading to the failure of 3′ flap–based OFM. The single-stranded DNA (ssDNA) binding protein RPA had little effect on Polδ-mediated 3′ flap cleavage or subsequent nick ligation (Fig. 2H), and it slightly enhanced formation of the ligated extended products (Fig. 2I).

Using reconstitution assays, we showed that the 3′ nuclease activities of Polδ and Lig I were sufficient to complete 3′ flap processing for OFM. Other nucleases in the nuclear extract (NE) might also be important in processing 3′ flaps, especially the hairpin-forming 3′ flap (fig. S7, A and B, and supplementary text S5). However, NE from *rad27*Δ cells, particularly those grown at 37°C, had reduced 3′ flap processing activity (fig. S7, A and B, and supplementary text S5). Because we observed no significant changes in the expression of major 3′ nucleases in yeast (fig. S8), we postulated that restrictive temperature stress could also induce molecular changes to inhibit 3′ flap processing, allowing 3′ flaps to invade into nearby homologous sequences, leading to alternative duplications.

We next determined how the *pol3* 458–477 ITD enabled *rad27*Δ cells to overcome lethal stress. Because the *pol3* 458–477 ITD did not change Polδ protein levels in *rad27*Δ cells (fig. S9), we tested if it affected biochemical properties of Polδ. We assayed the DNA polymerase and 3′ nuclease activities of a purified recombinant protein Polδ complex containing either a WT Pol3 subunit (WT Polδ) or a 458–477 ITD Pol3 subunit (hereafter called Polδ-ITD). Polδ-ITD could catalyze DNA synthesis but was less processive than WT Polδ during primer extension (Fig. 3A). Similarly, Polδ-ITD could effectively fill the gap, but it was less active than WT Polδ in displacing the downstream DNA oligonucleotide (Fig. 3B). In addition, Polδ-ITD had a weaker 3′ exonuclease activity on DNA duplexes compared with WT Polδ (fig. S10). However, Polδ-ITD had similar activity to WT Polδ in cleaving the 3′ flap and generating a ligatable nick (fig. S11). This activity likely allows cells that carry the *pol3* 458–477 ITD to have a similar capacity as WT cells for catalyzing 3′ flap processing for OFM. By contrast, a 3′ exonuclease-dead mutant, Polδ D520E (Asp^{520}→Glu), did not cleave the 3′ flap (fig. S11), which may explain why the Polδ D520E mutation is lethal at restrictive temperature and synthetically lethal with *rad27*Δ (*16*).

We further revealed that knock in of *pol3* mutations significantly reduced the mutation rate of *rad27*Δ cells, as measured by canavanine resistance (Can^r) (Fig. 3C), but did not affect the mutation rate of yeast cells with WT Rad27 (fig. S12). These *pol3* mutations nearly completely suppressed the occurrence of duplications (Fig. 3D). Consistent with the Can^r assay results, our WGS data confirmed that *pol3* mutations reduced the frequency of duplications and the overall mutation frequency (Fig. 3E). Duplication mutation rate correlates with the level of 5′ flap formation (*12*). Thus, our biochemical and genetic results demonstrate that *pol3* ITD and other point mutations can reverse the conditional lethality phenotype by limiting 5′ flap formation in *rad27*Δ cells.

To identify the signaling pathways that induced 3′ flap–mediated OFM and led to the generation of *pol3* ITD, we compared the transcriptomes of WT and *rad27*Δ cells grown at 37° or 30°C. We observed that genes regulated by the checkpoint kinases Mec1, Rad53, and Dun1 were significantly up-regulated in *rad27*Δ cells, especially those grown at 37°C (Fig. 4A); consistent with this, Western blot analysis confirmed that chromatin-associated Dun1 protein was increased in *rad27*Δ cells grown at 37°C (Fig. 4B). These results suggest activation of the Mec1-Rad53-Dun1 axis, the major signaling pathway that is activated to counteract genotoxic stress (*17*, *18*). We further showed that downstream targets of the up-regulated genes—including the stress response genes *HUG1*, *RNR2*, *RNR3*, and *RNR4* and the DNA repair gene *RAD51*—were synergistically induced by *rad27*Δ and restrictive temperature stress (fig. S13). *RAD51* is associated with inhibition of 3′ ssDNA degradation, which at least partially explains why degradation of 3′ flaps induced by NE from *rad27*Δ cells grown at 37°C was markedly less than degradation induced by WT NE (fig. S7, A and B).

To define the role of Dun1 signaling in stress-induced mutation and generation of revertants, we deleted the *DUN1* gene in WT and *rad27*Δ cells. We observed that knockout of *DUN1* (*dun1*Δ) in WT or *rad27*Δ cells had little effect on their survival (fig. S14), 3′ flap formation (Fig. 4C), or mutation rate at 30°C (Fig. 4D). However, *DUN1* deletion markedly reduced restrictive temperature stress–induced 3′ flap formation (Fig. 4C) and abolished restrictive temperature stress–induced mutations in *rad27*Δ cells (Fig. 4D). Consistent with this, *DUN1* deletion inhibited generation of *rad27*Δ revertants (Fig. 4E and supplementary text S6). Furthermore, all *rad27*Δ revertants in this experiment had *pol3* mutations, predominantly the *pol3* 458–477 ITD, but none of the *rad27*Δ *dun1*Δ revertants had *pol3* mutations (Fig. 4F). These findings suggest that Dun1 activation played an important role in the development of restrictive temperature stress–induced mutations that could reverse the lethal phenotype of *rad27*Δ cells. Consistent with this finding, blocking activation of Chk1, a Dun1 functional analog, significantly inhibited spontaneous lung cancer development in FEN1 mutant mice but not in WT mice (fig. S15 and supplementary text S7). An important function of Dun1 activation is to induce overexpression of *HUG1*, *RNR2*, *RNR3*, and *RNR4* for deoxyribonucleotide production. Increased deoxyribonucleotide concentrations changed the mode of action of Polδ and promoted the generation of ligated extended products in vitro (figs. S5, S16, and S17 and supplementary text S8). However, when we deleted *SML1*, the protein inhibitor of ribonucleotide reductase (*19*), to increase deoxyribonucleotide production, we did not observe increased mutation rates in *rad27*Δ cells (fig. S18), suggesting that an up-regulation of deoxyribonucleotide alone was not sufficient to promote alternative duplications.

To demonstrate the relevance of stress-induced 3′ flap–based OFM and alternative duplications in *rad27*Δ cells to human cancers, we used whole-exome sequencing (WES) to analyze alternative duplications in human tumors and mutant mice modeling human FEN1 mutations. Alternative duplications, similar to those in *rad27*Δ cells grown at restrictive temperature (i.e., 3′ flap OFM-related alternative duplications), were frequent in human B cell acute lymphoblastic leukemia (fig. S19, A to C, and supplementary text S9). In addition, the FEN1 A159V (Ala^{159}→Val) mutation, which occurs in human lung cancers (*20*), promoted 3′ flap OFM-related alternative duplications in mice (fig. S19D and supplementary text S9). Therefore, mutations in FEN1 or other OFM genes may lead to 3′ flap–based OFM and play crucial roles for cancer cell evolution, tumor growth, and resistance.

Our study defines error-prone processing of RNA-DNA primers during OFM (Fig. 4G). Induction of this mechanism generates alternative duplications and base substitutions. In WT cells, the displaced 5′ RNA-DNA flap is effectively cleaved either by Rad27 alone or by Dna2, which first cleaves the 5′ RNA-DNA flap in the middle, leaving a shorter 5′ DNA flap for Rad27 to subsequently cleave. When Rad27 is not available, other 5′ nucleases such as Dna2 alone or Exo1 are involved in inefficient 5′ flap processing (*21*, *22*). Resolution of 5′ flaps also

requires an alternative pathway that is mediated by the 3′ exonuclease activities of Polδ, which removes nucleotides from the 3′ end of an upstream Okazaki fragment, generating a gap for the unprocessed 5′ flap to reanneal for ligation (*16*, *23*). Restrictive temperature stress activates Dun1 signaling and stimulates de novo production of deoxyribonucleotides, which in turn inhibits the 3′ exonuclease activity, but not the flap nuclease activity of Polδ, and induces other DNA damage responses. These molecular changes push OFM toward transformation of an unprocessed 5′ flap into a 3′ flap, either through flap equilibration (*24*) or through the actions of helicases such as Sgs1 or Pif1, leading to a secondary structure that may result in alternative duplications, including Polδ-ITD, in revertant strains. In the revertants, Polδ mutations limit DNA displacement, thus suppressing 5′ flap formation or allowing more time for Dna2 or Exo1 to act on the 5′ flap and bypass the requirement for Rad27 (Fig. 4G).

REFERENCES AND NOTES

1. D. M. Fitzgerald, P. J. Hastings, S. M. Rosenberg, *Annu. Rev. Cancer Biol.* **1**, 119–140 (2017).
2. R. S. Galhardo, P. J. Hastings, S. M. Rosenberg, *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435 (2007).
3. T. Ogawa, T. Okazaki, *Annu. Rev. Biochem.* **49**, 421–457 (1980).
4. P. M. J. Burgers, T. A. Kunkel, *Annu. Rev. Biochem.* **86**, 417–438 (2017).
5. S. A. Nick McElhinny, D. A. Gordenin, C. M. Stith, P. M. Burgers, T. A. Kunkel, *Mol. Cell* **30**, 137–144 (2008).
6. S. Waga, B. Stillman, *Nature* **369**, 207–212 (1994).
7. M. R. Lieber, *BioEssays* **19**, 233–240 (1997).
8. S. H. Bae, K. H. Bae, J. A. Kim, Y. S. Seo, *Nature* **412**, 456–461 (2001).
9. Y. Liu, H. I. Kao, R. A. Bambara, *Annu. Rev. Biochem.* **73**, 589–615 (2004).
10. M. S. Reagan, C. Pittenger, W. Siede, E. C. Friedberg, *J. Bacteriol.* **177**, 364–371 (1995).
11. K. C. Sitney, M. E. Budd, J. L. Campbell, *Cell* **56**, 599–605 (1989).
12. D. X. Tishkoff, N. Filosi, G. M. Gaida, R. D. Kolodner, *Cell* **88**, 253–263 (1997).
13. A. Mariño-Enriquez *et al*., *Am. J. Surg. Pathol.* **42**, 335–341 (2018).
14. P. D. Kottaridis *et al*., *Blood* **98**, 1752–1759 (2001).
15. H. Kiyoi *et al*., *Leukemia* **12**, 1333–1337 (1998).
16. Y. H. Jin, R. Ayyagari, M. A. Resnick, D. A. Gordenin, P. M. Burgers, *J. Biol. Chem.* **278**, 1626–1633 (2003).
17. Z. Zhou, S. J. Elledge, *Cell* **75**, 1119–1127 (1993).
18. X. Zhao, R. Rothstein, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3746–3751 (2002).
19. A. Chabes, V. Domkin, L. Thelander, *J. Biol. Chem.* **274**, 36679–36683 (1999).
20. L. Zheng *et al*., *Nat. Med.* **13**, 812–819 (2007).
21. S. H. Bae, Y. S. Seo, *J. Biol. Chem.* **275**, 38022–38031 (2000).
22. P. T. Tran, N. Erdeniz, S. Dudley, R. M. Liskay, *DNA Repair* **1**, 895–912 (2002).
23. Y. H. Jin *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5122–5127 (2001).
24. Y. Liu, H. Zhang, J. Veeraraghavan, R. A. Bambara, C. H. Freudenreich, *Mol. Cell. Biol.* **24**, 4049–4064 (2004).

ACKNOWLEDGMENTS

We thank R. Kolodner for the yeast strains RDKY2672, RDKY2608, and RDKY2669; P. M. J. Burgers for the plasmids pBL335 (GST-Pol3), pBL338 (GAL1-Pol31), pBL340 (GAL10-Pol32), and pBL341 (Pol31/Pol32); L. Prakash and S. Prakash for the protease-deficient yeast strain YRP654 and the plasmids pBJ1445 (Flag-Pol3) and pBJ1524 (GST-Pol31/Pol32) to express the yeast recombinant Polδ complex (Pol3, Pol31, and Pol32); W.-D. Heyer for the anti-Dun1 antibody; and M. S. Wold for purified recombinant yeast replication protein A (RPA) complex. We thank H. Lou's laboratory members and H. Dai, D. Duenas, and M. E. Budd for technical assistance in mouse and yeast genetic experiments and stimulating discussions. We thank K. Walker and S. Wilkinson for critical reading and editing of the manuscript. **Funding:** This work was supported by NIH grants R50 CA211397 to L.Z. and R01 CA073764 and R01 CA085344 to B.S. Research reported in this publication includes work performed by City of Hope shared resources supported by the National Cancer Institute of the NIH under award number P30 CA033572. **Author contributions:** H.S., Z.L., A.S., Y.Z., and M.Z. conducted yeast genetic and biochemical experiments. E.Z., J.W., X.W., Z.H., and Z.G. conducted RNA sequencing (RNA-seq), WES, and WGS and performed data analysis. J.L.C. designed yeast genetic experiments and conducted data analysis. L.Z. conducted biochemical experiments, RNA-seq, and WGS data analysis; designed and coordinated most of the experiments; and wrote the first draft of the manuscript. B.S. supervised the entire project; designed and coordinated most of the experiments; and provided input into and finalized the manuscript. **Competing interests:** The authors declare no conflicts of interest in this study. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. Gene Expression Omnibus accession numbers for the mouse and yeast genomics datasets are GSE181154 and GSE178876, respectively.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj1013
Materials and Methods
Supplementary Text
Figs. S1 to S20
Tables S1 to S6
References (*25*–*51*)
MDAR Reproducibility Checklist

20 April 2021; accepted 14 October 2021
10.1126/science.abj1013

ORGANIC CHEMISTRY

A biomimetic S_H2 cross-coupling mechanism for quaternary sp^3-carbon formation

Wei Liu[1]†, Marissa N. Lavagnino[1]†, Colin A. Gould[1], Jesús Alcázar[2], David W. C. MacMillan[1]*

Bimolecular homolytic substitution (S_H2) is an open-shell mechanism that is implicated across a host of biochemical alkylation pathways. Surprisingly, however, this radical substitution manifold has not been generally deployed as a design element in synthetic C–C bond formation. We found that the S_H2 mechanism can be leveraged to enable a biomimetic sp^3-sp^3 cross-coupling platform that furnishes quaternary sp^3-carbon centers, a long-standing challenge in organic molecule construction. This heteroselective radical-radical coupling uses the capacity of iron porphyrin to readily distinguish between the S_H2 bond-forming roles of open-shell primary and tertiary carbons, combined with photocatalysis to generate both radical classes simultaneously from widely abundant functional groups. Mechanistic studies confirm the intermediacy of a primary alkyl–Fe(III) species prior to coupling and provide evidence for the S_H2 displacement pathway in the critical quaternary sp^3-carbon bond formation step.

Over the past five decades, transition metal–catalyzed cross-coupling has comprehensively transformed the landscape of molecule construction in the applied sciences, especially with respect to pharmaceuticals, agrochemicals, and functional materials (*1*, *2*). In particular, the combination of three mechanistic steps—oxidative addition, transmetalation, and reductive elimination—has served as a robust catalytic paradigm for C–C bond formation, enabling a highly modular yet general approach to fragment coupling (Fig. 1A). Although this paradigm has proven to be exceptionally successful for forging C(sp^2)–C(sp^2) bonds, it is important to recognize that each of these three elementary steps is less efficient when transition metals engage with secondary or tertiary alkyl fragments, limiting the development of a C(sp^3)–C(sp^3) cross-coupling platform of broad utility (*3*–*6*).

It is intriguing to consider that enzymatic formation of C(sp^3)–C(sp^3) bonds proceeds by fundamentally different open-shell pathways to achieve pivotal alkylation reactions (*7*, *8*). As one canonical example, methylcobalamin systems serve as nature's "free radical carrier" by stabilizing otherwise highly reactive methyl radicals (*9*, *10*). As such, in cobalamin-dependent radical *S*-adenosylmethionine (SAM) methyltransferases, transiently generated carbon-centered radicals can react with these alkyl-cobalt complexes via bimolecular homolytic substitution (S_H2) (Fig. 1B) (*11*). Although cobalamin provides critical stabilization to the reactive methyl radical, the methyl-cobalt bond remains notably weak (bond dissociation energy = ~37 kcal/mol), which underpins the kinetic preference for the S_H2 mechanism and heteroselective carbon-carbon bond formation (*12*). Elegant biosynthetic studies have shown that the rates of such enzymatic S_H2 reactions are extremely fast (~ 10^8 s^{-1}) and enable the formation of sterically congested

[1]Merck Center for Catalysis at Princeton University, Princeton, NJ 08544, USA. [2]Discovery Chemistry, Janssen Research and Development, Janssen-Cilag S.A., C/Jarama 75A, Toledo 45007, Spain.
***Corresponding author. Email: dmacmill@princeton.edu**
†These authors contributed equally to this work.

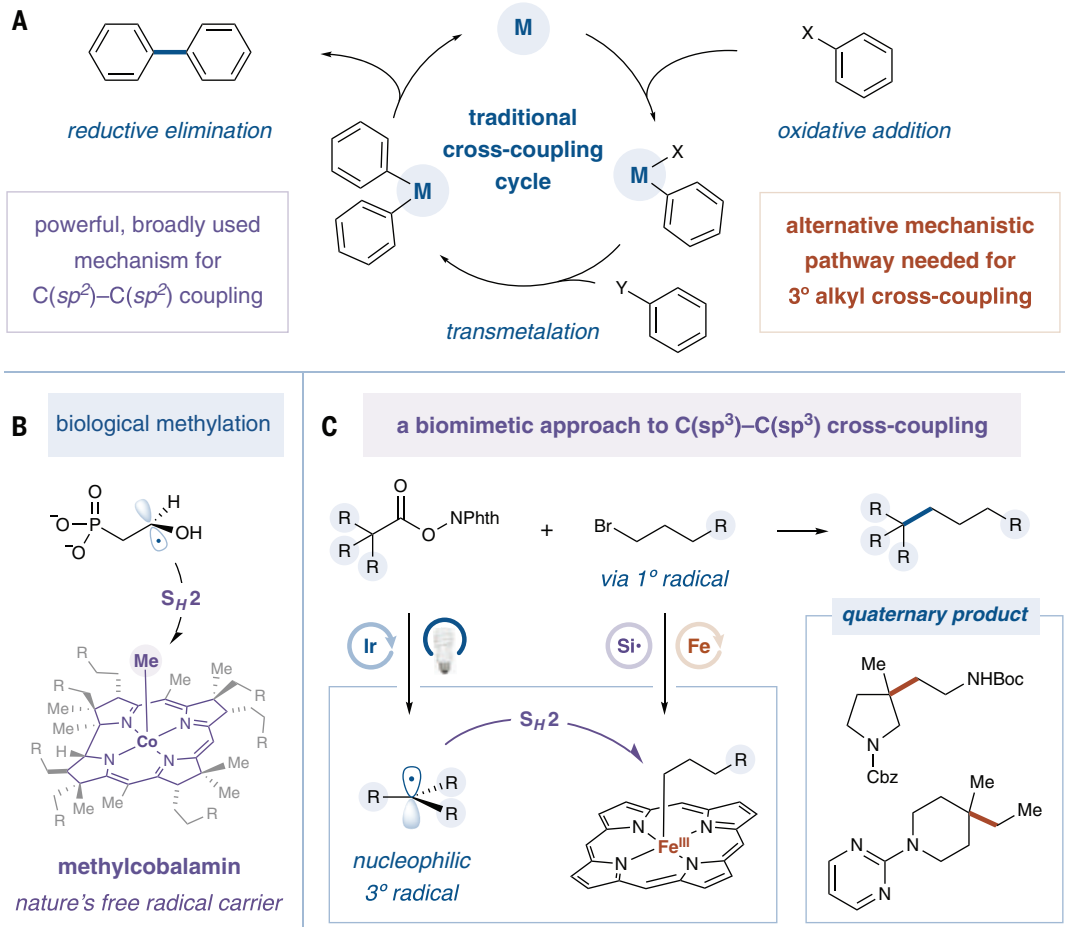


Fig. 1. Biomimetic C(sp^3)–C(sp^3) cross-coupling via dual iron/photoredox catalysis. (**A**) Representative catalytic cycle for transition metal–catalyzed cross-coupling of sp^2-hybridized fragments. (**B**) Reaction of carbon-centered radical with methylcobalamin via bimolecular homolytic substitution. (**C**) Design of a biomimetic approach to C–C bond formation via dual photoredox and iron catalysis. R, alkyl group; Me, methyl group; Phth, phthalimide; Boc, *tert*-butoxycarbonyl group; Cbz, benzyloxycarbonyl group.

quaternary C(sp^3) centers (*11*). However, despite broad biochemical relevance, S_H2-based cross-coupling paradigms remain effectively unknown within the laboratory setting, outside of stoichiometric organonickel methylation or intramolecular S_H cyclizations from seminal contributions of Sanford, Zhang, and others (*13*–*18*). Indeed, as Johnson stated in 1983 with respect to C–C bond formation, the S_H2 mechanism is "seldom postulated, rarely discussed, and frequently discarded as improbable" (*19*).

We recently asked whether a homolytic S_H2 pathway in combination with photoredox catalysis might be exploited to render an alternative catalysis paradigm for C(sp^3)–C(sp^3) bond formation (Fig. 1C). Previous bioinorganic studies have shown that both cobalt and iron porphyrins can serve as model systems of cobalamin, given that their respective alkyl-metal complexes possess weak metal-carbon bonds (*20*, *21*). These metalloporphyrins capture and release alkyl radicals reversibly, and the equilibrium is governed by the well-established bond dissociation energy of the metal-carbon bond (*22*). With this in mind, we recognized that such metalloporphyrin catalysts might effectively partition the roles of primary and tertiary radicals in a cross-coupling S_H2 reaction (Fig. 2A). More specifically, electron-rich tertiary radical ***3*** should be favored to induce S_H2 displacement of the primary alkyl fragment from 1° alkyl–Fe porphyrin ***5*** to generate heterocoupled C(sp^3)–C(sp^3) tertiary-primary linkages. However, the same 1° alkyl–Fe porphyrin ***5*** would be less susceptible to displacement by primary alkyl radical ***2*** given the reduced SOMO (singly occupied molecular orbital) nucleophilicity of primary radicals (*23*), a feature that should suppress the formation of 1°-1° homocoupled dimers. At the same time, 3° alkyl–metal porphyrin complex ***7*** is not formed in measurable equilibrium concentrations at room temperature (*24*), and its S_H2 displacement with other radicals (1°, 2°, or 3°) is kinetically slow as a result of induced nonbonding interactions [i.e., the pyramidalization of the 3° alkyl–Fe(III) intermediate] (*25*). As such, we postulated that the simultaneous generation of both primary and tertiary alkyl radicals in the presence of Fe porphyrin complexes should lead to heteroselective C(sp^3)–C(sp^3) bond formation in lieu of a statistical combination of open-shell processes.

Traditionally, alkyl-Fe or -Co systems are generated using Grignard reagents for alkyl transfer or via S_N2 pathways between low-valent metal porphyrins and alkyl halides, a viable yet relatively slow substitution step (k = ~10^2 s^{-1} for iron) (*26*). Furthermore, these alkyl-metal complexes are often heat- and oxygen-sensitive, restricting the options for open-shell alkyl nucleophile generation. As part of our design strategy, we recognized that photoredox catalysis should allow simultaneous generation of both primary and tertiary open-shell intermediates from widely abundant functional groups under mild conditions. For this first study, we selected a silyl radical–mediated halogen abstraction–radical capture (HARC) strategy (*27*–*29*) for the facile oxidative generation of alkyl radicals from primary alkyl bromides, after which access to electron-rich tertiary radicals from redox-active esters (readily derived from carboxylic acids) via reduction would ensure a net redox-neutral pathway (*30*).

It has long been recognized within medicinal chemistry that cyclic quaternary centers are conformationally restricted, a structural feature that is often linked to superior potency and metabolic stability in drug candidates (*31*, *32*). However, only a limited number of sp^3-sp^3 cross-coupling reports to date involve the formation of solely aliphatic quaternary

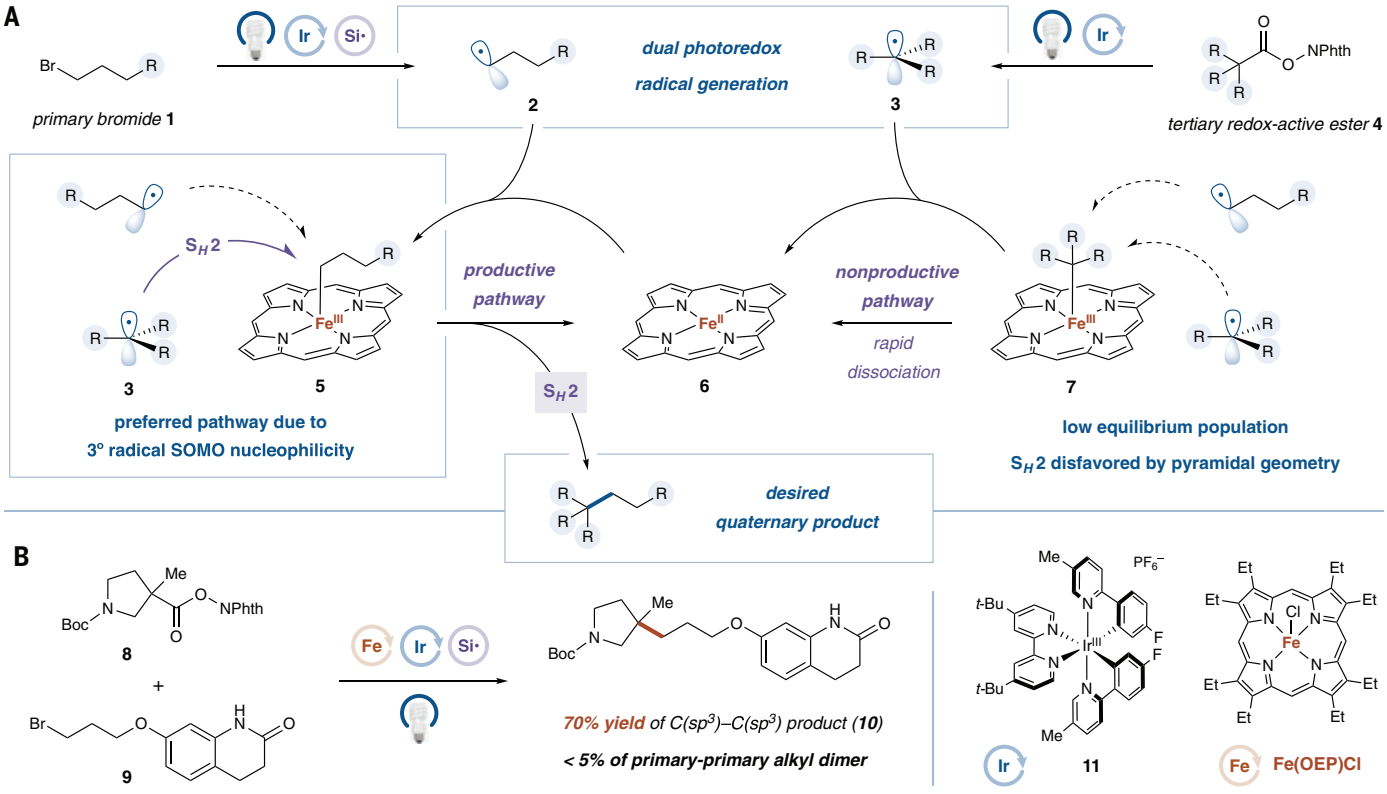


Fig. 2. Reaction design and development. (**A**) Proposed mechanism for the metallaphotoredox sp^3-sp^3 cross-coupling using iron porphyrin. (**B**) Representative reaction scheme. *t*-Bu, *tert*-butyl group; Et, ethyl group.

carbons, and these methods typically rely on highly reactive tertiary Grignard reagents or alkyl iodide electrophiles (*33*–*36*). We felt that the use of readily available redox-active esters and alkyl bromides as modular coupling fragments, in conjunction with the capacity of S_H2 for mechanistic partitioning, should lead to a generically useful $C(sp^3)$–$C(sp^3)$ cross-coupling method, thereby expanding the chemical space of sp^3-rich scaffolds that can be readily explored by medicinal chemists (*37*).

A description of our proposed mechanism for cross-coupling is outlined in Fig. 2A (see fig. S1 for a detailed proposal). Upon visible light excitation, the photocatalyst [Ir(FMeppy)$_2$(dtbbpy)][PF_6] [FMeppy = 2-(4-fluorophenyl)-5-(methyl)pyridine; dtbbpy = 4,4′-di-*tert*-butyl-2,2′-bipyridine] (**11**) would access a long-lived triplet excited-state species (lifetime τ = 1.1 μs) (*38*). This oxidizing Ir complex [half-wave reduction potential $E_{1/2}^{red}(*Ir^{III}/Ir^{II})$ = +0.77 V versus saturated calomel electrode (SCE) in CH_3CN] can undergo single electron transfer (SET) with the aminosilane reagent (peak potential of oxidation E_p^{ox} = +0.86 V versus SCE in *N*,*N*-dimethylacetamide and *tert*-amyl alcohol) to generate a reduced Ir(II) complex (*39*). The oxidized silane reagent would generate a reactive silyl radical, which readily abstracts a bromine atom from alkyl bromide **1** (*39*). The resulting primary alkyl radical **2** is expected to be captured by the Fe(II) porphyrin catalyst **6** at near diffusion-controlled rates to furnish 1° alkyl–Fe(III) intermediate **5** (*40*). Concurrently, the reduced Ir(II) complex [$E_{1/2}^{red}(Ir^{III}/Ir^{II})$ = −0.94 V versus SCE in CH_3CN] can reduce redox-active ester **4** via SET to furnish tertiary radical **3** upon extrusion of carbon dioxide and phthalimide (*30*). This matched combination of tertiary radical **3** with 1° alkyl–Fe(III) radicalphile **5** would lead to a successful S_H2 reaction, affording cross-coupled product and regenerating the Fe(II) catalyst.

With this mechanistic proposal in mind, we examined the cross-coupling between tertiary redox-active ester **8** and primary alkyl bromide **9**, both of which were selected on the basis of medicinal chemistry relevance (Fig. 2B). To our delight, we identified the commercial complex Fe(OEP)Cl [OEP = 2,3,7,8,12,13,17,18-octaethyl-21*H*,23*H*-porphine] as an effective S_H2 catalyst, in tandem with photocatalyst **11** and the aminosilane reagent $(TMS)_3SiNHAdm$ to deliver the quaternary carbon–bearing alkylation adduct **10** in 70% yield upon blue light irradiation. Control experiments revealed that all of the components used were necessary for optimal reaction performance; without Fe(OEP)Cl, only 13% yield of the desired product was observed, a result of free-radical background coupling (fig. S2).

Initial kinetic studies revealed that the reaction is zeroth-order in both of the fragment coupling substrates and first-order in photocatalyst and light intensity (see supplementary materials). However, an intriguing inverse order in the S_H2 catalyst Fe(OEP)Cl was observed. We subsequently determined that the iron porphyrin catalyst acts as an optical filter because of strong absorbance at 450 nm, thereby decreasing the photonic power available for the photoredox cycle in a reciprocal relationship to the concentration of the S_H2 catalyst. Indeed, with this information in hand, we recognized that similar levels of reaction efficiency should be achieved when the Fe(OEP)Cl loading is decreased (2 mol%) in proportion to light intensity, a hypothesis that was readily substantiated (fig. S4). The use of lower Fe porphyrin loadings allows for this coupling protocol to be scaled without loss in efficiency—a useful insight, especially when an apparatus with lower light intensity is used.

With optimal conditions in hand, we next examined the generality of our cross-coupling protocol with respect to the carboxylic acid component (Fig. 3). Bulky α-substitutions on pyrrolidine, such as isopropyl and benzyl groups, were well tolerated to furnish tertiary amine-bearing cross-coupled adducts in excellent yield (**15** and **16**, 71% and 80% yield),

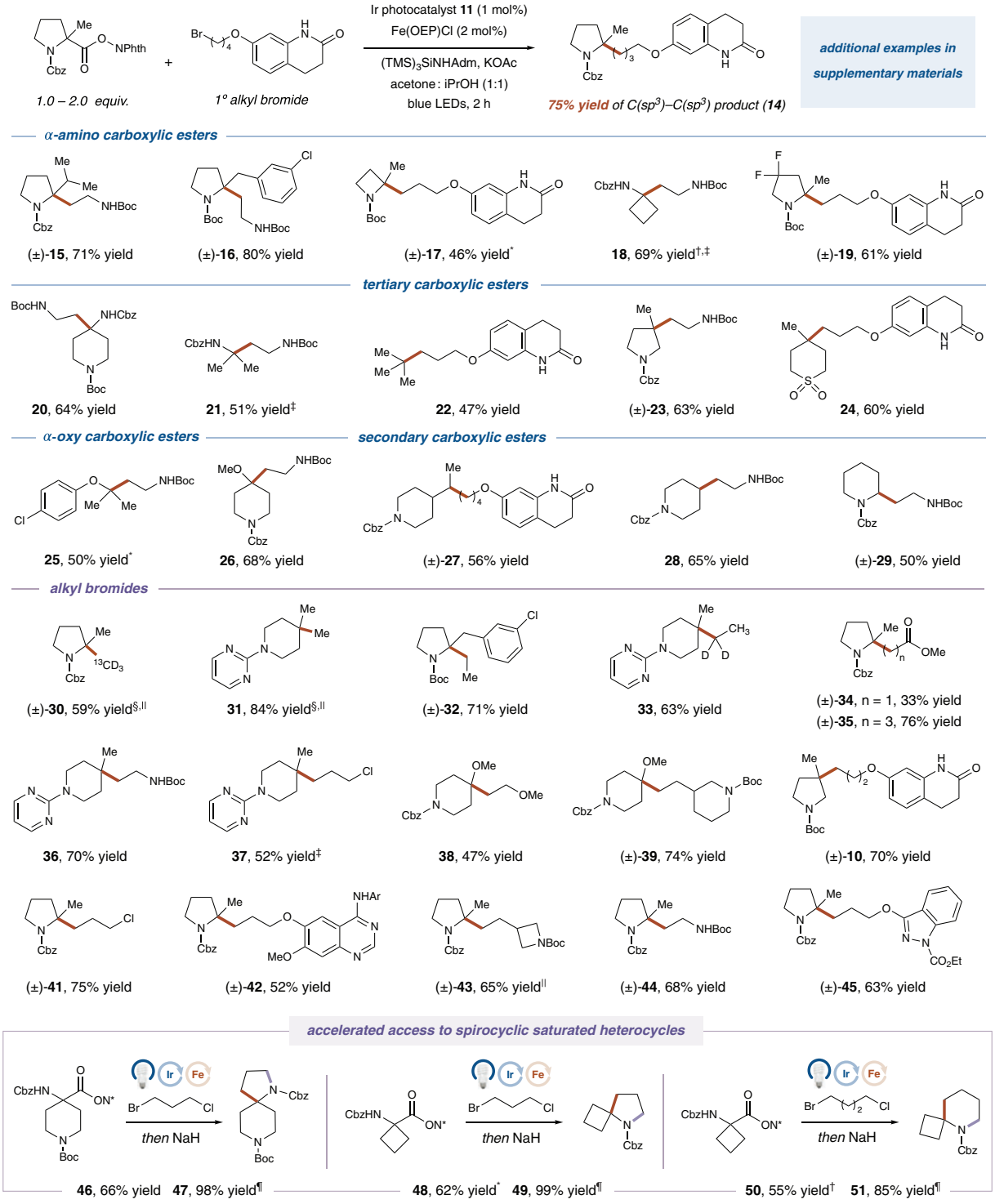


Fig. 3. Photoredox and iron-catalyzed C(sp^3)–C(sp^3) cross-coupling: Redox-active ester and alkyl bromide scope. All yields are isolated. See supplementary materials for detailed reaction conditions. TMS, trimethylsilyl group; Adm, 1-adamantyl; KOAc, potassium acetate; ^{i}PrOH, isopropanol; Ar, 3-chloro-4-fluorophenyl; N*, phthalimide. *With KOAc and $Zn(OAc)_2$ as bases. †With $Zn(OAc)_2$ as the base. ‡With $Ir(ppy)_2(dtbbpy)PF_6$ as the photocatalyst. §With 2 equivalents of methyl *p*-toluenesulfonate and tetrabutylammonium bromide. ||With $Ir[dF(CF_3)ppy]_2(dtbbpy)PF_6$ as the photocatalyst. ¶Sodium hydride, 60°C.

underscoring the capacity of the S_H2 mechanism to generally construct sterically hindered centers. Redox-active esters containing electron-deficient backbones (i.e., azetidine and difluoropyrrolidine) were found to be viable coupling partners and underwent alkylation in good yields (**17** to **19**, 46% to 69% yield). In addition, bulky α-functionalized exocyclic or acyclic amines could be accessed via cross-coupling in good efficiency (**20** and **21**, 64% and 51% yield). Furthermore, the use of tertiary redox-active esters enabled the formation of quaternary carbons, as represented in the acyclic *tert*-butyl moiety and a cyclic β-substituted pyrrolidine, as well as a

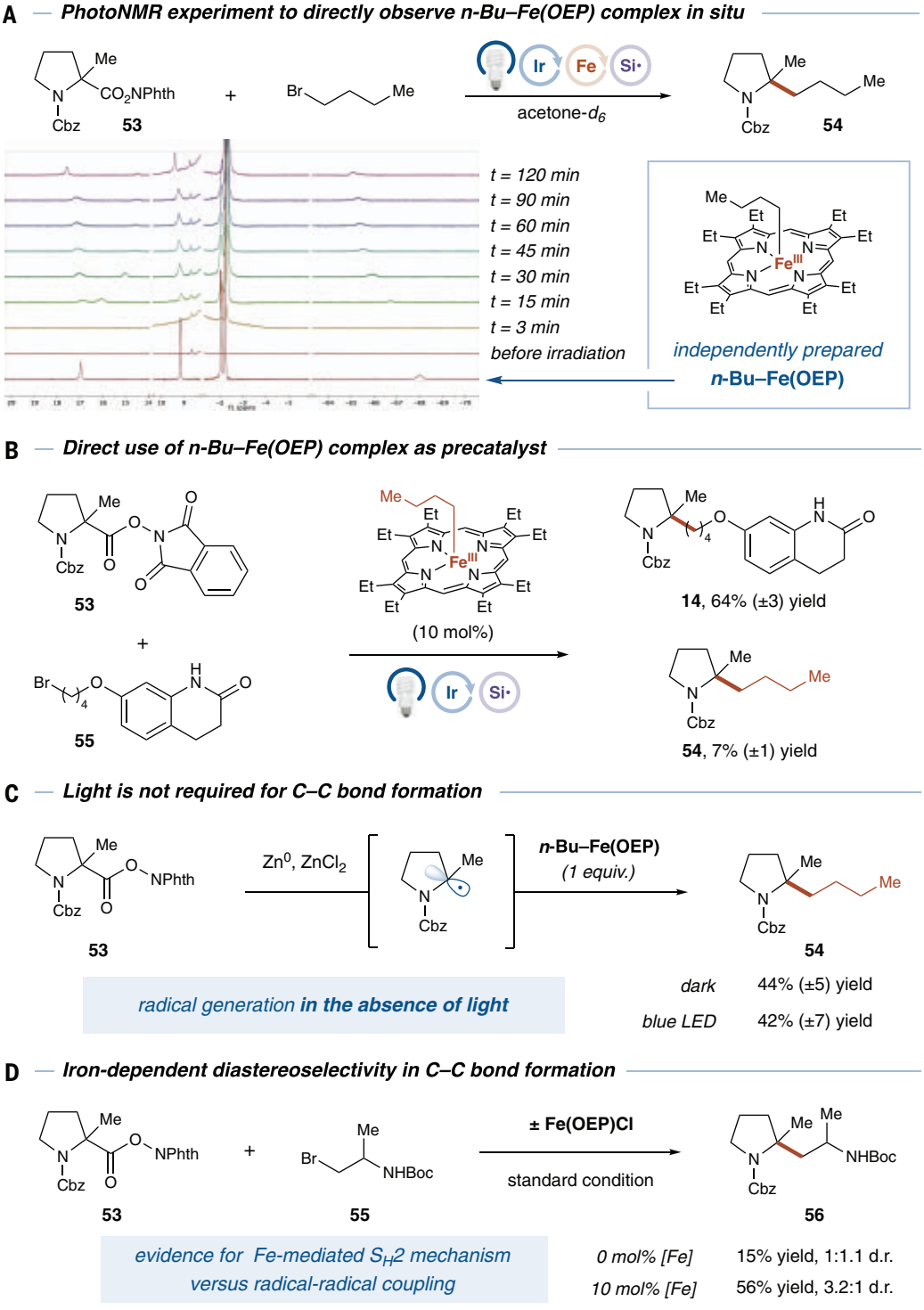


Fig. 4. Mechanistic studies for the proposed catalytic cycle and evidence of the intermediacy of alkyl-Fe(III) species. (**A**) Observation of key alkyl-Fe(III) species under reaction conditions. (**B**) Use of alkyl-Fe(III) complex as reaction precatalyst. (**C**) Formation of C–C bond under nonphotonic conditions. (**D**) Effect of iron porphyrin catalyst on diastereoselectivity of reaction. *n*-Bu, *n*-butyl group; $ZnCl_2$, zinc chloride.

medicinally important cyclic sulfone (**22** to **24**, 47% to 63% yield) (*41*). For α-oxy esters, radicals generated adjacent to both phenoxy and methoxy substituents could participate in cross-coupling with respectable efficiency, providing a new entry to the synthesis of hindered ethers (**25** and **26**, 50% and 68% yield). Finally, secondary redox-active esters could also be used in our metallaphotoredox protocol to couple with primary bromides in good yields (**27** to **29**, 50% to 65% yield). On the other hand, tertiary benzylic radical was found to be a challenging S_H2 reaction partner, presumably because of the diminished radical nucleophilicity (see fig. S12 for additional examples and limitations of substrate scope).

Next, we examined the scope of alkyl bromides using both α-heteroatom and tertiary redox-active esters as the representative coupling partners. Small alkyl fragments such as methyl

and ethyl could be introduced into sp^3 scaffolds in high efficiency (**30** to **33**, 59% to 84% yield). Methyl bromide was generated in situ via the combination of methyl tosylate and tetrabutylammonium bromide, whereas the use of methyl iodide led to diminished reactivity. No isomerization was observed in the alkylated product **33** when 1-bromoethane-1,1-d_2 was subjected to our reaction, demonstrating the orthogonality of the S_H2 mechanism to reductive elimination for C–C bond formation. Furthermore, 2-methylproline–derived redox-active esters could be alkylated with α- and γ-haloesters, providing straightforward access to homologated amino acids inaccessible via conjugate addition (**34** and **35**, 33% and 76% yield). A wide variety of functionalized alkyl bromides were successful coupling partners and furnished value-added products in good to high yields (**36** to **45**, 47% to 75% yield). The core heterocyclic fragments in aripiprazole (**10**), gefitinib (**42**), and benzydamine (**45**) were well tolerated in the cross-coupling, demonstrating the applicability of our method to medicinal chemistry campaigns. Finally, we used 1,3- and 1,4-bromo- and chloroalkyls as bifunctional linkers, which, after decarboxylative coupling, readily underwent intramolecular cyclization to directly construct medicinally relevant spirocyclic structures. This formal decarboxylative cycloaddition strategy was successfully applied to the synthesis of [5.6], [4.5], and [4.6] ring systems (**47**, **49**, and **51**), thereby providing a new and general approach to these synthetically challenging heterocycles from simple starting materials (*42*).

We performed detailed mechanistic experiments to support the proposed catalytic cycle and the intermediacy of 1° alkyl–Fe(III) species (Fig. 4 and fig. S9). Fluorescence quenching experiments confirmed the reductive quenching of the excited iridium photocatalyst **12** by aminosilane reagent **52** at a near diffusion-controlled rate ($k = 6.7 \times 10^8\ M^{-1}\ s^{-1}$), whereas the α-amino redox-active ester **53** was not found to be an effective quencher (fig. S9). To probe the intermediacy of the proposed alkyl–Fe(III) species, we used photo–nuclear magnetic resonance (NMR) techniques to monitor the cross-coupling between 1-bromobutane and α-amino redox-active ester **53** under our standard reaction conditions (Fig. 4A). By comparing these in situ spectra with an independently prepared *n*-Bu–Fe(OEP) complex (*43*), we directly observed the formation of the alkyl–iron porphyrin intermediate, and the concentration of this species was observed to slowly increase upon light exposure and to persist throughout the reaction. Furthermore, to demonstrate the catalytic relevance of the observed alkyl-Fe(III) species, we investigated the use of 10 mol% of the previously isolated *n*-Bu–Fe(OEP) adduct as both a catalytic intermediate and precatalyst in the cross-coupling of redox-active ester **53** and primary bromide **55** (Fig. 4B). Gratifyingly, the desired product **14** was observed in 64% yield, similar to the efficiency observed when Fe(OEP)Cl was used as the S_H2 precatalyst; notably, the *n*-butyl group was also incorporated into the alkylated product (**54**), providing direct evidence for the participation of the Fe(III)-alkyl species in the cross-coupling reaction.

Finally, given that the alkyl-Fe bonds of porphyrin complexes are known to homolyze under light irradiation to release alkyl radicals, it was unclear whether the C–C bond formation proceeds through free radical-radical coupling or the proposed S_H2 pathway. We sought to determine whether light is required in the C–C bond formation (Fig. 4C). When the independently generated *n*-Bu–Fe(OEP) complex was subjected to an α-amino radical arising from redox-active ester **53** under nonphotonic conditions (i.e., using zinc as the single-electron reductant) (*44*), the corresponding alkylation product was observed in good yield, indicating that the C–C bond formation is not dependent on photoexcitation of the S_H2 catalyst. Additionally, performing the same experiment under blue light irradiation led to the identical level of product formation—a result that aligns with the mechanistic interpretation that blue light is required for the photoredox cycle, yet is not involved in the C–C bond formation step. Furthermore, the iron porphyrin catalyst was able to achieve a degree of diastereocontrol during the cross-coupling of a β-chiral alkyl bromide **56** with redox-active ester **53** (Fig. 4D). Whereas free radical coupling without iron led to unselective diastereomer formation [diastereomeric ratio (d.r.), 1 to 1.1], the addition of iron porphyrin catalyst favored one major diastereomer in 3.2:1 diastereocontrol, providing further evidence for a concerted S_H2 mechanism and the participation of the iron-bound alkyl complex in the critical C–C bond-forming event.

REFERENCES AND NOTES

1. A. De Meijere, F. Diederich, *Metal-Catalyzed Cross-Coupling Reactions, Vol. 1* (Wiley-VCH, 2004).
2. M. R. Uehling, R. P. King, S. W. Krska, T. Cernak, S. L. Buchwald, *Science* **363**, 405–408 (2019).
3. J. Choi, G. C. Fu, *Science* **356**, eaaf7230 (2017).
4. S. Z. Tasker, E. A. Standley, T. F. Jamison, *Nature* **509**, 299–309 (2014).
5. T. Qin *et al*., *Science* **352**, 801–805 (2016).
6. C. P. Johnston, R. T. Smith, S. Allmendinger, D. W. C. MacMillan, *Nature* **536**, 322–325 (2016).
7. Q. Zhang, W. A. van der Donk, W. Liu, *Acc. Chem. Res.* **45**, 555–564 (2012).
8. M. R. Bauerle, E. L. Schwalm, S. J. Booker, *J. Biol. Chem.* **290**, 3995–4002 (2015).
9. J. Halpern, *Science* **227**, 869–875 (1985).
10. B. E. Daikh, R. G. Finke, *J. Am. Chem. Soc.* **114**, 2938–2943 (1992).
11. Y. Wang, T. P. Begley, *J. Am. Chem. Soc.* **142**, 9944–9954 (2020).
12. B. D. Martin, R. G. Finke, *J. Am. Chem. Soc.* **114**, 585–592 (1992).
13. J. R. Bour, D. M. Ferguson, E. J. McClain, J. W. Kampf, M. S. Sanford, *J. Am. Chem. Soc.* **141**, 8914–8920 (2019).
14. Y. Wang, X. Wen, X. Cui, X. P. Zhang, *J. Am. Chem. Soc.* **140**, 4792–4796 (2018).
15. X. Wang *et al*., *J. Am. Chem. Soc.* **143**, 11121–11129 (2021).
16. J. Xie *et al*., *J. Am. Chem. Soc.* **143**, 11670–11678 (2021).
17. M. Zhou, M. Lankelma, J. I. van der Vlugt, B. de Bruin, *Angew. Chem. Int. Ed.* **59**, 11073–11079 (2020).
18. J. C. Walton, *Acc. Chem. Res.* **31**, 99–107 (1988).
19. M. D. Johnson, *Acc. Chem. Res.* **16**, 343–349 (1983).
20. J. P. Collman, R. Boulatov, C. J. Sunderland, L. Fu, *Chem. Rev.* **104**, 561–588 (2004).
21. C. G. Riordan, J. Halpern, *Inorg. Chim. Acta* **243**, 19–24 (1996).
22. J. Halpern, *Polyhedron* **7**, 1483–1490 (1988).
23. F. Parsaee *et al*., *Nat. Rev. Chem.* **5**, 486–499 (2021).
24. A. L. Balch, R. L. Hart, L. Latos-Grazynski, T. G. Traylor, *J. Am. Chem. Soc.* **112**, 7382–7388 (1990).
25. D. Dodd, M. D. Johnson, B. L. Lockman, *J. Am. Chem. Soc.* **99**, 3664–3673 (1977).
26. D. Lexa, J. Mispelter, J. M. Savéant, *J. Am. Chem. Soc.* **103**, 6806–6812 (1981).
27. P. Zhang, C. C. Le, D. W. C. MacMillan, *J. Am. Chem. Soc.* **138**, 8084–8087 (2016).
28. C. Le, T. Q. Chen, T. Liang, P. Zhang, D. W. C. MacMillan, *Science* **360**, 1010–1014 (2018).
29. M. N. Lavagnino, T. Liang, D. W. C. MacMillan, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 21058–21064 (2020).
30. K. Okada, K. Okamoto, N. Morita, K. Okubo, M. Oda, *J. Am. Chem. Soc.* **113**, 9401–9402 (1991).
31. T. T. Talele, *J. Med. Chem.* **63**, 13291–13315 (2020).
32. F. Lovering, J. Bikker, C. Humblet, *J. Med. Chem.* **52**, 6752–6756 (2009).
33. T. Tsuji, H. Yorimitsu, K. Oshima, *Angew. Chem. Int. Ed.* **41**, 4137–4139 (2002).
34. C. T. Yang *et al*., *J. Am. Chem. Soc.* **134**, 11124–11127 (2012).
35. T. Iwasaki, H. Takagawa, S. P. Singh, H. Kuniyasu, N. Kambe, *J. Am. Chem. Soc.* **135**, 9604–9607 (2013).
36. S. A. Green, T. R. Huffman, R. O. McCourt, V. van der Puyl, R. A. Shenvi, *J. Am. Chem. Soc.* **141**, 7709–7714 (2019).
37. A. W. Dombrowski *et al*., *ACS Med. Chem. Lett.* **11**, 597–604 (2020).
38. M. S. Lowry *et al*., *Chem. Mater.* **17**, 5712–5719 (2005).
39. H. A. Sakai, W. Liu, C. C. Le, D. W. C. MacMillan, *J. Am. Chem. Soc.* **142**, 11691–11697 (2020).
40. D. Brault *et al*., *J. Am. Chem. Soc.* **102**, 1015–1020 (1980).
41. C. L. Hugelshofer *et al*., *Org. Lett.* **23**, 943–947 (2021).
42. E. M. Carreira, T. C. Fessard, *Chem. Rev.* **114**, 8257–8322 (2014).
43. P. Cocolios, G. Lagrange, R. Guilard, *J. Organomet. Chem.* **253**, 65–79 (1983).
44. J. Wang, B. P. Cary, P. D. Beyer, S. H. Gellman, D. J. Weix, *Angew. Chem. Int. Ed.* **58**, 12081–12085 (2019).

ACKNOWLEDGMENTS

Funding: Research reported in this publication was supported by National Institute of General Medical Sciences grant R35 GM134897-02, the Princeton Catalysis Initiative, and gifts from Janssen, Merck, Bristol Myers Squibb, GenMab, and Pfizer. C.A.G. thanks the Arnold and Mabel Beckman Foundation for a postdoctoral fellowship. **Author contributions:** W.L. and D.W.C.M. conceived of the work. All authors designed the experiments. W.L., M.N.L., and C.A.G. performed and analyzed the experiments. W.L., M.N.L., C.A.G., and D.W.C.M. prepared the manuscript. **Competing interests:** D.W.C.M. declares a financial interest with respect to the Integrated Photoreactor. **Data and materials availability:** Data are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4322
Materials and Methods
Figs. S1 to S12
NMR Spectra
References (*45*–*52*)

13 July 2021; accepted 14 October 2021
Published online 11 November 2021
10.1126/science.abl4322

NANOPHOTONICS

Continuous-wave frequency upconversion with a molecular optomechanical nanocavity

Wen Chen[1], Philippe Roelli[1]†, Huatian Hu[2], Sachin Verlekar[1], Sakthi Priya Amirtharaj[1], Angela I. Barreda[3], Tobias J. Kippenberg[1], Miroslavna Kovylina[4], Ewold Verhagen[5], Alejandro Martínez[4], Christophe Galland[1]*

Coherent upconversion of terahertz and mid-infrared signals into visible light opens new horizons for spectroscopy, imaging, and sensing but represents a challenge for conventional nonlinear optics. Here, we used a plasmonic nanocavity hosting a few hundred molecules to demonstrate optomechanical transduction of submicrowatt continuous-wave signals from the mid-infrared (32 terahertz) onto the visible domain at ambient conditions. The incoming field resonantly drives a collective molecular vibration, which imprints a coherent modulation on a visible pump laser and results in upconverted Raman sidebands with subnatural linewidth. Our dual-band nanocavity offers an estimated 13 orders of magnitude enhancement in upconversion efficiency per molecule. Our results demonstrate that molecular cavity optomechanics is a flexible paradigm for frequency conversion leveraging tailorable molecular and plasmonic properties.

Control and analysis of electromagnetic signals spanning the full spectrum from radio waves to x-rays governs technological progress in areas ranging from information processing, telecommunication networks, material characterization, spectroscopy, and imaging to remote sensing. The mid- and far-infrared (IR) frequency range, from a few to 100 THz, finds applications such as in homeland security, molecular analysis of gases, chemicals and biological tissues (*1*), thermal imaging and nondestructive material inspection (*2*), and astronomical surveys (*3*). However, IR detection technologies (*4*) do not rival with visible and near-infrared (VIS/NIR) detectors in terms of sensitivity, cost-effectiveness, and integration, motivating new approaches to perform IR spectroscopy with VIS/NIR detectors. Such methods include nonlinear interferometers (*5*, *6*) and coherent frequency upconversion (*7*, *8*), which is compatible with quantum technologies (*9*, *10*). Coherent upconversion of IR signals can be accomplished with bulk nonlinear optics using three-wave mixing processes, but delicate phase matching and propagation in centimeter-long crystals are needed to reach high efficiencies (*11*, *12*). Three-wave mixing may also occur at nanoscale interfaces and is used to probe the properties and dynamics of molecular layers with ultrafast nonlinear spectroscopy (*13*, *14*); however, such techniques require substantial peak powers only accessible with femto- or picosecond pulses.

Optomechanical cavities have recently emerged as promising candidates to realize quantum coherent frequency conversion (*9*, *10*, *15*). In a possible implementation, the signal of interest

[1]Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. [2]Hubei Key Laboratory of Optical Information and Pattern Recognition, Wuhan Institute of Technology, Wuhan 430205, China. [3]Institute of Applied Physics, Abbe Center of Photonics, Friedrich Schiller University Jena, 07745 Jena, Germany. [4]Nanophotonics Technology Center, Universitat Politècnica de València, 46022 Valencia, Spain. [5]Center for Nanophotonics, AMOLF, 1098 XG Amsterdam, Netherlands.
*Corresponding author. Email: chris.galland@epfl.ch
†Present address: Nano-optics Group, CIC nanoGUNE BRTA, E-20018 Donostia, San Sebastián, Spain.

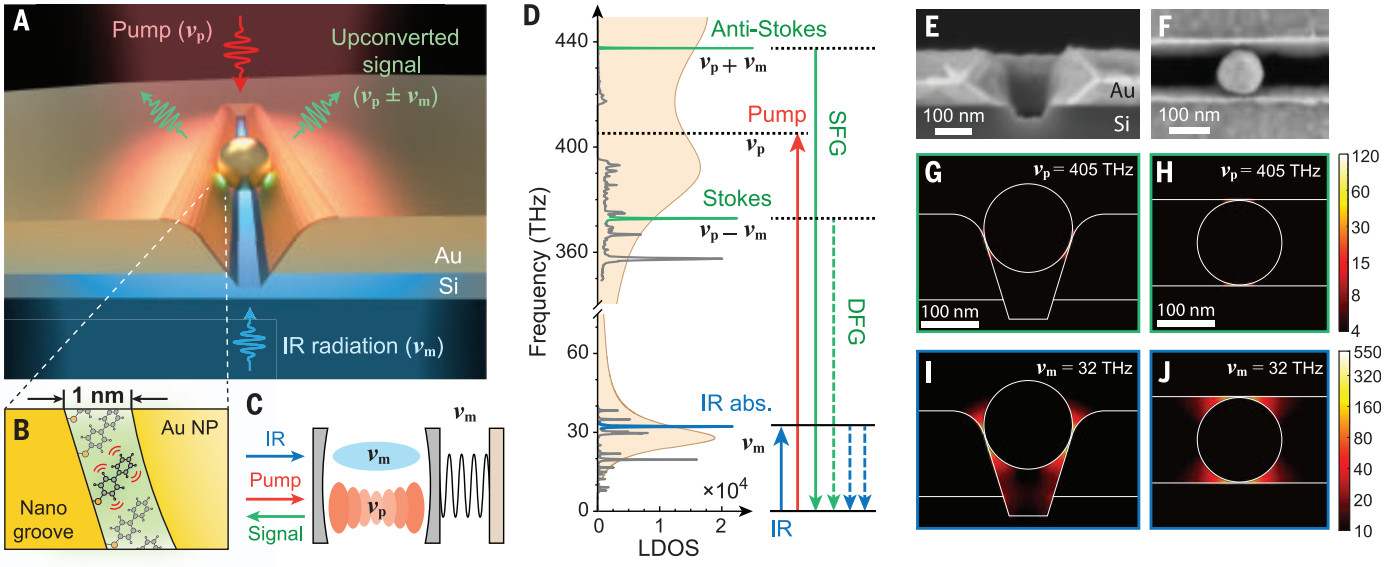


Fig. 1. Molecular optomechanical upconversion concept. (**A**) Illustration of a nanoparticle-in-groove cavity confining IR (frequency ν_m) and VIS (frequency ν_p) fields into a ~1-nm-thick BPhT molecular layer (**B**) Au NP, gold nanoparticle (~150 nm diameter). (**C**) The molecular vibration resonantly couples to the IR field and parametrically couples (through the Raman polarizability) to the VIS field, realizing an optomechanical cavity. (**D**) Vibrational levels and transitions involved in upconversion. LDOS, computed radiative local density of state inside the nanocavity. The measured Raman scattering and simulated molecular IR absorption (compare supplementary materials, section 1.5) are overlaid. SFG/DFG, sum frequency generation/difference frequency generation. (**E** and **F**) Cross-sectional scanning electron microscope (SEM) image of a fabricated nanogroove (E) and top view of a nanoparticle-in-groove (F). (**G** to **J**) Simulated electromagnetic field enhancement factors for incident plane waves at ν_p = 405 THz [740 nm; (G) and (H)] and ν_m = 32 THz [9.3 μm; (I) and (J)], both polarized orthogonally to the groove.

resonantly drives a mechanical oscillator, itself parametrically coupled to a laser-driven optical cavity, which results in modulation sidebands at the sum and difference frequencies (called anti-Stokes and Stokes sidebands, respectively). This approach offers a number of advantages, such as the resonant enhancement of nonlinear response at the mechanical frequency and the parametric enhancement of conversion efficiency with intracavity pump power. It is highly versatile and has been demonstrated with mechanical resonance frequencies ranging from kilohertz (*16*) to gigahertz (*17*–*21*). In a different approach, modulations on terahertz waves have been read out optically through a megahertz-frequency mechanical resonator (*22*). Molecular oscillators constitute a new frontier in cavity optomechanics (*23*, *24*) because they enable multi-terahertz resonant frequencies and room temperature quantum coherent operation (*25*). Moreover, they can be coupled to plasmonic nanocavities with deep-subwavelength mode volumes, thereby enabling optomechanical coupling rates in excess of 1 THz (*26*). Although plasmonic gap modes have been demonstrated to substantially enhance other nonlinear effects (*27*–*29*), frequency conversion devices based on molecular cavity optomechanics have yet to be demonstrated.

Here, we experimentally demonstrate the upconversion of continuous wave IR signal at ~32 THz (9.3 μm wavelength) into the visible domain using a subwavelength molecular optomechanical cavity at ambient conditions. Our upconversion scheme operates with microwatt-level continuous wave signal and pump beams and allows high-resolution spectroscopy of the IR signal because of the coherent nature of optomechanical transduction. This regime of operation is achieved by coupling a molecular monolayer to a doubly resonant plasmonic gap nanocavity, which supports deep-subwavelength mode volumes and simulated field enhancement factors in excess of 500 and 100 at IR and VIS frequencies, respectively, from which an overall enhancement of upconversion efficiency per molecule by >13 orders of magnitude compared with free space is predicted and experimentally validated.

A molecular optomechanical platform for upconversion was proposed in (*23*) and its theoretical performance was analyzed in (*8*), showing the feasibility of single-photon sensitivity at frequencies down to a few terahertz at ambient conditions. The mechanical resonator consists of a collective molecular vibration, which is parametrically coupled to the nanocavity through its Raman polarizability (*23*). This approach allows reaching mechanical frequencies in the 1 to 100 THz range. Our experiment is conceptually presented in Fig. 1, A to D. A dual-band "nanoparticle-in-groove" plasmonic nanocavity (Fig. 1, E and F) is realized by placing a single Au nanoparticle (150 nm nominal diameter) inside a nanogroove etched in a gold film and covered by a monolayer of biphenyl-4-thiol (BPhT) acting as spacer and molecular oscillator (Fig. 1, B and C). Being non-centrosymmetric, BPhT supports vibrational modes that are both IR and Raman active (fig. S13 and supplementary materials, section 1.5), and we used such a mode at 32.4 THz. Our structure is engineered to support colocalized plasmonic resonances at IR and VIS frequencies, which can be excited under normal incidence illumination (fig. S12 and supplementary materials, section 1.4) and correspond to near fields confined in the nanometer-wide gaps formed by the molecular layer (Fig. 1, G to J). The IR resonance frequency is governed by the length of the nanogroove (*30*) (fig. S2), which was chosen as 2 μm

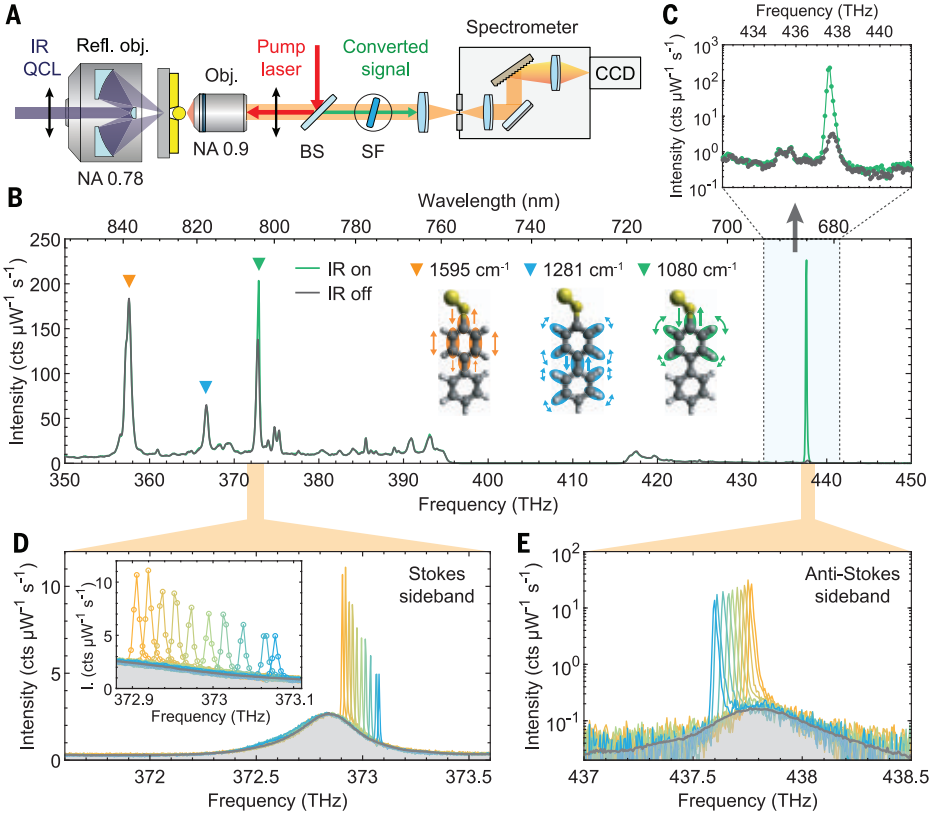


Fig. 2. Molecular optomechanical transduction from 32 THz to the visible domain. (**A**) Schematic of the setup. A reflective objective with numerical aperture (NA) 0.78 focuses the IR beam from a quantum cascade laser (QCL) through the Si substrate while a refractive objective (NA 0.9) focuses the visible pump beam and collects the Raman signal, which is directed to a spectrometer after blocking the pump light with spectral filters (SF). The polarization of pump and IR radiation is perpendicular to the nanogroove (black arrows). BS, beam splitter. (**B**) Low-resolution broadband Raman spectra from a single nanocavity under 10 μW pump power (740 nm wavelength) without (black line) and with (green line) incoming IR radiation (530 μW) near resonant with the BPhT vibration at 32.4 THz (1080 cm^{-1}). The inset shows the vibration patterns of some Raman-active modes labeled with orange, blue, and green triangles. Acquisition time was 10 s. (**C**) Zoom on anti-Stokes sideband plotted on a logarithmic vertical scale. (**D** and **E**) High-resolution (0.23 cm^{-1}, 7 GHz) Stokes (D) and anti-Stokes (E, logarithmic vertical scale) spectra measured when tuning the IR frequency (range limited by the QCL), all normalized to 175 μW of incoming IR power (pump: 10 μW). The gray lines and areas are the spectra without IR radiation and their Lorentzian fits. The inset of (D) shows an enlarged view of the resolution-limited upconverted peak (compare figs. S7 and S8). Acquisition times for the upconverted signals, spontaneous Stokes spectra, and anti-Stokes spectra were 10, 300, and 600 s, respectively.

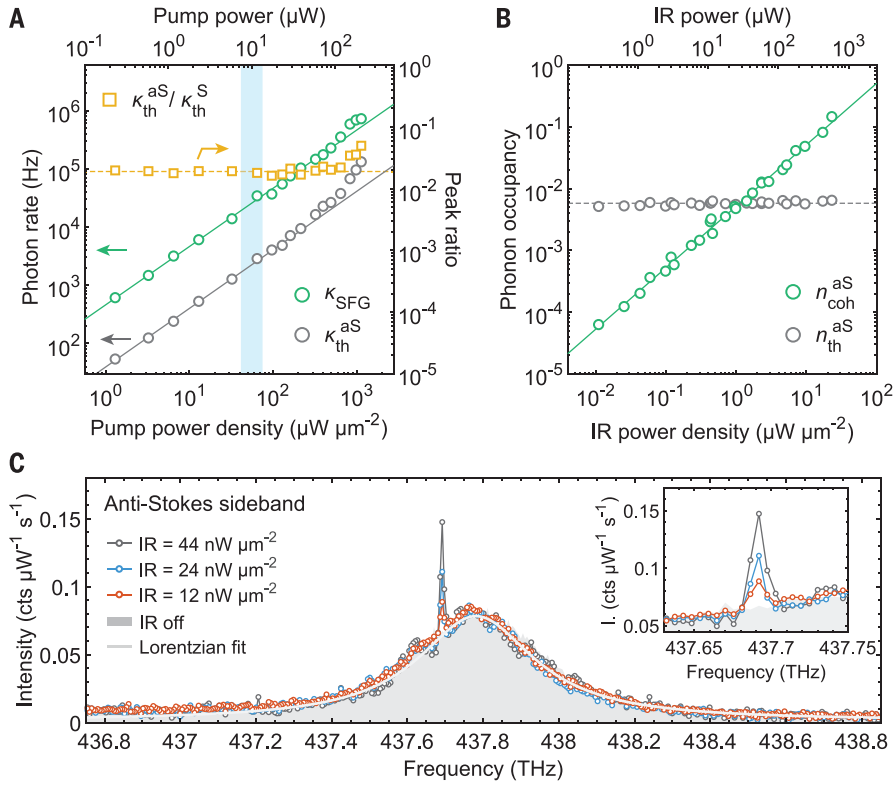


Fig. 3. Dependence of thermal and upconverted signals on IR and VIS powers. (**A**) Gray and green circles represent the rates of thermal (κ_{th}^{aS}) and upconverted (κ_{SFG}; for 600 μW IR power) anti-Stokes photons, respectively, as a function of VIS pump power (tuned at 740 nm). The ratio of thermal anti-Stokes to Stokes signals (without IR input) is plotted as orange squares. Solid lines are linear functions; the dashed line is a constant. The blue shaded area denotes the power range used in other upconversion measurements. (**B**) IR-driven anti-Stokes coherent occupancy n_{coh}^{aS} (green circles) extracted from the sharp peak area [compare with (C)], together with the thermal occupancy inferred from the area of the broad anti-Stokes emission (gray circles) as a function of incident IR power (VIS pump power 10 μW, tuned at 740 nm). The green line is a linear fit; the dashed line is a constant. To convert peak area to phonon occupancy, we assumed that the molecules were at 25°C without IR input. (**C**) Detection of submicrowatt IR signals using high-resolution spectroscopy on the anti-Stokes sideband. Acquisition time was 600 s.

to match the vibration frequency of BPhT at $\nu_m \simeq 32.4$ THz.

When a pump laser tuned at 405 THz (740 nm) is focused on the sample (Fig. 2A), the parametric interaction with the molecular vibrations generates Raman sidebands at lower (Stokes) and higher (anti-Stokes) frequencies (Fig. 2B). The giant enhancement factor described above makes it possible to detect the Raman signal from few hundred molecules, as estimated from the VIS mode area (fig. S12) and molecular layer density (see the supplementary materials, section 1.6). Without IR beam incident on the device (black solid line in Fig. 2B), the Stokes signal is dominated by spontaneous emission of phonon-photon pairs, whereas the anti-Stokes signal originates from the upconversion of thermal vibrations (*31*). At room temperature (T = 25°C), the thermal occupancy of the vibrational mode at ν_m = 32 THz is

$$n_{th} = \left(\exp\left(\frac{h\nu_m}{k_B T}\right) - 1\right)^{-1} \simeq 5.8 \times 10^{-3}$$

where h and k_B are Planck's and Boltzman's constants, respectively.

When the IR beam from a quantum cascade laser was focused through the Si substrate onto the back side of the device (Fig. 2A), we observed two new peaks at the corresponding Raman shift on the Stokes and anti-Stokes sidebands (green line in Fig. 2, B and C), the linewidth of which was much narrower than the natural linewidth of the spontaneous Raman scattering peaks (Fig. 2, D and E). Altogether, these observations are compatible with a coherent upconversion process as pictured in Fig. 1D. As we tuned the frequency of the incoming field, the upconverted signal shifted accordingly, and its measured linewidth was found to be limited by that of our spectrometer, with a value of 7 GHz or 0.23 cm^{-1} (compare fig. S8), which is well below that of single-molecule Raman linewidths typically observed at room temperature (see the supplementary materials, section 2.2). The relative conversion efficiency versus detuning is plotted in fig. S7 and confirms that upconversion is assisted by the vibrational mode.

We interpret these results as being the manifestation of optomechanical transduction, in which a collective molecular vibrational mode is resonantly driven by the nanocavity-enhanced incoming IR field. Because of the ~50 times larger mode volume in the IR versus VIS domain (compare Fig. 1, G to J, and fig. S12), many of the molecules covering the groove may be vibrationally excited yet remain silent in Raman scattering. In the following, we consider only the subset of molecules that are in the region where IR and VIS near fields strongly overlap, which contains a few hundred molecules at most (see calculation in the supplementary materials, section 1.6). We can describe their collective vibrational state by a displaced thermal state with a coherent amplitude α and a corresponding coherent occupancy $n_{coh} = |\alpha|^2$. This coherent, IR-driven oscillation is mapped onto the Raman sidebands of the pump laser, where the IR signal can be analyzed and detected using Si-based detectors. Without IR drive, the Stokes and anti-Stokes photon rates are proportional to $1 + n_{th}$ and n_{th}, respectively; with IR drive, these rates become proportional to $1 + n_{th} + n_{coh}^{S}$ and $n_{th}^{aS} + n_{coh}$, respectively. We added the superscripts "S" and "aS" to highlight that the observed value of n_{coh} depends on the overlap between Raman scattered and IR near fields. Knowing n_{th}, we can extract $n_{coh}^{S,aS}$ from the measured Raman count rate with and without IR drive or, equivalently, from the area subtended by the sharp upconverted peak versus that of the broad spontaneous or thermal emission. Moreover, under the approximation $n_{th} \ll 1$, the Stokes sideband offers a self-calibrated measurement of n_{coh}^{S}. For most nanocavities, we found $0.1 \leq n_{coh}^{S} \leq 0.5$ for 500 to 600 μW IR power (compare table S2), corresponding to 20 to 24 μW/μm^2 IR power density at the spot's center (fig. S17). This figure is compatible with an enhancement of IR absorption cross section per molecule by more than five orders of magnitude, as predicted from field enhancement simulations and detailed in the supplementary materials, section 1.6.

When varying the power of the pump laser at 740 nm, we observed a linear dependence

of the upconverted Raman signal from 0.1 to 100 μW (green circles in Fig. 3A), which is consistent with the expected parametric upconversion process. In this power range, the spontaneous Stokes and thermal anti-Stokes signals also grow linearly with pump power, and their ratio remains constant (gray circles and yellow squares in Fig. 3A; see also fig. S10). An optomechanical description of plasmon-enhanced Raman scattering (*8*, *23*, *32*) predicts three main regimes for the pump power dependence of the spontaneous Stokes and anti-Stokes signals: (i) at low power, a linear regime dominated by thermal noise; then (ii) a quadratic increase of anti-Stokes intensity as the vibrational population increases linearly with laser power due to quantum back action, yet remains small against unity (also called vibrational pumping) (*24*, *26*, *33*); and finally (iii) a phonon-stimulated regime is expected, dominated by dynamical back-action amplification of the vibration (*23*), where both Stokes and anti-Stokes powers diverge and the harmonic potential approximation breaks down. The data in Fig. 3A are compatible with regime (i) except >100 μW, where signatures of regime (ii) may be inferred; however, the behavior is not always reversible at such powers, suggesting that permanent changes in the nanocavity affected the observations. Figure 3B shows that the upconverted anti-Stokes signal scales linearly with incoming mid-IR power, as expected for a resonant drive well below saturation. We also plotted with gray circles the area of the thermal incoherent anti-Stokes peak, which shows a slight increase with IR power. As detailed in fig. S16, we deduced that the effective temperature of the molecules did not rise by >10°C for the highest IR powers.

Finally, we quantified the external IR to visible conversion efficiency of our device by spectrally filtering the anti-Stokes sideband and sending it to a single-photon-counting module with independently calibrated detection efficiency (results are summarized in table S2). We inferred that the upconverted anti-Stokes photon rate collected by the objective reaches up to 200 kHz per nanocavity for 600 μW incident IR power (corresponding to $n_{IR} \simeq 2.8 \times 10^{16}$ photons/s) at 10 μW pump power. Direct comparison of these incoming and upconverted fluxes yields a conversion efficiency from incoming IR photon to outgoing visible photon collected by our objective on the order of 10^{-12} (fig. S5), or 10^{-7}/W of pump power. Only 6% of the power emitted in the near field was collected by our objective, as simulated in Fig. S12E, so that the internal efficiency is at least 15 times larger. Despite the low efficiency, the coherent nature of the process allowed us to reliably detect incoming IR powers densities down to tens of nanowatts per square micrometer, as shown in Fig. 3C, a figure that would further improve with the increasing resolution of the spectrometer.

We present an optomechanical nanocavity leveraging molecular vibrations that are both IR and Raman active for coherent frequency conversion between the mid-infrared and visible domains. Subwavelength device dimensions elude the need for phase matching and permit the extension of spectral coverage by the mere choice of molecule (*34*) (figs. S14 and S15) and adjustment of plasmonic resonance (fig. S2). We identified several parameters that can be improved to increase conversion efficiency, notably a better overlap between IR and VIS near fields, a larger number of nanocavity-coupled molecules, and a higher resilience to Raman pump power. Operating our device in the vibrational strong coupling regime may enable efficient bidirectional IR ↔ VIS operation (see the supplementary materials, section 2.1). Finally, our concept is compatible with photonic integrated circuits and with the realization of chip-scale pixel arrays used as IR spectrometers and hyperspectral imagers.

Note added in proof: Related experiments are independently reported (*35*).

REFERENCES AND NOTES

1. S. De Bruyne, M. M. Speeckaert, J. R. Delanghe, *Crit. Rev. Clin. Lab. Sci.* **55**, 1–20 (2018).
2. F. Ciampa, P. Mahmoodi, F. Pinto, M. Meo, *Sensors* **18**, 609 (2018).
3. T. L. Roellig *et al*., *J. Astron. Telesc. Instrum. Syst.* **6**, 041503 (2020).
4. A. Rogalski, *Infrared and Terahertz Detectors* (CRC Press, ed. 3, 2019).
5. D. A. Kalashnikov, A. V. Paterova, S. P. Kulik, L. A. Krivitsky, *Nat. Photonics* **10**, 98–101 (2016).
6. C. Lindner *et al*., *Opt. Express* **29**, 4035–4047 (2021).
7. A. Barh, P. J. Rodrigo, L. Meng, C. Pedersen, P. Tidemand-Lichtenberg, *Adv. Opt. Photonics* **11**, 952–1019 (2019).
8. P. Roelli, D. Martin-Cano, T. J. Kippenberg, C. Galland, *Phys. Rev. X* **10**, 031057 (2020).
9. N. Lauk *et al*., *Quantum Sci. Technol.* **5**, 020501 (2020).
10. N. J. Lambert, A. Rueda, F. Sedlmeir, H. G. L. Schwefel, *Advanced Quantum Technologies* **3**, 1900077 (2020).
11. G. Temporão *et al*., *Opt. Lett.* **31**, 1094–1096 (2006).
12. Y.-P. Tseng, C. Pedersen, P. Tidemand-Lichtenberg, *Opt. Mater. Express* **8**, 1313–1321 (2018).
13. S. Roke, G. Gonella, *Annu. Rev. Phys. Chem.* **63**, 353–378 (2012).
14. C. S. Tian, Y. R. Shen, *Surf. Sci. Rep.* **69**, 105–131 (2014).
15. Y. Chu, S. Gröblacher, *Appl. Phys. Lett.* **117**, 150503 (2020).
16. T. Bagci *et al*., *Nature* **507**, 81–85 (2014).
17. R. W. Andrews *et al*., *Nat. Phys.* **10**, 321–326 (2014).
18. A. Vainsencher, K. J. Satzinger, G. A. Peairs, A. N. Cleland, *Appl. Phys. Lett.* **109**, 033107 (2016).
19. K. C. Balram, M. I. Davanço, J. D. Song, K. Srinivasan, *Nat. Photonics* **10**, 346–352 (2016).
20. M. Forsch *et al*., *Nat. Phys.* **16**, 69–74 (2020).
21. M. Mirhosseini, A. Sipahigil, M. Kalaee, O. Painter, *Nature* **588**, 599–603 (2020).
22. C. Belacel *et al*., *Nat. Commun.* **8**, 1578 (2017).
23. P. Roelli, C. Galland, N. Piro, T. J. Kippenberg, *Nat. Nanotechnol.* **11**, 164–169 (2016).
24. M. K. Schmidt, R. Esteban, F. Benz, J. J. Baumberg, J. Aizpurua, *Faraday Discuss.* **205**, 31–65 (2017).
25. S. Tarrago Velez, V. Sudhir, N. Sangouard, C. Galland, *Sci. Adv.* **6**, eabb0260 (2020).
26. F. Benz *et al*., *Science* **354**, 726–729 (2016).
27. M. P. Nielsen, X. Shi, P. Dichtl, S. A. Maier, R. F. Oulton, *Science* **358**, 1179–1181 (2017).
28. S. Ummethala *et al*., *Nat. Photonics* **13**, 519–524 (2019).
29. Y. Salamin *et al*., *Nat. Commun.* **10**, 5550 (2019).
30. C. Huck *et al*., *ACS Photonics* **2**, 1489–1497 (2015).
31. S. T. Velez *et al*., *Phys. Rev. X* **9**, 041007 (2019).
32. M. K. Schmidt, R. Esteban, A. González-Tudela, G. Giedke, J. Aizpurua, *ACS Nano* **10**, 6291–6298 (2016).
33. K. Kneipp *et al*., *Phys. Rev. Lett.* **76**, 2444–2447 (1996).
34. Z. Koczor-Benda, P. Roelli, C. Galland, E. Rosta, "Molecular vibration explorer" (Materials Cloud, 2021); https://molecular-vibration-explorer.matcloud.xyz/voila/render/index.ipynb.
35. A. Xomalis *et al*., *Science* **374**, 1268–1271 (2021).
36. Data for: W. Chen *et al*., Continuous-wave frequency upconversion with a molecular optomechanical nanocavity, Zenodo (2021); https://doi.org/10.5281/zenodo.5558767.

ACKNOWLEDGMENTS

C.G. is indebted to V. Sudhir for valuable comments and fruitful discussions about the results. W.C. and C.G. thank H. Altug for providing access to an FTIR spectrometer and acknowledge support from the IPHYS mechanical workshop, characterization platform, and EPFL CMi cleanroom. **Funding:** This work received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement nos. 829067 (FET Open THOR), 820196 (ERC CoG QTONE), and 732894 (HOT). C.G. acknowledges support from the Swiss National Science Foundation (project nos. 170684 and 198898). This work is part of the research program of the Netherlands Organisation for Scientific Research (NWO). A.I.B. acknowledges financial support by the Alexander von Humboldt Foundation. **Author contributions:** W.C. designed and fabricated the devices, performed the experiments, analyzed the data, and created the main figures. P.R. performed calculations of molecular parameters, assisted in the early stages of the experiments, and assisted in data analysis. H.H. performed the electromagnetic simulations of nanocavities and contributed to nanocavity design. S.V. assisted in recording and analyzing the photon-counting data. S.P.A. contributed to setting up the experimental apparatus. A.I.B., M.K., and A.M. conceived and fabricated the first generation of nanogroove cavities. T.J.K. contributed to early ideas leading to this work and commented on the manuscript. E.V. and A.M. discussed the results and contributed to writing and improving the manuscript. C.G. designed and supervised the study, analyzed the data, and wrote the manuscript with the assistance of W.C., P.R., H.H., E.V., and A.M. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data supporting this report are available in the Zenodo repository (*36*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk3106
Materials and Methods
Supplementary Text
Figs. S1 to S17
Tables S1 and S2
References (*37*–*79*)

6 July 2021; accepted 12 October 2021
10.1126/science.abk3106

NANOPHOTONICS

Detecting mid-infrared light by molecular frequency upconversion in dual-wavelength nanoantennas

Angelos Xomalis[1], Xuezhi Zheng[1,2], Rohit Chikkaraddy[1], Zsuzsanna Koczor-Benda[3], Ermanno Miele[1,4,5], Edina Rosta[3], Guy A. E. Vandenbosch[2], Alejandro Martínez[6], Jeremy J. Baumberg[1]*

Coherent interconversion of signals between optical and mechanical domains is enabled by optomechanical interactions. Extreme light-matter coupling produced by confining light to nanoscale mode volumes can then access single mid-infrared (MIR) photon sensitivity. Here, we used the infrared absorption and Raman activity of molecular vibrations in plasmonic nanocavities to demonstrate frequency upconversion. We converted approximately 10-micrometer-wavelength incoming light to visible light by surface-enhanced Raman scattering (SERS) in doubly resonant antennas that enhanced upconversion by more than 10^{10}. We showed 140% amplification of the SERS anti-Stokes emission when an MIR pump was tuned to a molecular vibrational frequency, obtaining lowest detectable powers of 1 to 10 microwatts per square micrometer at room temperature. These results have potential for low-cost and large-scale infrared detectors and spectroscopic techniques.

Infrared spectroscopy delivers information that is difficult to obtain from other frequency bands, such as atmospheric absorption of molecules (greenhouse gases) or thermally emitted radiation from Earth (meteorological maps or imaging wildfires) (*1*–*5*). Although the development of mid-IR (MIR) sources is evolving, a bottleneck continues to be producing low-noise room temperature detectors (*6*). One proposed scheme is to directly upconvert MIR photons into high-energy visible photons that are efficiently detected, potentially delivering single-photon semiconductor-based detectors (*7*–*9*). Analogous wavelength conversion from microwave to optical frequencies has used expensive fabrication and cryogenic temperatures (*10*, *11*), as well as $LiNbO_3$ resonators (*12*, *13*). To access the efficiencies required, strongly enhanced light-matter interactions are paramount. Thus, plasmonic devices and planar resonant metasurfaces that confine light have been of interest for MIR-integrated detection and biosensing (*14*–*16*).

A promising approach for detecting infrared radiation through frequency upconversion is by molecular optomechanical coupling (*17*). Optomechanical interactions allow coherent conversion of signals between the optical and mechanical domains (Fig. 1). Nanocavities containing vibrating molecules act as mechanical oscillators, with MIR-absorbing infrared vibrational modes probed by a visible laser through their Raman scattering (Fig. 1B). The required interactions can be boosted by using the tight light localization inside plasmonic nanocavities <100 nm across, which yield detectable signals even from single vibrational bonds (*18*). The interaction of light and matter in these subnanometer mode volumes gives extreme optomechanical coupling with single MIR-photon sensitivity in principle, but so far this has only been studied theoretically (*17*). The noise-equivalent power of hybrid nanocavity-molecular detectors is predicted to be 100-fold lower than commercial uncooled detectors.

Of vital importance for upconversion efficiency is the optimal spatial overlap of visible and infrared radiation. Plasmonic nanoparticles allow extreme light confinement at visible frequencies, and at longer wavelengths light localization is challenging but can be achieved with suitable designs (*19*, *20*). Achieving light confinement simultaneously in both visible and MIR spectral regions requires a hybrid dual resonator (*21*). Here, this was fulfilled by creating doubly resonant antennas that focus long and short wavelengths into the same active region, allowing extreme optomechanical coupling (Fig. 1D). Their construction combines bottom-up and top-down methods that allow for ease of fabrication and cost-effective, large-scale arrays of devices.

To demonstrate MIR detection, we performed surface-enhanced Raman spectroscopy (SERS)

[1]NanoPhotonics Centre, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, UK. [2]Department of Electrical Engineering (ESAT-TELEMIC), KU Leuven, Leuven, Belgium. [3]Department of Physics and Astronomy, University College London, London, UK. [4]Department of Chemistry, University of Cambridge, Cambridge, UK. [5]The Faraday Institution, Harwell Science and Innovation Campus, Oxford, UK. [6]Nanophotonics Technology Center, Universitat Politècnica de València, Valencia, Spain.
*Corresponding author. Email: jjb12@cam.ac.uk

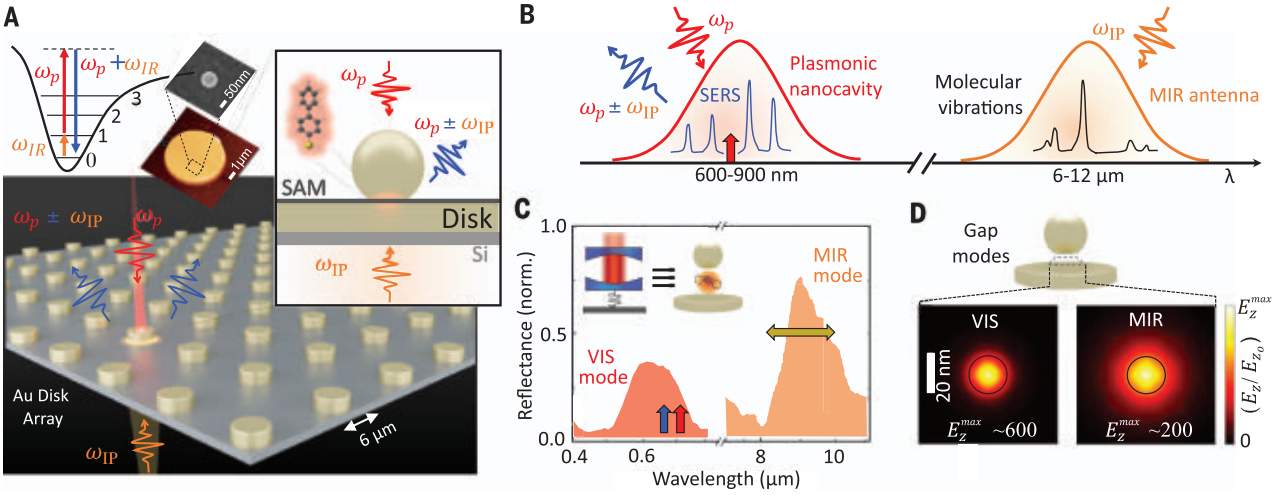


Fig. 1. Dual-wavelength antenna and frequency upconversion. (**A**) Pump (MIR)–probe (visible) detection configuration. Inset shows upconversion process, AFM (disk) and SEM (nanoparticle) images, and a self-assembled monolayer of BPT creating a 1.3 nm cavity between the 60 nm Au nanoparticle and the 6 μm disk. (**B**) Scheme of MIR to visible light upconversion through molecular optomechanics. (**C**) Experimental reflectance of nanoparticle-on-resonator (NPoR) resonances at both visible (red) and MIR (orange) wavelengths. Arrows indicate SERS probe wavelength (785 nm, red), inelastic scattered light (blue), and MIR tuning range (8.5 to 12.6 μm, yellow). Inset shows equivalence of optomechanical cavity and NPoR. (**D**) Near-field normalized maps of MIR and visible gap modes of NPoR. Black circle shows a 20 nm nanoparticle facet.

on self-assembled molecular monolayers with discrete vibrational absorption modes in the λ = 6 to 12 μm range. Coupling requires matching the optical (infrared absorption) and mechanical (molecular vibration) energies. Biphenyl-4-thiol (BPT) was chosen (inset, Fig. 1A) because it provides vibrations that are simultaneously active in both IR absorption and Raman and binds strongly and consistently to Au. Integrated into a dual-wavelength Au antenna, called a nanoparticle-on-resonator (NPoR), this strongly confines visible and MIR light within the same active region (*21*), accessing single-molecule optomechanical nonlinearities (*22*, *23*). The Au disk resonators (diameter 6 μm) have a fundamental resonant mode around λ = 10 μm and high-order modes in the visible spectrum (*21*). Onto these is self-assembled a molecular monolayer of BPT, with 60 nm Au nanoparticles drop-cast on top. The molecule length sets the 1.3 nm spacing (*18*), giving resonances that are experimentally measured with visible and MIR light (Fig. 1C). Comparison with simulations shows field enhancements E/E_0 > 500 (visible) and >200 (MIR) (fig. S4A) (*21*), providing a more favorable geometry than previously devised for (simulating) molecular upconversion (*17*). A modified microscope focuses visible and MIR lasers onto the same NPoR nanodevice (with >40 NPoRs tested here). The 1080 cm^{-1} molecular vibration was observed in SERS anti-Stokes emission, with amplitude that increased linearly when pumped directly with MIR radiation tuned to the same energy.

Our experiments used synchronized visible and quantum cascade laser (QCL) rectangular pulses (0.4 μs) to collect SERS spectra with and without the MIR light (Fig. 2). These confirmed the prediction of frequency upconversion (*17*) using the ν = 1080 cm^{-1} BPT mode, which is both infrared and Raman active (Figs. 3A and 4A). Measuring SERS from NPoRs shows the expected BPT vibrations on both the Stokes and anti-Stokes sides of the laser (Fig. 3B), which are stable and repeatable over long periods. The QCL is then tuned to the same photon energy $h\nu$ (orange, Fig. 3B) and an infrared pump power dependence recorded (Fig. 3C). We found that the anti-Stokes SERS was 100% higher from $NPoR_1$ when QCL average powers of 5 μW/$μm^2$ were incident [Fig. 3D; using peak area ratio $AS(QCL_{on})/AS(QCL_{off})$ with background-subtracted anti-Stokes peaks, peak power is 12 times larger; see materials and methods section S2]. The expected linear dependence of frequency upconversion with pump power was similar for the different NPoRs (red points, Fig. 3C). The lowest detectable light intensity of these dual-wavelength plasmonic antennas was ~1 μW/$μm^2$ (Fig. 3C), whereas the lock-in detection synchronized technique here showed that the response speed was submicrosecond,

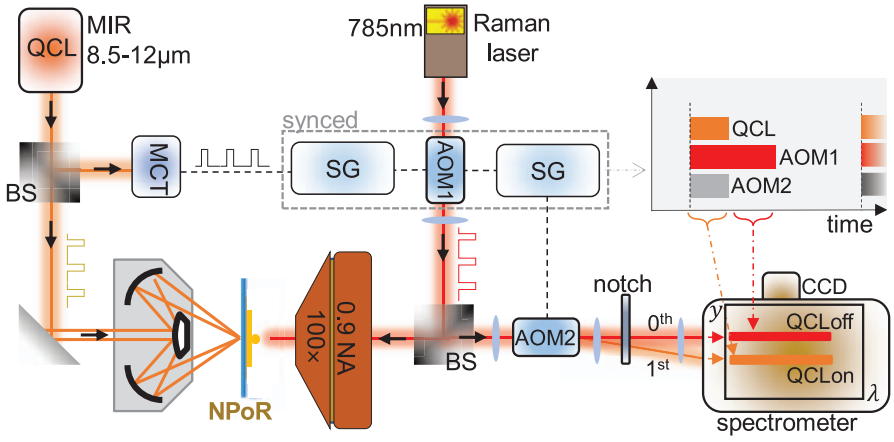


Fig. 2. MIR and visible spectroscopy. Dual microscope combines visible probe and MIR pump for frequency upconversion of molecules in nanogaps: AOM, acousto-optical modulator; MCT, mercury-cadmium-telluride detector; BS, beam splitter; SG, signal generator. Inset: Timing sequence of each repetition of QCL (pump) to Raman laser modulation (AOM1). AOM2 deflects each SERS spectrum to different vertical positions *y* on the spectrometer slit, extracting Raman spectra versus λ for QCL on/off.

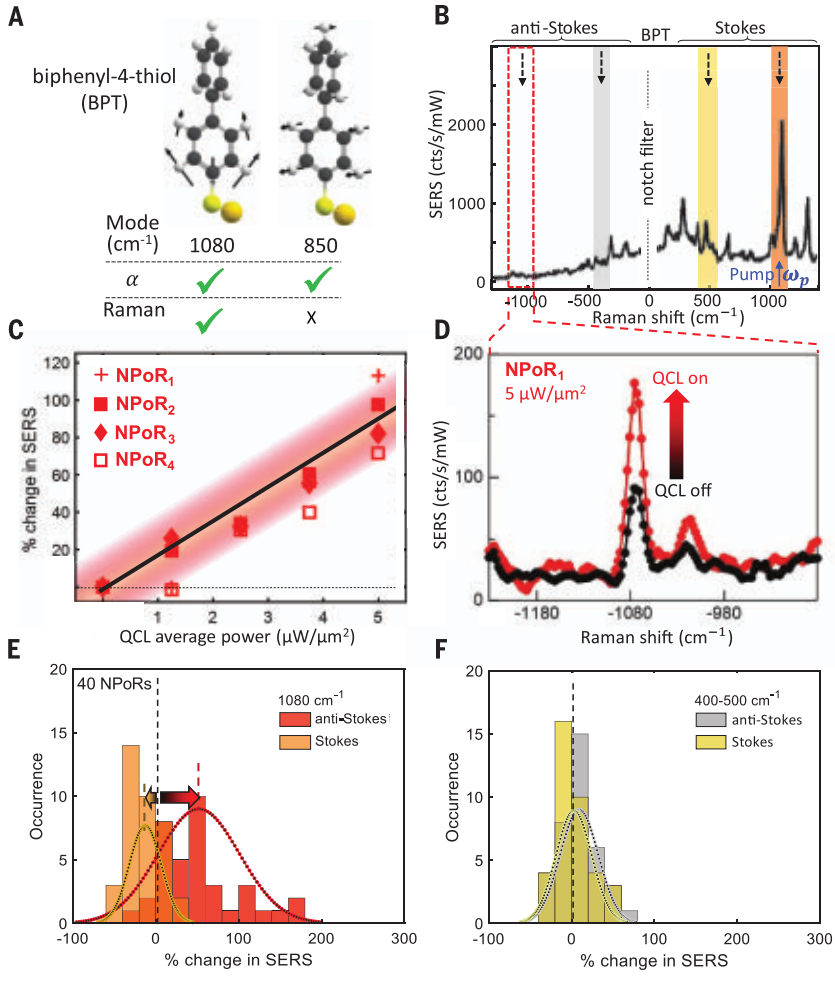


Fig. 3. Upconversion of MIR to visible photons in doubly resonant plasmonic antennas. (**A**) Vibrations of BPT showing frequencies with strong infrared absorption (α) or Raman. (**B**) BPT SERS spectrum from the 785 nm probe alone. Shaded regions mark pump (orange, 1080 cm^{-1}) and monitored frequency bands (arrows). (**C**) Power dependence for four NPoRs. (**D**) Raw spectra showing the ν = 1080 cm^{-1} anti-Stokes increase when MIR pump is on (red). (**E** and **F**) MIR-induced change in SERS of 40 NPoRs at Stokes and anti-Stokes peaks at 1080 cm^{-1} (E) and 400 to 500 cm^{-1} (F).

much faster than the QCL pulse repetition rate (5 μs).

To better quantify the upconversion efficiency, we measured the percentage change of SERS on 40 NPoRs, where each NP is located at different positions on each disk antenna. These showed an average 52% increase of anti-Stokes at ν (red, Fig. 3E) for 5 μW/μm^2 MIR average power, whereas the Stokes at ν showed a decrease of 13% (red). No systematic correlation with the nanoparticle position on the disk was apparent, although it likely controls in-coupling of both visible and MIR light into the nanogap.

To confirm the frequency upconversion mechanism, the percentage SERS changes were also extracted for the 400 to 500 cm^{-1} spectral region (yellow and gray shaded areas for Stokes and anti-Stokes, respectively, in Fig. 3B). These low-frequency vibrational modes showed no discernible change within the ±10% signal noise (Fig. 3F). This lack of low wave number signal shows that the signal was not simply thermal heating (fig. S5), as was also suggested by the submicrosecond response, but rather was a nonequilibrium response. If simple heating were involved, then a trebling of anti-Stokes at 1080 cm^{-1} would give a 60% increase at 450 cm^{-1}, which was not observed (see the supplementary materials, section S8).

To understand the frequency-selective dependence, we calculate the product [in m^3/(mol·sr)] of infrared absorption and Raman intensity of BPT, averaged over all orientations for each normal mode (Fig. 4A and supplementary materials, section S1) (*24*). This clearly showed that the optimum overlap of optical and vibrational modes was at 1080 cm^{-1}, and that the dipoles were all well aligned with the vertical E field in the nanogap at both visible and MIR wavelengths. To confirm this, we tuned the QCL from 795 to 1170 cm^{-1} in 15 cm^{-1} steps, ensuring a constant 5 μW/μm^2 incident on the sample. Although the NPoR device showed a resonant anti-Stokes increase of 140% at 1080 cm^{-1}, it showed no increase elsewhere across the frequency scan (red, Fig. 4B), and neither did the Stokes signal (orange). No clear change in SERS intensity was seen for the 400 to 500 cm^{-1} lines across this MIR tuning on the same NPoR (Fig. 4C). These data clearly distinguish the direct resonant pumping of the optimum 1080 cm^{-1} mode.

The quantum efficiency of these devices was estimated by calibrating to the thermal scale of anti-Stokes emission. At room temperature (T = 300 K), with MIR powers of P =100 μW (intensity I = 5 μW/μm^2) at 1080 cm^{-1} ($h\nu$ = 0.13 eV) and assuming decay times from the first vibrational state of τ = 1 ps (*22*, *25*), using the measured anti-Stokes increase of Δζ = 100% (Fig. 3C) gives the fraction of MIR photons arriving that result in an upconverted vibrational response as (see the supplementary materials, section S9)

$$\eta = \Delta\zeta \exp\left\{-\frac{h\nu}{k_B T}\right\}\left[\left(\frac{P}{h\nu}\right)\tau\right]^{-1} \quad (1)$$

corresponding to photon quantum efficiency η ~2 × 10^{-6} in this first generation of devices. The induced occupation of the first vibrational level was estimated to be $\Delta\zeta\exp\{-h\nu/k_B T\}$ ~ 1%. Theoretical estimates showed a similar efficiency (*17*)

$$\eta_t = \eta_{IR}\eta_o \frac{g^2\tau}{\kappa_{IR}} \sim 1 \times 10^{-6} \quad (2)$$

using the measured MIR linewidth (Fig. 1C) to get the antenna loss rate κ_{IR} ~2.7 THz, an antenna efficiency η_{IR} ~0.5, and an optomechanical coupling g ~2 GHz for BPT molecules in the nanocavity gap (*22*, *23*). This assumes that the optical cross-section of the dual-wavelength antenna matches the incident MIR focus. We estimated the overlap efficiency of nanocavity modes at MIR and Raman wavelengths as η_o = 66% (see the supplementary materials, section 10), covering ~260 BPT molecules. The main inefficiency in η_{IR} was in the fraction of MIR photons giving significant field inside the NPoR gap to vibrationally excite a molecule. Improving the Q factor of the antenna, for example, by using hybridization with photonic cavities, is needed for further enhancements (*26*). Of 40 measured NPoR devices, 75% showed an upconversion response above the noise (Fig. 3E).

We have also shown that it is possible to fabricate these integrated NPoR detectors using SiN waveguides on standard 4-inch Si wafers (*27*) in a cheap and scalable combination of top-down and bottom-up lithographies (fig. S3). Prospects for multiband operation are promising [selection of optimal molecules is required to broaden the vibrational range and responsivity (*24*)] because lower-frequency anti-Stokes emission has already been observed at 250 cm^{-1} (Fig. 3B, λ = 40 μm or 7.5 THz). Using alternative molecules embedded in NPoRs, SERS lines observed at ~160 cm^{-1} can access targets for astronomical detectors (OH line at 4.7 THz and lower). Although the lifetime of such devices is not yet fully characterized, it already exceeds 1 month, extended by suitable encapsulation to exclude oxygen. The rapid relaxation of nonresonant molecules in the virtual Raman process is encouraging for engineering robust performance. We emphasize that further increases in sensitivity can come from exploiting single-atom picocavities, which deliver 100-fold larger SERS signals from the enhanced light localization around single Au adatoms (*28*, *29*), with simple estimates based on Eq. 2 using the measured g ~5 THz (*22*, *23*) giving near unity upconversion efficiencies. This makes current efforts to stabilize picocavities significant and optimizes the overlap of MIR light in the same volume.

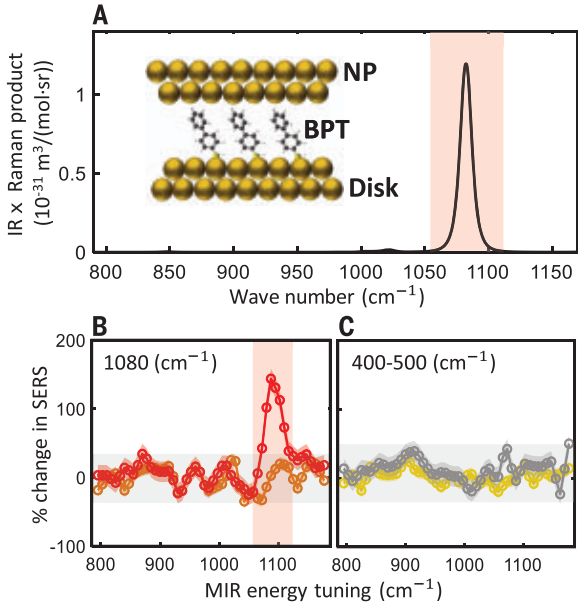


Fig. 4. MIR tuning dependence of upconversion in NPoR plasmonic construct. (**A**) Calculated product of molecular infrared absorption and Raman cross-section for BPT in plasmonic nanogaps (inset). (**B** and **C**) Percentage change in SERS from illuminated NPoR versus MIR frequency of 1080 cm^{-1} Stokes (orange) and anti-Stokes (red) peaks (B), as well as 400 to 500 cm^{-1} Stokes (yellow) and anti-Stokes (gray) peaks (C).

REFERENCES AND NOTES

1. M. J. Elrod, *J. Chem. Educ.* **76**, 1702 (1999).
2. A. Kuze, H. Suto, M. Nakajima, T. Hamazaki, *Appl. Opt.* **48**, 6716–6733 (2009).
3. R. S. Allison, J. M. Johnston, G. Craig, S. Jennings, *Sensors* **16**, 1310 (2016).
4. W. R. Bandeen, R. Hanel, J. Licht, R. Stampfl, W. Stroud, *J. Geophys. Res.* **66**, 3169–3185 (1961).
5. R. Hanel *et al.*, *J. Geophys. Res.* **77**, 2629–2641 (1972).
6. A. Rogalski, *Infrared Phys. Technol.* **43**, 187–210 (2002).
7. J. S. Dam, P. Tidemand-Lichtenberg, C. Pedersen, *Nat. Photonics* **6**, 788–793 (2012).
8. I. Kviatkovsky, H. M. Chrzanowski, E. G. Avery, H. Bartolomaeus, S. Ramelow, *Sci. Adv.* **6**, eabd0264 (2020).
9. S. Junaid *et al.*, *Optica* **6**, 702–708 (2019).
10. M. Forsch *et al.*, *Nat. Phys.* **16**, 69–74 (2020).
11. J. Bochmann, A. Vainsencher, D. D. Awschalom, A. N. Cleland, *Nat. Phys.* **9**, 712–716 (2013).
12. W. Jiang *et al.*, *Nat. Commun.* **11**, 1166 (2020).
13. L. Shao *et al.*, *Optica* **6**, 1498–1505 (2019).
14. F. Yesilkoy *et al.*, *Nat. Photonics* **13**, 390–396 (2019).
15. A. Tittl *et al.*, *Science* **360**, 1105–1109 (2018).

16. B. Schwarz *et al*., *Nat. Commun.* **5**, 4085 (2014).
17. P. Roelli, D. Martin-Cano, T. J. Kippenberg, C. Galland, *Phys. Rev. X* **10**, 031057 (2020).
18. J. J. Baumberg, J. Aizpurua, M. H. Mikkelsen, D. R. Smith, *Nat. Mater.* **18**, 668–678 (2019).
19. O. D. Miller *et al*., *Opt. Express* **24**, 3329–3364 (2016).
20. R. Chikkaraddy, A. Xomalis, L. A. Jakob, J. J. Baumberg, Mid-infrared-perturbed molecular vibrational signatures in plasmonic nanocavities. arXiv:2108.10171 [physics.optics] (2021).
21. A. Xomalis *et al*., *Nano Lett.* **21**, 2512–2518 (2021).
22. A. Lombardi *et al*., *Phys. Rev. X* **8**, 011016 (2018).
23. F. Benz *et al*., *Science* **354**, 726–729 (2016).
24. Z. Koczor-Benda *et al*., *Phys. Rev. X* (2021).
25. S. Yampolsky *et al*., *Nat. Photonics* **8**, 650–656 (2014).
26. A. I. Barreda, M. Zapata-Herrera, I. Palstra, L. Mercadé, J. Aizpurua, A. F. Koenderink, A. Martínez, Hybrid photonic-plasmonic cavities based on the nanoparticle-on-a-mirror configuration. arXiv:2106.01931 [physics.optics] (2021).
27. J. Losada *et al*., *IEEE J. Sel. Top. Quantum Electron.* **25**, 1–6 (2019).
28. T. Wu, W. Yan, P. Lalanne, *ACS Photonics* **8**, 307–314 (2021).
29. T. Wu, M. Gurioli, P. Lalanne, *ACS Photonics* **8**, 1522–1538 (2021).
30. Data for: A. Xomalis, X. Zheng, R. Chikkaraddy, Z. Koczor-Benda, E. Miele, E. Rosta, G. A. E. Vandenbosch, A. Martínez, J. J. Baumberg, Detecting mid-infrared light by molecular frequency upconversion in dual-wavelength nanoantennas, University of Cambridge Repository (2021); https://doi.org/10.17863/CAM.77640.

ACKNOWLEDGMENTS

Funding: This work was supported by the European Union's Horizon 2020 Research and Innovation Program under grant agreements 829067 (THOR), 861950 (POSEIDON), and 883703 (PICOFORCE) and by the Engineering and Physical Sciences Research Council (EPSRC) (Cambridge NanoDTC grants EP/L015978/1, EP/L027151/1, EP/S022953/1, EP/P029426/1, and EP/R020965/1). X.Z. acknowledges support from KU Leuven Internal Funds C14/19/083, IDN/20/014, KA/20/019, and FWO G090017N. R.C. acknowledges support from Trinity College, University of Cambridge. Z.K.B. and E.R. acknowledge funding from the EPSRC (EP/R013012/1, EP/L027151/1) and ERC project 757850 BioNet. We are grateful to the UK Materials and Molecular Modelling Hub, which is partially funded by EPSRC (EP/P020194/1), for computational resources. **Author contributions:** A.X. fabricated the devices, performed the experiments, and analyzed the data. Z.X. and G.V. did electromagnetic calculations. R.C. and A.X. did full-wave simulations. S.K.B. performed DFT calculations. E.M. and A.X. performed SERS of NPoR on Si chips. E.R., A.M., and J.J.B. designed and supervised the work. All authors discussed the results, provided feedback, and contributed to the writing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the study are present in the main text or the supplementary materials. Source data can be found at the University of Cambridge Repository (*30*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk2593
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*31*–*42*)

6 July 2021; accepted 12 October 2021
10.1126/science.abk2593

EXOPLANETS

GJ 367b: A dense, ultrashort-period sub-Earth planet transiting a nearby red dwarf star

Kristine W. F. Lam[1,2]*†, Szilárd Csizmadia[2]†, Nicola Astudillo-Defru[3], Xavier Bonfils[4], Davide Gandolfi[5], Sebastiano Padovan[2,6], Massimiliano Esposito[7], Coel Hellier[8], Teruyuki Hirano[9]‡, John Livingston[10], Felipe Murgas[11,12], Alexis M. S. Smith[2], Karen A. Collins[13], Savita Mathur[11,12], Rafael A. Garcia[14,15], Steve B. Howell[16], Nuno C. Santos[17,18], Fei Dai[19], George R. Ricker[20], Roland Vanderspek[20], David W. Latham[13], Sara Seager[20,21,22], Joshua N. Winn[23], Jon M. Jenkins[16], Simon Albrecht[24], Jose M. Almenara[4], Etienne Artigau[4], Oscar Barragán[25], François Bouchy[26], Juan Cabrera[4], David Charbonneau[13], Priyanka Chaturvedi[7], Alexander Chaushev[1], Jessie L. Christiansen[27], William D. Cochran[28], José R. De Meideiros[29], Xavier Delfosse[4], Rodrigo F. Díaz[30], René Doyon[31], Philipp Eigmüller[2], Pedro Figueira[32,17], Thierry Forveille[4], Malcolm Fridlund[33,34], Guillaume Gaisné[4], Elisa Goffo[5,7], Iskra Georgieva[33], Sascha Grziwa[35], Eike Guenther[7], Artie P. Hatzes[7], Marshall C. Johnson[36], Petr Kabáth[37], Emil Knudstrup[24], Judith Korth[35,38], Pablo Lewin[39], Jack J. Lissauer[16,40], Christophe Lovis[26], Rafael Luque[11,12], Claudio Melo[32], Edward H. Morgan[20], Robert Morris[16,41], Michel Mayor[26], Norio Narita[42,43,44,11], Hannah L. M. Osborne[45], Enric Palle[11,12], Francesco Pepe[26], Carina M. Persson[33], Samuel N. Quinn[13], Heike Rauer[2,1,46], Seth Redfield[47], Joshua E. Schlieder[48], Damien Ségransan[26], Luisa M. Serrano[5], Jeffrey C. Smith[16,41], Ján Šubjak[37,49], Joseph D. Twicken[16,41], Stéphane Udry[26], Vincent Van Eylen[45], Michael Vezie[20]

Ultrashort-period (USP) exoplanets have orbital periods shorter than 1 day. Precise masses and radii of USP exoplanets could provide constraints on their unknown formation and evolution processes. We report the detection and characterization of the USP planet GJ 367b using high-precision photometry and radial velocity observations. GJ 367b orbits a bright (*V*-band magnitude of 10.2), nearby, and red (M-type) dwarf star every 7.7 hours. GJ 367b has a radius of 0.718 ± 0.054 Earth-radii and a mass of 0.546 ± 0.078 Earth-masses, making it a sub-Earth planet. The corresponding bulk density is 8.106 ± 2.165 grams per cubic centimeter—close to that of iron. An interior structure model predicts that the planet has an iron core radius fraction of 86 ± 5%, similar to that of Mercury's interior.

Red dwarf stars of spectral type M (M dwarfs) are cool stars with effective temperatures (T_{eff}) below ~4000 K (*1*). They have masses and radii less than ~60% the size of those of the Sun and are the most abundant type of stars in the solar neighborhood (*2*–*4*). It has been estimated that M dwarfs host an average of 2.5 ± 0.2 small planets [planet radius R_p < 4 Earth-radii ($R_\oplus$)] with periods <100 days (*5*). Because of the small stellar radius, the transit signal produced by a planet orbiting an M dwarf is larger than that of a planet of the same size orbiting a solar-type star (G dwarf). The radial velocity (RV) signal induced by a planet is also larger for an M dwarf host than for that of a G dwarf, as a result of the lower stellar mass. M dwarfs therefore provide an opportunity to search for exoplanets with a small radius and low mass. However, young M dwarfs often have high stellar activity, which gives rise to noise in the RV observations (*6*). RV analysis can be complicated even for old, inactive M dwarfs because their slow rotation periods have harmonics in the range of periods where small planets are sought (*7*).

GJ 367 (also cataloged as TOI-731) is an M dwarf located 9.41 pc from the Sun (*8*) with a brightness of 10.153 magnitudes in the optical *V* band and 5.78 magnitudes in the infrared *K* band. We observed this star with the High Accuracy Radial Velocity Planet Searcher (HARPS) spectrograph (*9*) and determined its stellar properties. GJ 367 has an effective temperature of T_{eff} = 3519 ± 70 K, a stellar mass M_s = 0.454 ± 0.011 solar masses ($M_\odot$), a stellar radius R_s = 0.457 ± 0.013 solar radii ($R_\odot$), and a stellar luminosity L_s = 0.0288 ± 00029 solar luminosities ($L_\odot$) (*9*) (Table 1).

The Transiting Exoplanet Survey Satellite (TESS) (*10*) observed GJ 367 during sector 9 of its survey. TESS acquired optical photometry at 2-min cadence for 27 days from 28 February 2019 to 26 March 2019. The light curve (brightness as a function of time) was extracted using the Science Processing Operations Center (SPOC) pipeline (*11*). This revealed a planet candidate with an orbital period of 0.32 days and a radius of 0.75 $R_\oplus$, which was designated TOI-731.01 by the TESS Science Office on the basis of the SPOC transit search and data validation results. We also searched for transit signals using the Détection Spécialisée de Transits (DST) algorithm (*12*), which indicated a transit-like signal every 0.32 days and a transit depth of ~0.03%, corresponding to the transit of a sub-Earth–sized planet (Fig. 1).

We performed several tests to ensure that the candidate was not a false positive. Comparison of photometric data using varying aperture sizes showed no correlation between the aperture size and transit depth, indicating that the transit signal is not from another source blended with GJ 367. We performed

follow-up ground-based photometry and found no contamination from eclipsing binaries up to 2.5 arc sec from the target star (*9*). The follow-up photometry shows that nearby stellar sources contribute 9.5 ± 1.2% of the flux within the TESS optimal aperture. This contamination reduces the transit depth, causing analysis of the TESS light curve to underestimate the planet's radius by ~5% (*9*). We account for this dilution factor in a transit model to obtain the true planet radius (Table 1). The density of the host star, ρ_s, was derived from the transit light curve (*9*), finding 7.64 ± 3.51 g cm^{-3}, which is consistent with the value $\rho_s = 6.71^{+0.61}_{-0.55}$ g cm^{-3} determined from the spectral analysis discussed above (*9*).

In a further test, we performed a frequency analysis of the HARPS RV measurements and stellar activity indicators (*9*). The periodogram of the RVs has a peak at orbital frequency (*f*) of 3.103 d^{-1} (*P* = 0.322 days) that has no counterpart in the periodograms of the activity indicators (fig. S4), consistent with a planetary origin. A further 45-day signal is present in the RV periodogram and in the activity indicators. Our analysis of archival photometry from the Wide Angle Search for Planets (WASP) survey indicates a stellar rotational period of 48 ± 2 days (*9*). GJ 367's Ca(II) activity index is log R'_{HK} = −5.214 ± 0.074, which corresponds to an estimated stellar rotation period of 58.0 ± 6.9 days (*9*). This indicates that the 45-day signal likely originates from active regions on the stellar surface. We conclude that the 0.322-day period is the result of an ultrashort-period (USP) planet, GJ 367b.

Using a priori information on the host star properties from our spectral analysis, we derived the physical properties of the GJ 367 system using a Bayesian Markov chain Monte Carlo (MCMC) code, Transit and Light Curve

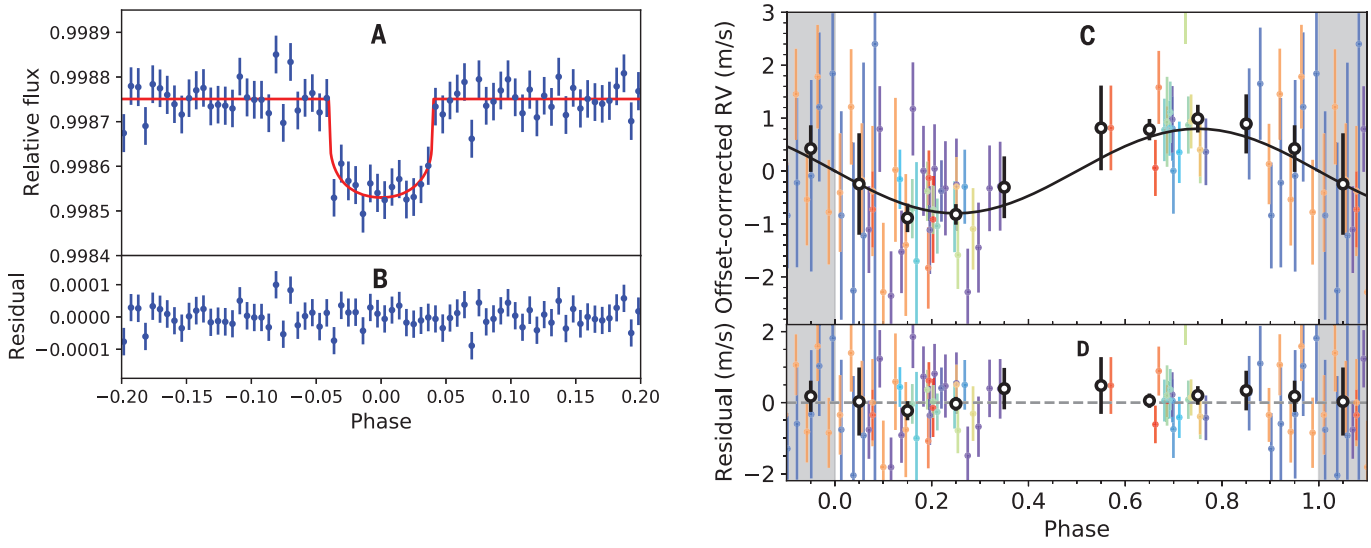


Fig. 1. Phase-folded RV and light curve of GJ 367. (**A**) Phase-folded, 2.6-min binned TESS light curve (blue circles) of GJ 367 with the best-fitting transit model (red line). Error bars show the 1-sigma uncertainties of the binned values. (**B**) The residuals of the light curve. A noise-correction model has been applied to the data (*9*). (**C**) Phase-folded HARPS RV data for GJ 367. Different color dots correspond to different corrections applied to the RV model (*9*). Black open circles are the RV data phase-binned in intervals of 0.10. The solid black line shows the best-fitting RV model, which has a semiamplitude of 79.8 ± 11.0 cm s^{-1}. (**D**) The corresponding residuals of the RV data. In (C) and (D), the RV orbital phase limits extend beyond phases 0 to 1 (shaded gray regions), so the first and last data points are duplicated.

[1]Centre for Astronomy and Astrophysics, Technical University Berlin, 10585 Berlin, Germany. [2]Institute of Planetary Research, German Aerospace Center, 12489 Berlin, Germany. [3]Departamento de Matemática y Física Aplicadas, Universidad Católica de la Santísima Concepción, Concepción, Chile. [4]Université Grenoble Alpes, Centre national de la recherche scientifique, Institut de Planétologie et d'Astrophysique de Grenoble, F-38000 Grenoble, France. [5]Dipartimento di Fisica, Università degli Studi di Torino, I-10125, Torino, Italy. [6]WorkGroup Solutions GmbH at European Organisation for the Exploitation of Meteorological Satellites, 64295 Darmstadt, Germany. [7]Thüringer Landessternwarte Tautenburg, D-07778 Tautenberg, Germany. [8]Astrophysics Group, Keele University, Staffordshire, ST5 5BG, UK. [9]Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Tokyo, Japan. [10]Department of Astronomy, University of Tokyo, Tokyo, Japan. [11]Instituto de Astrofísica de Canarias, 38205 La Laguna, Tenerife, Spain. [12]Departamento de Astrofísica, Universidad de La Laguna, 38206 La Laguna, Tenerife, Spain. [13]Center for Astrophysics, Harvard and Smithsonian, Cambridge, MA, USA. [14]Institut de Recherche sur les Lois Fondamentales de l'Universe, Commissariat à l'Énergie Atomique et aux énergies alternatives, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France. [15]Astrophysique, Instrumentation et modélisation, Commissariat à l'Énergie Atomique et aux énergies alternatives, Centre National de la recherche scientifique, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France. [16]NASA Ames Research Center, Moffett Field, CA, USA. [17]Instituto de Astrofísica e Ciênciasdo Espaço, Universidade do Porto, Centro de Astrofísica da Universidade do Porto, 4150-762 Porto, Portugal. [18]Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal. [19]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. [20]Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [21]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [22]Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA. [23]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA. [24]Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark. [25]Subdepartment of Astrophysics, Department of Physics, University of Oxford, Oxford, OX1 3RH, UK. [26]Geneva Observatory, University of Geneva, 1290 Versoix, Switzerland. [27]Infrared Processing and Analysis Center, Caltech, Pasadena, CA, USA. [28]Center for Planetary Systems Habitability and McDonald Observatory, The University of Texas, Austin, TX, USA. [29]Departamento de Física, Universidade Federal do Rio Grande do Norte, 59072-970 Natal, RN, Brazil. [30]International Center for Advanced Studies and Instituto de Ciencias Físicas (Consejo Nacional de Investigaciones Científicas y Técnicas), Escuela de Ciencia y Tecnología - Universidad Nacional de San Martín, Campus Miguelete, Buenos Aires, Argentina. [31]Institut de Recherche sur les Exoplantes, Dpartement de Physique, Universit de Montral, Montral, QC, H3C 3J7, Canada. [32]European Southern Observatory, Vitacura, Santiago, Chile. [33]Department of Space, Earth and Environment, Chalmers University of Technology, Onsala Space Observatory, 439 92 Onsala, Sweden. [34]Leiden Observatory, University of Leiden, 2300 RA, Leiden, Netherlands. [35]Rheinisches Institut für Umweltforschung an der Universität zu Köln, D-50931 Köln, Germany. [36]Las Cumbres Observatory, Goleta, CA, USA. [37]Astronomical Institute, Czech Academy of Sciences, 25165 Ondřejov, Czech Republic. [38]Department of Space, Earth and Environment, Astronomy and Plasma Physics, Chalmers University of Technology, 412 96 Gothenburg, Sweden. [39]The Maury Lewin Astronomical Observatory, Glendora, CA, USA. [40]Geological Sciences Department, Stanford University, CA, USA. [41]Search for Extraterrestrial Intelligence Institute, Mountain View, CA, USA. [42]Komaba Institute for Science, The University of Tokyo, Tokyo, Japan. [43]Japan Science and Technology Agency, Precursory Research for Embryonic Science and Technology, Tokyo, Japan. [44]Astrobiology Center, Tokyo, Japan. [45]Mullard Space Science Laboratory, University College London, Dorking, Surrey, RH5 6NT, UK. [46]Institute of Geological Sciences, Freie Universität Berlin, D-12249 Berlin, Germany. [47]Astronomy Department and Van Vleck Observatory, Wesleyan University, Middletown, CT, USA. [48]NASA Goddard Space Flight Center, Greenbelt, MD, USA. [49]Astronomical Institute of Charles University, 180 00 Prague, Czech Republic.
***Corresponding author. Email: kristine.lam@dlr.de** †These authors contributed equally to this work. ‡Present address: Astrobiology Center, 2-21-1 Osawa, Mitaka, Tokyo, Japan.

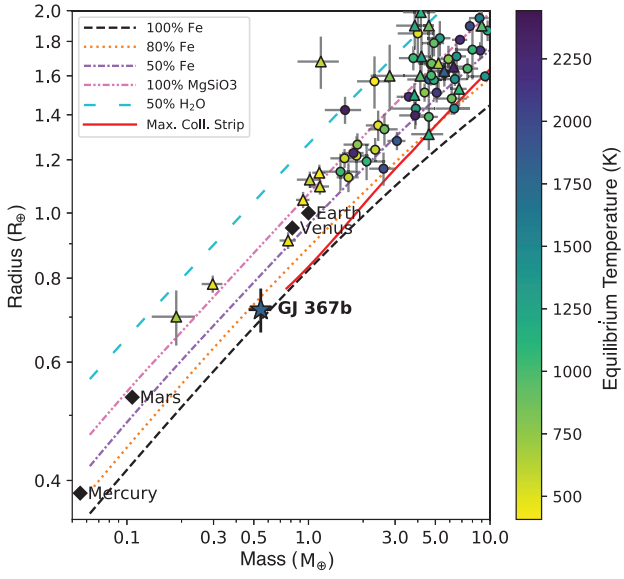
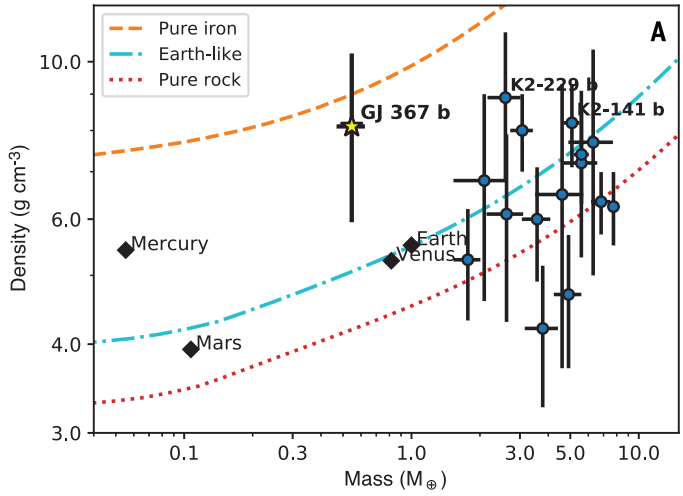


Fig. 2. Masses and radii of small planets (<2 $R_\oplus$) that have both quantities measured with precision ≤30%. Symbols indicate masses determined with RVs (circles) and transit timing variations (triangles), Solar System planets (diamonds), and GJ 367b (star). Error bars show 1-sigma uncertainties. Exoplanet symbols are color coded according to the equilibrium dayside temperatures (color bar). Theoretical mass-radius relations for two-layer rocky planets (*22*) are shown with lines corresponding to different core mass fractions. These cores consist of pure iron, pure rock (100% MgSiO3), or a two-layer core with a mixture of iron and rock or rock and H_2O ice, as indicated in the legend. The solid red line denotes the lower limit on planet radius after collisional stripping (*31*). GJ 367b is likely an iron-dominated planet.

Table 1. Properties of host star GJ 367 and planet GJ 367b. The stellar parameters were derived from the spectral analysis of the HARPS data (*9*). Planet parameters were obtained from the joint model fitting of the TESS photometry and HARPS RVs (*9*). Reported values are the medians of the posterior probability distributions with uncertainties of the 34th and 68th percentiles of those distributions.

Parameter	Value
Star GJ 367 (TOI 731)	
Right ascension (J2000 equinox)	$09^h44^m29.84^s$
Declination (J2000 equinox)	−45°46′35.43″
TESS-band magnitude	8.032 ± 0.007
V-band magnitude	10.153 ± 0.044
Parallax* (milli–arc sec)	106.272 ± 0.056
Distance, *d* (pc)	9.410 ± 0.005
Effective temperature, T_{eff} (K)	3522 ± 70
Stellar mass, M_s ($M_\odot$)	0.454 ± 0.011
Stellar radius, R_s ($R_\odot$)	0.457 ± 0.013
Stellar density, ρ_s ($\rho_\odot$)	$4.76^{+0.43}_{-0.39}$
Metallicity, [Fe/H]	−0.01 ± 0.12
Surface gravity, log *g*	4.777 ± 0.026
Luminosity, L_s ($L_\odot$)	0.0288 ± 0.0027
Spectral type	M1.0V
Planet GJ 367b	
Epoch, T_0 [barycentric Julian date (BJD_{TDB})]	2458544.1348 ± 0.0004
Orbital period, *P* (days)	$0.321962^{+0.000010}_{-0.000012}$
Planet-to-star radius ratio, R_p/R_s	$0.0143^{+0.0096}_{-0.0010}$
Scaled orbital semimajor axis, a/R_s	$3.41^{+0.06}_{-0.07}$
Impact parameter, *b*	$0.55^{+0.03}_{-0.04}$
RV semiamplitude†, *K* (cm s^{-1})	79.8 ± 11.0
Systemic RV‡, v_γ (km s^{-1})	47.9258 ± 0.0003
Eccentricity, *e*	0
Transit duration, T_{14} (min)	$36.9^{+1.0}_{-0.9}$
Orbital semimajor axis, *a* (astronomical units)	0.0071 ± 0.0002
Orbital inclination, *i* (°)	80.75 ± 0.64
Planet mass, M_p ($M_\oplus$)	0.546 ± 0.078
Planet radius, R_p ($R_\oplus$)	0.718 ± 0.054
Planet bulk density, ρ_p (g cm^{-3})	8.106 ± 2.165
Equilibrium dayside temperature§, assuming an Earth-like bond albedo (A_b = 0.3), T_{eq} (K)	1597 ± 39
Equilibrium dayside temperature§, assuming zero bond albedo, T_{eq} (K)	1745 ± 43

*A correction of +61 milli–arc sec was applied to the Gaia parallax (*9*). †RV induced by the orbiting planet. ‡RV of the star-planet system with respect to the observer. The uncertainty only reflects the internal precision of HARPS and does not account for systematic effects, such as gravitational redshift. §Assuming no atmospheric circulation.



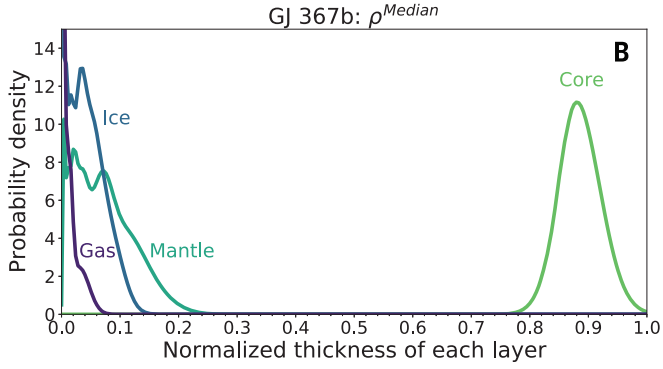


Fig. 3. Bulk composition of USP planets and predicted interior structure of GJ 367b. (**A**) Mass-density diagram for USP (P_{orb} < 1 day) exoplanets with low mass (<10 $M_\oplus$) and measured mass precisions ≤30%. Inner Solar System planets are shown as black diamonds. Planet interior composition models (*22*) are shown with lines indicated in the legend. The bulk densities of low-mass USP planets are usually consistent with terrestrial compositions (pure rock or Earth-like). GJ 367b is more consistent with pure iron and an interior similar to that of Mercury. (**B**) The predicted relative thicknesses of each interior layer of GJ 367b from the neural network model (*32*). The core is assumed to be a liquid Fe-FeS alloy. The mantle is assumed to be composed of olivine and orthopyroxene enstatite in the upper mantle and bridgmanite and magnesiowüstite in the lower mantle. The ice layer is assumed to be water ice VII, and the gas layer consists of hydrogen and helium. The interior composition of GJ 367b was computed using the median mass and radius measurements (corresponding to the derived median planet density ρ^{Median} = 8.106 g cm^{-1}). We infer an iron core filling 86 ± 5% of the planet radius with <1% ice and gas, similar to the interior of Mercury, which has an iron core radius fraction of 83 ± 2% (*30*). If we take the lowest density of GJ 367b permitted by the observations, 5.941 g cm^{-1}, the planet iron core radius fraction is still higher than that of Earth (fig. S7).

Modeller (*13*), to model the transit photometry and RV data simultaneously (*9*). Table 1 reports the physical properties of the planetary system from this analysis. The transit depth of 212 ± 42 parts per million (ppm) and RV semiamplitude K = 79.8 ± 11.0 cm s^{-1} correspond to a planetary radius of 0.718 ± 0.054 $R_\oplus$ and a planetary mass of 0.546 ± 0.078 Earth-masses ($M_\oplus$). Figure 1 shows the phase-folded light curve and RV measurements of GJ 367 along with the corresponding best-fitting transit and RV models. We find that GJ 367b is a sub-Earth planet with a high expected signal-to-noise metric for emission spectroscopy (see supplementary text). The planet receives high stellar irradiation because of its close proximity to the host star, ~576 times the incident flux on Earth. This corresponds to a dayside temperature of 1745 ± 43 K (assuming zero Bond albedo), which is high enough to evaporate any primordial atmosphere (*14*–*16*) and begin to melt or vaporize any silicates or metallic iron (*17*).

The measured mass and radius of GJ 367b imply a bulk density of 8.106 ± 2.165 g cm^{-3}. The bulk composition of a planet can be estimated from theoretical mass-radius relations (*18*–*21*). Figure 2 shows the mass and radius distribution of small planets (R_p below 2 $R_\oplus$) along with theoretical predictions for rocky planets (*21*, *22*). GJ 367b has a mass and radius consistent with an interior dominated by an iron core. This is similar to two larger USP planets, K2-229b (*23*) and K2-141b (*24*, *25*), which have enhanced iron fractions (Fig. 3A). Other known planets with similar sizes to GJ 367b, such as Kepler-138 b (*26*, *27*) and TRAPPIST-1 d (*28*, *29*), have lower densities and longer orbital periods and are exposed to lower stellar irradiation, so they may be less susceptible to loss of an atmosphere (*14*).

We used a neural network to investigate possible interior structures of GJ 367b (*9*). At the best-fitting density, the neural network indicates that GJ 367b is predominantly made of iron (Fig. 3B), composed of 86 ± 5% iron core (by radius), <10% water ice and/or a H and He gas envelope, and the remainder as silicate mantle. This composition is similar to that of Mercury, which the neural network predicts would have an iron core radius fraction of 81 ± 4% (*9*). This is consistent with the measured Mercury core radius of 2030 ± 37 km (*30*), which corresponds to a core radius fraction of 83 ± 2%. For comparison, the interior structures of Mercury and other terrestrial planets predicted by our analysis are shown in fig. S8.

REFERENCES AND NOTES

1. C. Cifuentes *et al.*, *Astron. Astrophys.* **642**, A115 (2020).
2. E. E. Salpeter, *Astrophys. J.* **121**, 161 (1955).
3. T. J. Henry *et al.*, *Astron. J.* **132**, 2360–2371 (2006).
4. G. Chabrier, *Publ. Astron. Soc. Pac.* **115**, 763–795 (2003).
5. C. D. Dressing, D. Charbonneau, *Astrophys. J.* **807**, 45 (2015).
6. A. Suárez Mascareño *et al.*, *Astron. Astrophys.* **639**, A77 (2020).
7. S. H. Saar, R. A. Donahue, *Astrophys. J.* **485**, 319–327 (1997).
8. A. G. A. Brown *et al.*, *Astron. Astrophys.* **616**, A1 (2018).
9. Materials and methods are available as supplementary materials.
10. G. R. Ricker *et al.*, in *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*, H. A. MacEwen *et al.*, Eds., vol. 9904 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Proceedings* (2016), pp. 99042B.
11. J. M. Jenkins *et al.*, in *Software and Cyberinfrastructure for Astronomy IV*, G. Chiozzi, J. C. Guzman, Eds., vol. 9913 of *SPIE Proceedings* (2016), pp. 99133E.
12. J. Cabrera, S. Csizmadia, A. Erikson, H. Rauer, S. Kirste, *Astron. Astrophys.* **548**, A44 (2012).
13. S. Csizmadia, *Mon. Not. R. Astron. Soc.* **496**, 4442–4467 (2020).
14. J. E. Owen, Y. Wu, *Astrophys. J.* **847**, 29 (2017).
15. E. D. Lopez, J. J. Fortney, *Astrophys. J.* **792**, 1 (2014).
16. J. E. Owen, Y. Wu, *Astrophys. J.* **775**, 105 (2013).
17. A. Léger *et al.*, *Icarus* **213**, 1–11 (2011).
18. D. Valencia, D. D. Sasselov, R. J. O'Connell, *Astrophys. J.* **665**, 1413–1420 (2007).
19. L. Zeng, D. Sasselov, *Publ. Astron. Soc. Pac.* **125**, 227–239 (2013).
20. S. Seager, M. Kuchner, C. A. Hier-Majumder, B. Militzer, *Astrophys. J.* **669**, 1279–1297 (2007).
21. L. Zeng, D. D. Sasselov, S. B. Jacobsen, *Astrophys. J.* **819**, 127 (2016).
22. L. Zeng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9723–9728 (2019).
23. A. Santerne *et al.*, *Nat. Astron.* **2**, 393–400 (2018).
24. O. Barragán *et al.*, *Astron. Astrophys.* **612**, A95 (2018).
25. L. Malavolta *et al.*, *Astron. J.* **155**, 107 (2018).
26. D. Jontof-Hutter, J. F. Rowe, J. J. Lissauer, D. C. Fabrycky, E. B. Ford, *Nature* **522**, 321–323 (2015).
27. J. M. Almenara, R. F. Díaz, C. Dorn, X. Bonfils, S. Udry, *Mon. Not. R. Astron. Soc.* **478**, 460–486 (2018).
28. M. Gillon *et al.*, *Nature* **533**, 221–224 (2016).
29. M. Gillon *et al.*, *Nature* **542**, 456–460 (2017).
30. D. E. Smith *et al.*, *Science* **336**, 214–217 (2012).
31. R. A. Marcus, D. Sasselov, L. Hernquist, S. T. Stewart, *Astrophys. J. Lett.* **712**, L73–L76 (2010).
32. P. Baumeister *et al.*, *Astrophys. J.* **889**, 42 (2020).

ACKNOWLEDGMENTS

We acknowledge use of observations from the LCOGT network. We made use of data from the European Space Agency (ESA) mission Gaia (www.cosmos.esa.int/gaia), processed by the Gaia Data

Processing and Analysis Consortium (DPAC) (www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement. We acknowledge the use of public TESS Alert data from pipelines at the TESS Science Office and at the TESS Science Processing Operations Center. This work was supported by the KESPRINT collaboration, an international consortium devoted to the characterization and research of exoplanets discovered with space-based missions. Some of the observations were made at Gemini South using the high-resolution imaging instrument Zorro, funded by the NASA Exoplanet Exploration Program and built at the NASA Ames Research Center by S. B. Howell, N. Scott, E. P. Horch, and E. Quigley. Zorro was mounted on the Gemini South telescope of the international Gemini Observatory, a program of NSF's OIR Lab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation on behalf of the Gemini partnership: the National Science Foundation (United States); the National Research Council (Canada); Agencia Nacional de Investigacin y Desarrollo (Chile); Ministerio de Ciencia, Tecnologa e Innovacin (Argentina); Ministrio da Cifncia, Tecnologia, Inovaes e Comunicaes (Brazil); and the Korea Astronomy and Space Science Institute (Republic of Korea). This is University of Texas Center for Planetary Systems Habitability Contribution no. 0041. **Funding:** K.W.F.L., Sz.Cs., M.E., S.G., A.P.H., and H.R. were supported by Deutsche Forschungsgemeinschaft grants PA525/18-1, PA525/19-1, PA525/20-1, HA3279/12-1, and RA714/14-1 within the DFG Schwerpunkt SPP 1992, Exploring the Diversity of Extrasolar Planets. Sz.Cs. is supported by Deutsche Forschungsgemeinschaft Research Unit 2440, Matter Under Planetary Interior Conditions: High Pressure Planetary and Plasma Physics. T.H. was supported by JSPS KAKENHI grant no. JP19K14783, and N.N. was supported by JSPS KAKENHI grant nos. JP18H01265 and JP18H05439 and JST PRESTO grant no. JPMJPR1775. J.L. is supported by JSPS KAKENHI grant no. JP20K14518. R.A.G. acknowledges support from PLATO and GOLF CNES grants. P.K. acknowledges support from the MSMT grant LTT20015. S.M. acknowledges support by the Spanish Ministry of Science and Innovation with the Ramon y Cajal fellowship no. RYC-2015-17697 and grant no. PID2019-107187GB-I00. N.C.S. was supported by Fundaçao para a Ciência e a Tecnologia through national funds and by FEDER through COMPETE2020 - Programa Operacional Competitividade e Internacionalização grants UID/FIS/04434/2019, UIDB/04434/2020, UIDP/04434/2020, PTDC/FIS-AST/32113/2017 and POCI-01-0145-FEDER-032113, and PTDC/FIS-AST/28953/2017 and POCI-01-0145-FEDER-028953. X.D. and G.G. acknowledge funding in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02), Origin of Life project of the Université Grenoble-Alpes. J.R.D.M. acknowledges grants from CNPq, CAPES, and FAPERN Brazilian agencies. N.A.-D. acknowledges support from FONDECYT project 3180063. Resources for the production of the SPOC data products were provided by the NASA High-End Computing (HEC) program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. S.A. is supported by the Danish National Research Foundation (grant agreement no. DNRF106). D.G. and L.M.S. acknowledge financial support from the Cassa di Risparmio di Torino foundation under grant no. 2018.2323, "Gaseous or rocky? Unveiling the nature of small worlds." **Author contributions:** K.W.F.L. contributed to the planet detection, transit light curve analysis, gyrochronology age estimate, and tidal evolution calculations and led the writing of the paper. Sz.Cs. performed the joint light curve and RV analysis using the Transit and Light Curve Modeller (TLCM). N.A.-D. and X.B. led the HARPS RV follow-up program and reduced and analyzed the RV data. D.G. performed the frequency analysis of the RV and activity indicators. F.D., O.B., A.P.H., and R.L. analyzed the RV data. S.P. determined the interior composition of the planet and modeled the precursor gaseous planet. M.E. performed the Rapid Eye Mount (REM) robotic telescope observations and analysis. M.E. and A.M.S.S. derived the light curve dilution factor. C.H., K.W.F.L., S.M., and R.A.G. determined the stellar rotation period using WASP data. S.B.H. analyzed the speckle imaging data. T.H. and M.F. performed the spectral analysis using SpecMatch-Emp, and T.H. derived the stellar parameters using an MCMC simulation. J.L. performed stellar characterization using isochrone fitting. F.M. performed the LCOGT and HARPS observations. K.A.C. coordinated the TESS SG1 working group. K.A.C. and P.L. analyzed the ground-based photometric observations. N.C.S. performed spectral characterization using ODUSSEAS. M.C.J. determined the Galactic space velocities of the star. K.W.F.L. and S.R. calculated the emission spectroscopy metric. J.K. performed transit-timing variation (TTV) analysis. J.C., P.E., and S.G. analyzed the light curve for planet detection. E.Gu. analyzed the activity indicators. G.R.R., R.V., D.W.L., S.S., J.N.W., and J.M.J. led and organized the TESS mission, including the observations, processing of the data, working groups coordination, target selection, and dissemination of the data products. E.H.M., M.V., and J.E.S. are members of the TESS POC who coordinated and scheduled the TESS science observations and conducted instrument planning. R.M., J.C.S., and J.D.T. are members of the SPOC who performed data calibration, light curve production, and transit planet detection. D.C., J.C., and S.N.Q. are members of the TSO who reviewed the data and performed planet candidate vetting. J.M.A., E.A., F.B., D.C., J.R.D.M., X.D., R.F.D., R.D., P.F., T.F., G.G., C.L., C.M., F.P., D.S., and S.U. are coinvestigators of the proposal, which provided the HARPS observations of the target and supported the interpretation of results. S.A., P.C., A.C., W.D.C., E.Go., I.G., P.K., E.K., J.J.L., N.N., H.L.M.O., E.P., C.M.P., H.R., L.M.S., J.Š., and V.V.E. are part of the KESPRINT consortium and contributed to the interpretation of the results. All authors contributed to the preparation of the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The TESS photometric observations are available at the Mikulski Archive for Space Telescopes (MAST) at https://exo.mast.stsci.edu under target name TOI 731.01. The raw HARPS spectroscopic data are available on the ESO Science Archive Facility http://archive.eso.org/cms.html under ESO program IDs 072.C-0488, 082.C-0718, 183.C-0437 (primary investigators: M. Mayor and X. Bonfils), and 1102.C-0339 (primary investigator: X. Bonfils). The ground-based photometry obtained by the LCO telescope and REM as well as the Gemini imaging data are available on the Exoplanet Follow-up Observing Program (ExoFOP) website https://exofop.ipac.caltech.edu/tess/ under target name TOI 731.01. The raw Gemini data are available at https://archive.gemini.edu/searchform under program ID GS-2021A-LP-105. The archival WASP data are available on the NASA Exoplanet Archive https://exoplanetarchive.ipac.caltech.edu/docs/SuperWASPMission.html under object name GJ 367. Our reduced RVs and activity indices are listed in tables S1 and S2 and in machine-readable form in data S1 and S2. The TLCM is available at www.transits.hu/.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aay3253
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S3
References (*33*–*127*)
Data S1 and S2

31 December 2020; accepted 12 October 2021
10.1126/science.aay3253

CARBON CYCLE

Strong Southern Ocean carbon uptake evident in airborne observations

Matthew C. Long[1]*, Britton B. Stephens[1], Kathryn McKain[2,3], Colm Sweeney[3], Ralph F. Keeling[4], Eric A. Kort[5], Eric J. Morgan[4], Jonathan D. Bent[1,4]†, Naveen Chandra[6]‡, Frederic Chevallier[7], Róisín Commane[8], Bruce C. Daube[9], Paul B. Krummel[10], Zoë Loh[10], Ingrid T. Luijkx[11], David Munro[2,3], Prabir Patra[12], Wouter Peters[11,13], Michel Ramonet[7], Christian Rödenbeck[14], Ann Stavert[10], Pieter Tans[3], Steven C. Wofsy[9,15]

The Southern Ocean plays an important role in determining atmospheric carbon dioxide (CO_2), yet estimates of air-sea CO_2 flux for the region diverge widely. In this study, we constrained Southern Ocean air-sea CO_2 exchange by relating fluxes to horizontal and vertical CO_2 gradients in atmospheric transport models and applying atmospheric observations of these gradients to estimate fluxes. Aircraft-based measurements of the vertical atmospheric CO_2 gradient provide robust flux constraints. We found an annual mean flux of −0.53 ± 0.23 petagrams of carbon per year (net uptake) south of 45°S during the period 2009–2018. This is consistent with the mean of atmospheric inversion estimates and surface-ocean partial pressure of CO_2 (pCO_2)–based products, but our data indicate stronger annual mean uptake than suggested by recent interpretations of profiling float observations.

Ocean water-column carbon inventories suggest that the Southern Ocean accounts for more than 40% of the cumulative global ocean uptake of anthropogenic CO_2 (*1*, *2*). However, estimates of contemporary net Southern Ocean air-sea carbon fluxes based on surface-ocean partial pressure of CO_2 (pCO_2) observations or atmospheric inversions remain highly uncertain (*3*–*8*). Recent interpretations of profiling float observations have introduced further complications, proposing a profound revision of the Southern Ocean carbon budget, with reduced summertime uptake and larger wintertime outgassing (*9*, *10*). Given the Southern Ocean's critical role as a sink for anthropogenic CO_2, as well as

[1]National Center for Atmospheric Research, Boulder, CO, USA. [2]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA. [3]Global Monitoring Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA. [4]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA. [5]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA. [6]National Institute of Environmental Studies, Tsukuba, Japan. [7]Laboratoire des Sciences du Climat et de l'Environnement, IPSL-LSCE, CEA-CNRS-UVSQ, UMR8212 91191, France. [8]Department of Earth and Environmental Sciences, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY, USA. [9]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA. [10]Climate Science Centre, CSIRO Oceans and Atmosphere, Aspendale, Victoria, Australia. [11]Department of Meteorology and Air Quality, Environmental Sciences Group, Wageningen University, Netherlands. [12]Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Japan. [13]Centre for Isotope Research, University of Groningen, Groningen, Netherlands. [14]Max Planck Institute for Biogeochemistry, Jena, Germany. [15]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA.
***Corresponding author. Email: mclong@ucar.edu**
†Present address: Picarro, Inc., Santa Clara, CA, USA.
‡Present address: Research Institute for Global Change, JAMSTEC, Yokohama, Japan.

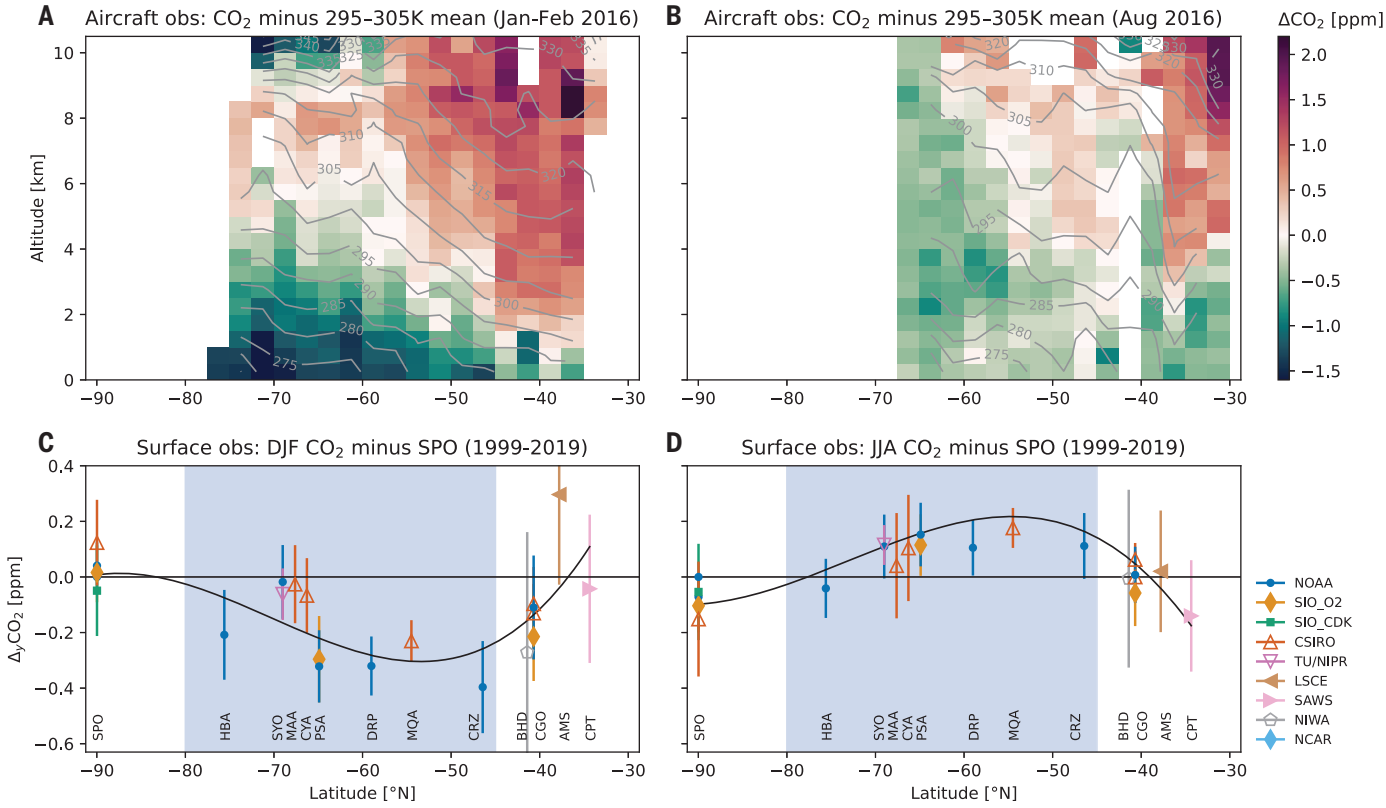


Fig. 1. Observed patterns in atmospheric CO_2 over the Southern Ocean. (**A** and **B**) Cross sections observed by aircraft during (A) ORCAS, in January to February 2016, and (B) ATom-1, in August 2016. Colors show the observed CO_2 dry air mole fraction relative to the average observed within the 295–305 K potential temperature range south of 45°S on each campaign. Contour lines show the observed potential temperature. See figs. S1 and S2 for flight tracks and cross-sectional plots for all campaigns, and figs. S3 and S4 for simulated fields. (**C** and **D**) Compilation of mean CO_2 observed at surface monitoring stations minus the National Oceanic and Atmospheric Administration (NOAA) in situ record at the South Pole Observatory (SPO) during the period 1999–2019 for (C) summer (DJF) and (D) winter (JJA). The black line is a spline fit provided simply as a visual guide. Blue shading denotes the latitude band in which we designate "Southern Ocean stations." See table S1 and fig. S5 for station locations and temporal coverage. SM includes additional methodological details.

indications that regional fluxes vary substantially on decadal time scales (*7*, *11*, *12*), it is essential to develop more-robust constraints on Southern Ocean air-sea CO_2 exchange. Observations of atmospheric CO_2 provide an opportunity for doing so, as the atmosphere effectively integrates flux signals over large surface regions. Atmospheric inversion models provide a formal statistical method to estimate fluxes that optimally satisfy atmospheric observational constraints, given circulation simulated by data-constrained atmospheric transport models (*4*, *13*, *14*). However, global-scale atmospheric inversion models have not converged on consistent Southern Ocean fluxes, as they suffer from inaccuracies in the simulated transport, reliance on uncertain "prior" flux estimates, and requirements to meet tighter constraints elsewhere in the world, where signals are stronger and measurements less sparse (*4*, *13*–*17*).

In this study, we derived "emergent constraints" on regional air-sea fluxes by relating fluxes in a collection of models to observable gradients in CO_2 in the atmosphere directly overlying the Southern Ocean. We used observations from nine deployments of three recent aircraft projects: the HIAPER Pole-to-Pole Observations (HIPPO) project (*18*), the O_2/N_2 Ratio and CO_2 Airborne Southern Ocean (ORCAS) study (*19*), and the Atmospheric Tomography (ATom) mission (*20*) (see supplementary materials, hereafter SM). We also examined 44 atmospheric CO_2 records from surface monitoring stations in the high-latitude Southern Hemisphere, selecting and filtering the highest-quality data (SM). Collectively, these observations show a distinct pattern in the seasonal variability of atmospheric CO_2 overlying the Southern Ocean, most notably characterized by lower-troposphere CO_2 depletion in austral summer and neutral to weakly positive enhancement in austral winter (Figs. 1 and 2, A to C). To isolate CO_2 gradients driven by Southern Ocean fluxes, we examined CO_2 anomalies relative to a local reference, using potential temperature (θ) to delineate boundaries in the vertical dimension (SM). We defined a metric quantifying the vertical CO_2 gradient, $\Delta_\theta CO_2$, as the difference between the median value of CO_2 observed south of 45°S, where $\theta < 280$ K, and that in the mid- to upper-troposphere, where $295\ K < \theta < 305$ K. The aircraft observations suggest that the amplitude of seasonal variation in CO_2 is minimized within this upper θ range relative to the rest of the column (fig. S7); it is also above the vertical extent of wintertime, near-surface homogeneity (Fig. 2A) and below altitudes substantially influenced by the stratosphere, making it a good reference for detecting regional air-sea flux signals (see SM). Similarly, we defined a metric of the horizontal surface gradient, $\Delta_y CO_2$, as the difference between CO_2 averaged across stations in the core latitudes of summertime CO_2 drawdown (Fig. 1, C and D, shaded region) and that at the South Pole Observatory (SPO). $\Delta_\theta CO_2$ is strongly negative in the austral summer, followed by near-neutral conditions in the austral winter through spring (Fig. 2B). Correspondingly, $\Delta_y CO_2$ also indicates summertime drawdown at the surface and weakly positive to near-neutral conditions in winter (Fig. 2C), although the amplitude of seasonal variation in $\Delta_\theta CO_2$ is more than

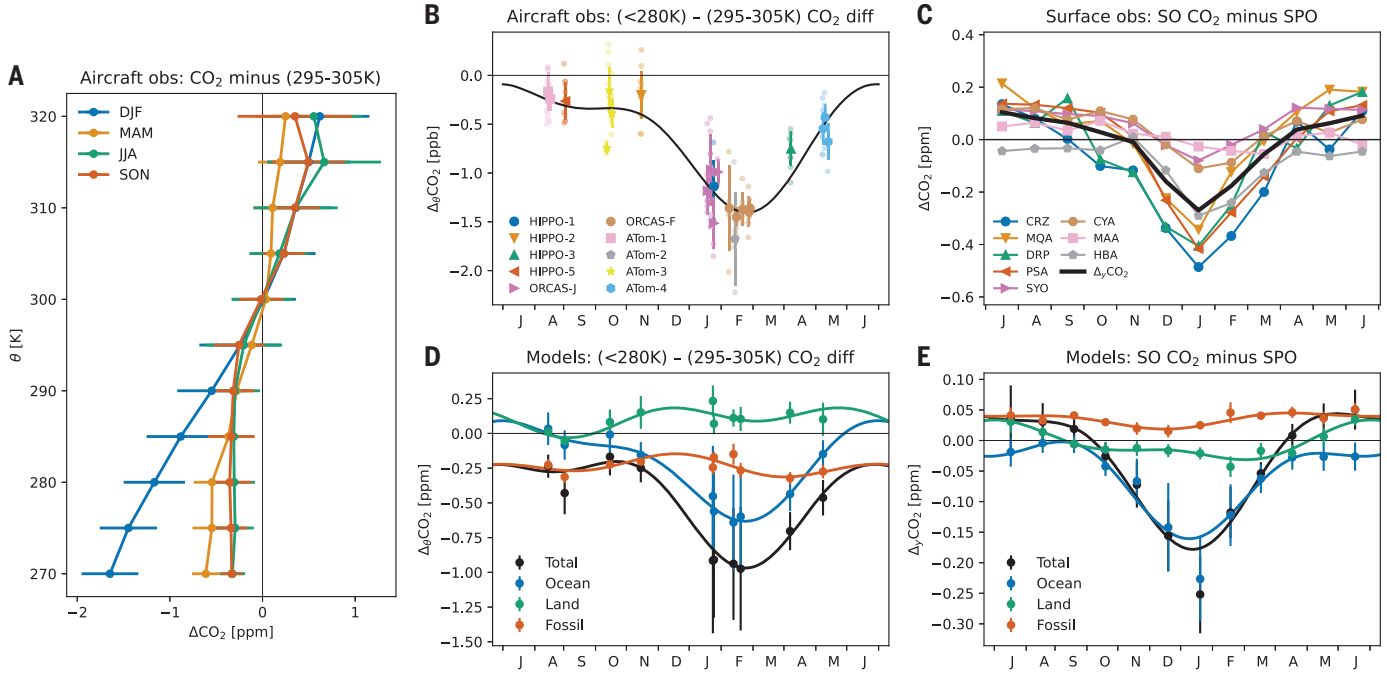


Fig. 2. Seasonal evolution of atmospheric CO_2 over the Southern Ocean. (**A**) Vertical profiles of CO_2 observations collected by aircraft south of 45°S, binned on 5 K potential temperature (θ) bins and averaged by season (whiskers show standard deviation; fig. S6 shows model comparison). MAM, austral fall (March to May); SON, austral spring (September to November). (**B**) The vertical gradient ($\Delta_\theta CO_2$) in CO_2 measured from aircraft south of 45°S. Small points show $\Delta_\theta CO_2$ for individual profiles; larger points show the median and standard deviation (whiskers) for each flight. The black line shows a two-harmonic fit to the flight-median points. (**C**) Monthly climatology (1999–2019) of the latitudinal gradient in CO_2 measured by surface stations (Fig. 1); the black line shows the station mean metric ($\Delta_y CO_2$). Separate laboratory records at Syowa Station (SYO) and Palmer Station (PSA) have been averaged. The seasonal evolution of (**D**) $\Delta_\theta CO_2$ and (**E**) $\Delta_y CO_2$ simulated in a collection of atmospheric inversion models (table S3). The points show the median across the models, and whiskers show the standard deviation. The colors correspond to the total CO_2 (black) and CO_2 tracers responsive to only ocean (blue), land (green), and fossil (red) surface fluxes. Note that the *y* axis bounds differ by panel.

three times larger than that in $\Delta_y CO_2$. Variation in drawdown intensity across stations contributing to $\Delta_y CO_2$ likely reflects differential sampling of air exposed to strong ocean productivity signals (fig. S4).

We developed inferences about air-sea CO_2 fluxes from these gradient metrics by examining a collection of atmospheric inverse models that simulate time-varying, three-dimensional CO_2 fields sampled to replicate observations (SM). The inverse models demonstrate that seasonality in $\Delta_\theta CO_2$ and $\Delta_y CO_2$ is dominated by Southern Ocean air-sea fluxes. Although land and fossil fuel fluxes are small south of 45°S, extraregional contributions do influence local gradients via transport from the north. The models explicitly simulate CO_2 tracers responsive only to ocean (CO_2^{ocn}), land (CO_2^{lnd}), and fossil fuel (CO_2^{ff}) fluxes and subject to identical transport fields. The simulations of these tracers indicate that the influence of land fluxes generally opposes the effect of fossil fuel emissions for both gradient metrics, and the seasonality in the land and fossil fuel tracers is much weaker than the ocean-derived signal (Fig. 2, D and E, and fig. S6). The negative vertical (positive horizontal) gradient in fossil fuel CO_2 is consistent with elevated CO_2 concentrations in the equatorward portion of the domain, particularly at high altitude (Fig. 1, A and B, and figs. S2 to S4). Ancillary measurements of methane-mixing ratios confirm that this feature reflects long-range transport of emission signals from land, but that it has little influence on $\Delta_\theta CO_2$ (figs. S6 and S8). Additional evidence that fossil fuel emissions make only small contributions to the annual mean and seasonality in $\Delta_y CO_2$ comes from ancillary observations of sulfur hexafluoride—which provides an analog for fossil fuel CO_2 (*21*) and shows very little spatial or temporal structure over the Southern Ocean (fig. S9).

To develop quantitative flux estimates, we related simulated $\Delta_\theta CO_2^{ocn}$ and $\Delta_y CO_2^{ocn}$ to regionally integrated, temporally averaged air-sea flux in each modeling system (Fig. 3). In addition to inverse models, we included forward atmospheric transport integrations forced with spatially explicit surface-ocean pCO_2-based flux datasets (SM). Ultimately, each model realization was a forward simulation producing three-dimensional CO_2 fields from which we computed gradient metrics consistent with the model's surface fluxes and atmospheric transport. The relationship between the fluxes and simulated gradient metrics across the collection of models enabled using the observed gradients to constrain Southern Ocean fluxes. We assumed that the relevant surface-influence region can be approximated as the area south of a particular latitude and focused on fluxes integrated over the region south of 45°S, noting that the flux products indicate strong meridional gradients and seasonality in the zonal mean fluxes south of 30°S (fig. S18, A and B). We averaged the regional fluxes over individual seasons to regress against the surface mole fraction observations and over 90 days before each aircraft campaign (see SM for sensitivity tests, including an assessment of different region boundaries and a similar analysis based on gradients in total CO_2). There is a robust relationship between the CO_2 flux south of 45°S and $\Delta_\theta CO_2^{ocn}$ across the models (Fig. 3, A and B). The sensitivity of $\Delta_\theta CO_2^{ocn}$ to fluxes varies seasonally, as indicated by a change in slope between seasons. December to February (DJF) is distinct in having a smaller slope (higher sensitivity); the other seasons individually have larger slopes that are similar to each other, thus we grouped data from

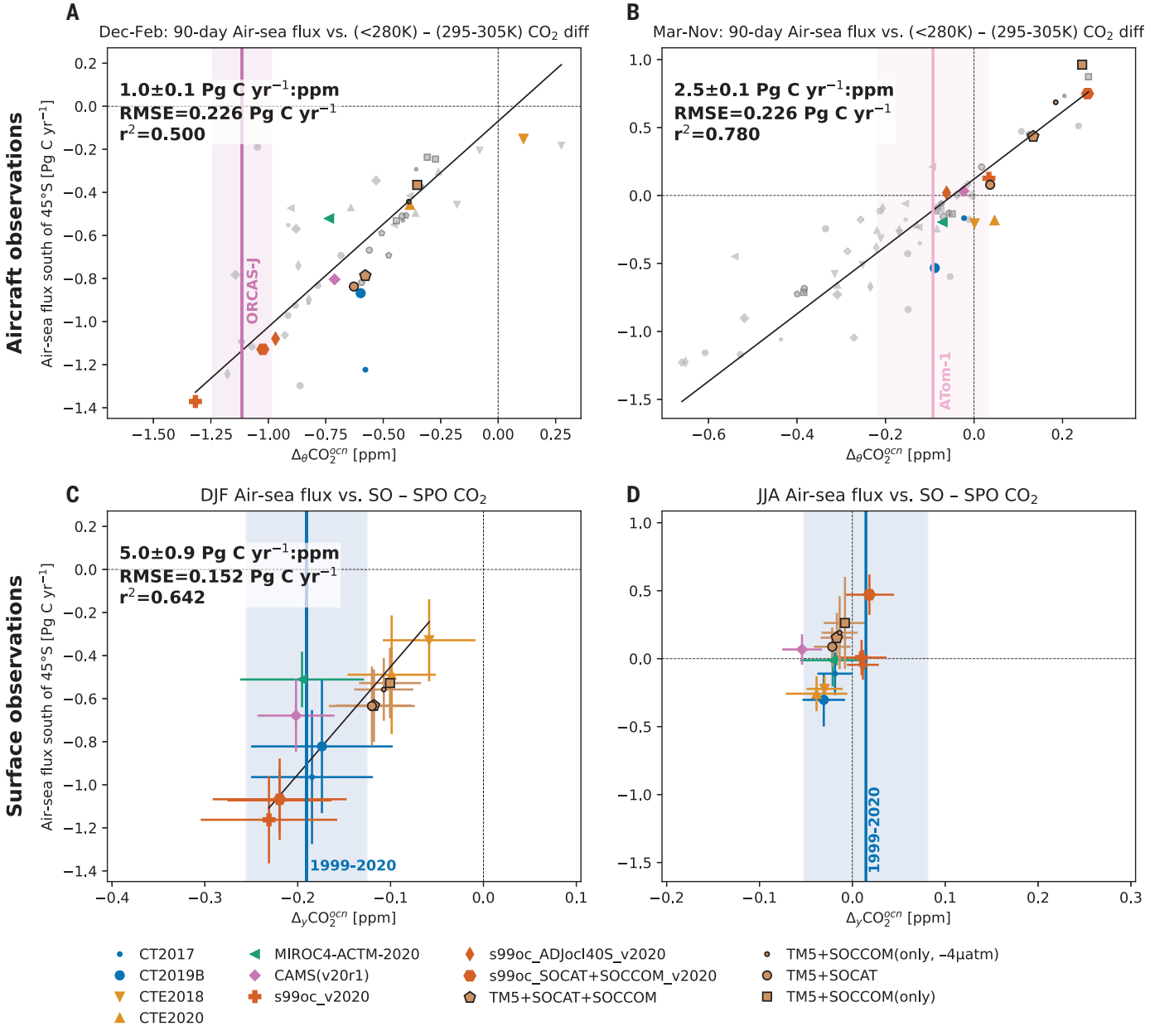


Fig. 3. Emergent constraints on air-sea fluxes south of 45°S. (**A** and **B**) 90 day–mean air-sea fluxes south of 45°S versus $\Delta_\theta CO_2^{ocn}$ from model simulations (see SM) replicating aircraft observations collected during (A) December to February and (B) March to November. Colored vertical lines show an observed value of $\Delta_\theta CO_2$ [ORCAS during January in (A) and ATom-1 in (B)] corrected for land and fossil fuel influence, with shading indicating both analytical uncertainty and model spread in the correction (see SM); colored points highlight the model samples from these particular campaigns, while gray points show data from other campaigns in the (A) December to February or (B) March to November timeframe. Figures S10 and S11 show similar plots for each individual aircraft campaign. (**C** and **D**) Seasonal-mean surface fluxes versus $\Delta_y CO_2^{ocn}$ computed from models for (C) summer (DJF) and (D) winter (JJA) over the period 1999–2019. Points correspond to individual models; whiskers denote the standard deviation of interannual variability. Light blue vertical lines show the observed $\Delta_y CO_2$ corrected for land and fossil fuel influence; shading shows analytical uncertainty and model spread in the correction (see SM; fig. S12, A and B, shows $\Delta_y CO_2$ time series). The sign convention for fluxes is positive upward. Diagonal lines, where significant, show the best-fit line to all data points shown; inset text shows an estimate of the slope with standard error (SM), and goodness-of-fit statistics are also shown. Table S3 provides detailed information on the model products, defining the acronyms used in the legend. Note that the axis bounds differ by panel. See fig. S16 for a version of this plot based on total CO_2.

campaigns flown in March to November together (Fig. 3B). For the surface data, we find a significant positive relationship between the regional air-sea flux and $\Delta_y CO_2$ in DJF across the models (Fig. 3C); the flux-$\Delta_y CO_2$ relationship dwindles in strength during nonsummer months, however, and there is no significant relationship in austral winter [June to August (JJA)] (Fig. 3D). The spread enabling these relationships results from the diversity of flux estimates, while the scatter about the fits represents both different realizations of atmospheric transport and spatiotemporal mismatch between the true surface influence function and our coarse spatiotemporal approximation. The smaller slope for the aircraft data in DJF is consistent with greater atmospheric stability (reduced vertical mixing) over the cold ocean

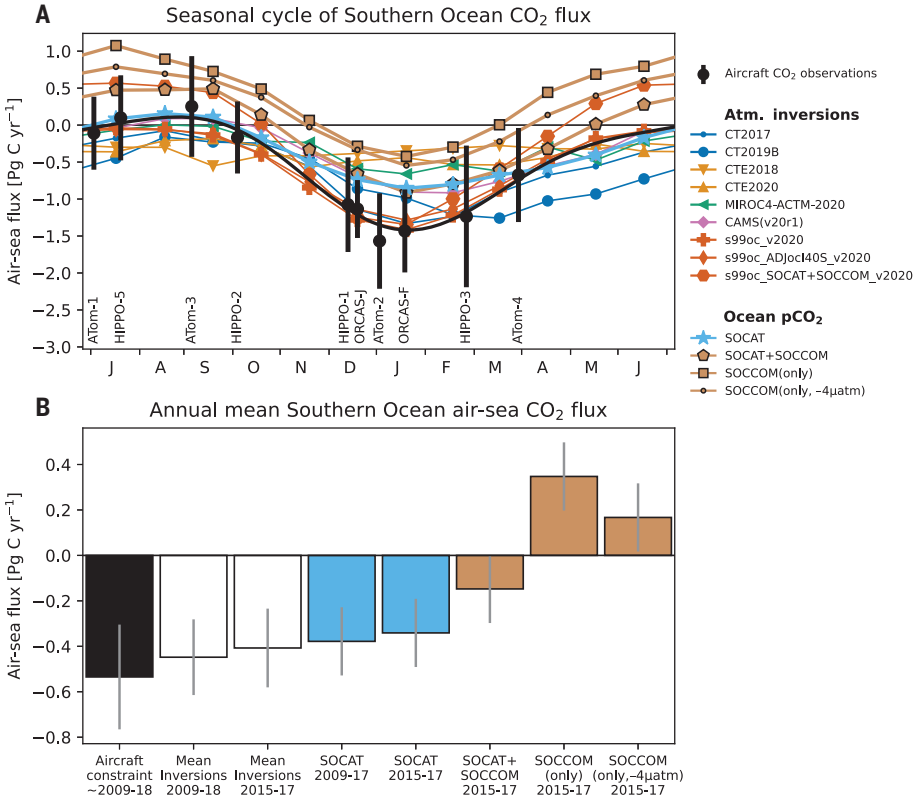


Fig. 4. Observationally based estimates of Southern Ocean air-sea fluxes. (**A**) The seasonal cycle of air-sea CO_2 flux south of 45°S estimated from aircraft campaigns (black points, labels), plotted at the center of the 90-day window for which the emergent flux constraint was calibrated. Whiskers show the standard deviation derived from propagating analytical and statistical uncertainties; the black line shows a two-harmonic fit used to estimate the annual mean flux. The colored lines give the seasonal cycle from atmospheric inversion systems as well as the neural network extrapolation (*22*) of the Surface Ocean CO_2 Atlas (SOCAT) Pco_2 observations (*31*) and profiling float observations from the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) project (*32*). Fluxes are averaged over the period 2009–2018, except for the three neural network–based flux estimates (*27*) incorporating SOCCOM observations, which are averaged over the period 2015–2017. (**B**) Annual mean flux estimated in this study (leftmost bar) including uncertainty (whisker), along with the mean and standard deviation (whiskers) across the inversion systems shown in (A) as well as the surface-ocean Pco_2-based methods; averaging time periods are noted in the axis labels (both SOCAT flux estimates were derived using neural network training over the full observational period). The uncertainty estimate on the SOCAT and SOCCOM fluxes is approximated from (*10*), which reported ±0.15 Pg C year^{-1} as the "method uncertainty" associated with the neural network–based flux estimates for the whole Southern Ocean (south of 44°S). Note that while s99oc_v2020 and s99c_SOCAT+SOCCOM_v2020 are global inversions, their ocean fluxes are prescribed, not optimized using atmospheric observations (see SM); similarly, the CAMS(v20r1) ocean fluxes remain close to its SOCAT Pco_2-based prior.

during austral summer, intensifying the flux signal in the lower troposphere; more-energetic vertical mixing in other seasons, as well as stronger horizontal flow, results in diminished sensitivity in $\Delta_\theta CO_2$ and no clear relationship between fluxes and the surface station–based $\Delta_y CO_2$ metric in winter.

Vertical lines in Fig. 3 show representative observations of each gradient metric corrected for land and fossil fuel contributions; the intersection of these lines with the flux-gradient fit provides a quantitative flux estimate. Applying this emergent constraint for each aircraft campaign yields 10 flux estimates spread over 7 months of the year; these data suggest that the Southern Ocean is a strong sink for CO_2 in austral summer, with fluxes that are near-neutral during winter (Fig. 4A). Applying a two-harmonic fit to the data, we estimated an annual mean flux spread over the aircraft observing period (2009–2018) of −0.53 ± 0.23 petagrams of carbon (Pg C) year^{-1} (Fig. 4B). The seasonal cycle of fluxes estimated from aircraft campaigns largely agrees with flux estimates derived from the Surface Ocean CO_2 Atlas (SOCAT) Pco_2 data product, using either neural network interpolation (*22*) or the Jena mixed-layer scheme (*23*) (Fig. 4A). Similarly, the aircraft-based fluxes agree with the multimodel mean of inverse estimates, although inversions tend to underestimate summer uptake (fig. S12C), overestimate winter uptake (fig. S12D), and show greater than 100% disagreement on the annual mean flux. We have not explicitly accounted for interannual variability or trends in the fluxes over the period of aircraft data collection (fig. S12, C and D), although we expect this to be a relatively small effect, as seasonal coverage between HIPPO and ATom is relatively uniform (Fig. 4A). The flux estimates obtained from the surface atmospheric CO_2 gradient in summer are consistent with the aircraft-based estimates (fig. S12C) but have larger uncertainty—indeed, the magnitude of the $\Delta_y CO_2$ signal is small relative to analytical uncertainty (SM), a particular challenge in this region, where sites are remote, conditions are harsh, and intercomparison between the multiple laboratories maintaining CO_2 records is limited (*24*, *25*). Despite the large uncertainty, however, trends in $\Delta_y CO_2$ are consistent with increasing Southern Ocean uptake since 2005 (*7*, *26*) (see SM); for instance, $\Delta_y CO_2$ declined by about 0.16 ± 0.03 parts per million (ppm) decade^{-1} over the period 2005–2019 for both DJF and JJA, and while there was no significant flux-gradient relationship in JJA (Fig. 3D), the associated $\Delta_y CO_2$ -based flux estimates suggest the DJF flux was −0.5 ± 0.7 Pg C year^{-1} from 2005 to 2009, −1.1 ± 0.9 Pg C year^{-1} from 2010 to 2014, and −1.3 ± 1.1 Pg C year^{-1} from 2015 to 2019 (fig. S12C). Notably, the aircraft-based flux estimates indicate stronger annual mean uptake than fluxes incorporating Pco_2 estimates from the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) profiling float pH measurements (*10*, *27*). The primary SOCCOM flux product we examined (SOCAT+SOCCOM) is derived from neural network interpolation including both ship-based surface-ocean Pco_2 observations as well as float-derived Pco_2 estimates [see SM and (*27*)]; this product yields weaker annual mean uptake but tracks the individual aircraft campaign flux estimates within uncertainty (Fig. 4A). The other two SOCCOM-based products presented here are sensitivity runs (*10*, *27*) that selectively exclude ship-based Pco_2 observations in the Southern Ocean (see SM). While these products have a seasonal phase and amplitude similar to those of the aircraft flux estimates, they indicate greater outgassing in winter and less ingassing during summer than the aircraft-based fluxes (Fig. 4). Such large fluxes should have clear atmospheric signatures (Fig. 3, A and B), but no such signatures are evident in any of the Southern Ocean aircraft campaigns (Fig. 2, A and B, and figs. S2, S10, and S11).

Our analysis demonstrates that Southern Ocean air-sea fluxes impart a coherent pattern in atmospheric CO_2 as measured by aircraft. The surface station network is only detectably sensitive to variations in fluxes during austral summer and hampered by measurement noise commensurate with flux signals. Our results highlight the difficulty global atmospheric inversions have in capturing meaningful estimates of Southern Ocean fluxes using existing surface data constraints. It is important to note that a robust emergent constraint (Fig. 3) requires a diverse collection of models to avoid results being affected by biases specific to a single model or unidentified transport biases common across models. The collection of models we included use four different underlying meteorological reanalysis datasets, four different transport models, and differ in spatial resolution and treatment of vertical transport (table S3); moreover, they make up a substantial proportion of the models commonly used for long-term global CO_2 inversions. However, inclusion of additional model solutions would improve confidence in our result by increasing the number of independent realizations of transport. Despite this potential limitation, aircraft observations leverage the broad integrative power of the atmosphere, which provides an advantage over estimating fluxes from surface ocean pCO_2 observations: The ocean surface is heterogeneous, making representative sampling difficult; air-sea fluxes computed from pCO_2 estimates depend on an uncertain gas exchange parameterization (*28*); and float-based estimates have additional uncertainty associated with estimating pCO_2 itself (*29*). However, we resolved fluxes only over a broadly defined Southern Ocean region; finer-scale spatial features present in surface-ocean pCO_2 data can provide important mechanistic insight, reinforcing the need for more high-quality, widely distributed ocean observations to advance process understanding. Uncertainty regarding Southern Ocean carbon uptake is a critical limitation in current understanding of the global carbon cycle (*30*). Our results can be used to validate Earth system models and inversion-based assessments of the Southern Hemisphere carbon budget. Critically, integral constraints on the atmospheric CO_2 budget require balanced fluxes; therefore, our result of strong Southern Ocean uptake alleviates the need to identify missing Southern Hemisphere land or subtropical ocean sinks, as suggested by the float observations. Finally, our analysis has important implications for effective monitoring of the Southern Ocean carbon sink. A regular program of aircraft observations could provide a cost-effective approach to drastically improve estimates of the carbon budget for the Southern Ocean and globally, helping to fulfill a societal requirement for clear understanding of mechanisms driving variation in atmospheric CO_2.

REFERENCES AND NOTES

1. S. Khatiwala, F. Primeau, T. Hall, *Nature* **462**, 346–349 (2009).
2. T. DeVries, *Global Biogeochem. Cycles* **28**, 631–647 (2014).
3. A. Lenton *et al*., *Biogeosciences* **10**, 4037–4054 (2013).
4. P. Peylin *et al*., *Biogeosciences* **10**, 6699–6720 (2013).
5. S. Crowell *et al*., *Atmos. Chem. Phys.* **19**, 9797–9831 (2019).
6. T. Takahashi *et al*., *Deep Sea Res. Part II Top. Stud. Oceanogr.* **56**, 554–577 (2009).
7. P. Landschützer *et al*., *Science* **349**, 1221–1224 (2015).
8. N. Gruber *et al*., *Global Biogeochem. Cycles* **23**, GB1005 (2009).
9. A. R. Gray *et al*., *Geophys. Res. Lett.* **45**, 9049–9057 (2018).
10. S. M. Bushinsky *et al*., *Global Biogeochem. Cycles* **33**, 1370–1388 (2019).
11. G. A. McKinley *et al*., *Nature* **530**, 469–472 (2016).
12. N. Gruber, P. Landschützer, N. S. Lovenduski, *Ann. Rev. Mar. Sci.* **11**, 159–186 (2019).
13. W. Peters *et al*., *J. Geophys. Res.* **110**, D24304 (2005).
14. I. G. Enting, in *Inverse Methods in Global Biogeochemical Cycles*, P. Kasibhatla*et al*., vol. 114 of *Geophysical Monograph Series* (American Geophysical Union, 2000), pp. 19–31.
15. A. S. Denning, I. Y. Fung, D. Randall, *Nature* **376**, 240–243 (1995).
16. A. E. Schuh *et al*., *Global Biogeochem. Cycles* **33**, 484–500 (2019).
17. S. Basu *et al*., *Atmos. Chem. Phys.* **18**, 7189–7215 (2018).
18. S. C. Wofsy, *Philos. Trans. R. Soc. A* **369**, 2073–2086 (2011).
19. B. B. Stephens *et al*., *Bull. Am. Meteorol. Soc.* **99**, 381–402 (2018).
20. S. C. Wofsy *et al.*, ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, version 2, ORNL Distributed Active Archive Center (2021); https://doi.org/10.3334/ORNLDAAC/1925.
21. L. S. Geller *et al*., *Geophys. Res. Lett.* **24**, 675–678 (1997).
22. P. Landschützer, N. Gruber, D. C. E. Bakker, *Global Biogeochem. Cycles* **30**, 1396–1417 (2016).
23. C. Rödenbeck *et al*., *Biogeosciences* **11**, 4599–4613 (2014).
24. K. A. Masarie *et al*., *J. Geophys. Res.* **106**, 20445–20464 (2001).
25. R. J. Francey, J. S. Frederiksen, L. P. Steele, R. L. Langenfelds, *Atmos. Chem. Phys.* **19**, 14741–14754 (2019).
26. D. R. Munro *et al*., *Geophys. Res. Lett.* **42**, 7623–7630 (2015).
27. P. Landschützer, S. M. Bushinsky, A. R. Gray, A combined globally mapped carbon dioxide (CO_2) flux estimate based on the Surface Ocean CO_2 Atlas Database (SOCAT) and Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) biogeochemistry floats from 1982 to 2017 (NCEI Accession 0191304), NOAA National Centers for Environmental Information (NCEI) (2019); https://doi.org/10.25921/9hsn-xq82.
28. R. Wanninkhof, *Limnol. Oceanogr. Methods* **12**, 351–362 (2014).
29. N. L. Williams *et al*., *Global Biogeochem. Cycles* **31**, 591–604 (2017).
30. J. G. Canadell *et al.*, in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte *et al*., (Cambridge Univ. Press, 2021), chap. 5.
31. D. C. E. Bakker *et al*., *Earth Syst. Sci. Data* **6**, 69–90 (2014).
32. K. S. Johnson *et al*., *J. Geophys. Res. Oceans* **122**, 6416–6436 (2017).
33. UCAR/NCAR - Earth Observing Laboratory, NSF/NCAR GV HIAPER Aircraft (2005); https://doi.org/10.5065/D6DR2SJP.
34. S. C. Wofsy *et al.*, HIPPO Merged 10-Second Meteorology, Atmospheric Chemistry, and Aerosol Data, version 1.0, UCAR/NCAR - Earth Observing Laboratory (2017); https://doi.org/10.3334/CDIAC/HIPPO_010.
35. S. Wofsy *et al.*, HIPPO NOAA Flask Sample GHG, Halocarbon, and Hydrocarbon Data, version 1.0, UCAR/NCAR - Earth Observing Laboratory (2017); https://doi.org/10.3334/CDIAC/HIPPO_013.
36. S. Wofsy *et al.*, HIPPO MEDUSA Flask Sample Trace Gas and Isotope Data, version 1.0, UCAR/NCAR - Earth Observing Laboratory (2017); https://doi.org/10.3334/CDIAC/HIPPO_014.
37. B. Stephens, ORCAS Merge Products, version 1.0, UCAR/NCAR - Earth Observing Laboratory (2017); https://doi.org/10.5065/D6SB445X.
38. M. C. Long *et al.*, Southern Ocean Air-Sea Carbon Fluxes from Aircraft Observations: Modeling Datasets, version 1.0, UCAR/NCAR - DASH Repository (2021); https://doi.org/10.5065/fepv-0z52.
39. M. C. Long, B. B. Stephens, Southern Ocean Air-Sea Carbon Fluxes from Aircraft Observations: Analysis code, National Center for Atmospheric Research (2021);https:/doi.org/10.5065/6vnv-1x08.

ACKNOWLEDGMENTS

Many people contributed to the collection of data used in this study. We thank R. Jimenez, J. Pittman, S. Park, B. Xiang, G. Santoni, M. Smith, and J. Budney for HIPPO and ATom Harvard QCLS and OMS CO_2 data; T. Newberger, F. Moore, and G. Diskin for ORCAS and ATom NOAA Picarro and PFP CO_2 data; A. Watt, S. Shertz, B. Paplawsky, and S. Afshar for HIPPO, ORCAS, and ATom NCAR AO2 and NCAR/Scripps Medusa CO_2 data; J. Elkins, F. Moore, and E. Hintsa for ATom-1 N_2O data; R.-S. Gao and R. Spackman for HIPPO O_3 data; T. Ryerson, J. Peischl, I. Bourgeois, and C. Thompson for ATom O_3 data; M. Zondlo, M. Diao, and S. Beaton for HIPPO and ORCAS H_2O data; and G. Diskin and J. DiGangi for ATom H_2O data. We thank the flight crew and support staff for the NSF/NCAR GV (*33*), which is part of NSF's Lower Atmosphere Observing Facilities, and the NASA DC-8, which is supported by the NASA Airborne Science Program and Earth Science Project Office. We thank the institutions and investigators responsible for the surface station records used in the study, including NOAA, CSIRO, Scripps, NIWA, SAWS, Program 416 from the French Polar Institute (IPEV), Tohoku University, and National Institute of Polar Research (NIPR), and in particular E. Dlugokencky, R. Langenfelds, S. Walker, G. Brailsford, S. Nichol, C. Labuschagne, W. Joubert, and S. Morimoto. We thank K. Lindsay, J.-F. Lamarque, and F. Vitt for useful modeling discussions. We thank R. Wanninkhof and two anonymous reviewers for their comments that improved this manuscript. CarbonTracker (CT2017, CT2019B) results provided by NOAA ESRL, Boulder, Colorado, USA, from the website http://carbontracker.noaa.gov. We thank A. Jacobson for the TM5 transport simulations. We thank the SOCAT and SOCCOM science teams and specifically P. Landschützer, S. Bushinsky, and A. Gray for providing flux estimates. **Funding:** This material is based on work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. Computing resources were provided by the Climate Simulation Laboratory at NCAR's Computational and Information Systems Laboratory (CISL). Data were collected using NSF's Lower Atmosphere Observing Facilities, which are managed and operated by NCAR's Earth Observing Laboratory. Additional sources of funding include NSF-PLR-1502301, NSF-ATM-0628388, NSF-ATM-0628519, NSF-ATM-0628575, NSF-PLR-1501993, NSF-PLR-1501292, NSF-PLR-1501997, NSF-AGS-1547626, NSF-AGS-1547797, NSF-AGS-1623745, NSF-AGS-1623748, NASA-NNX17AE74G, NASA-NNX15AJ23G, NASA-NNX16AL92A, and NOAA-NA15OAR4320071. **Author contributions:** M.C.L., B.B.S., K.M., and C.S. conceived of the study. B.B.S., E.J.M., J.D.B., R.F.K., E.A.K., B.C.D., R.C., S.C.W., K.M., C.S., P.T., A.S., Z.L., P.B.K., M.R., and D.M. collected and helped interpret CO_2 observational data. C.R., N.C., P.P., F.C., I.T.L., and W.P. produced and helped interpret model output. M.C.L. and B.B.S. performed the analysis and wrote the paper. All authors contributed to interpretation of the results and provided feedback on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Links providing access to the surface observational datasets used in this study are available in table S1; the aircraft data are available in references for HIPPO (*34*–*36*), ORCAS (*37*), and ATom (*20*). All model data are available via the UCAR/NCAR - Digital Asset Services Hub (DASH) Repository (*38*). Code to reproduce this calculation is documented and accessible (*39*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi4355
Materials and Methods
Supplementary Text
Figs. S1 to S19
Tables S1 to S4
References (*40*–*83*)

9 March 2021; accepted 18 October 2021
10.1126/science.abi4355

ORGANIC CHEMISTRY

Ligand-controlled divergent dehydrogenative reactions of carboxylic acids via C–H activation

Zhen Wang[1]†, Liang Hu[1]†, Nikita Chekshin[1], Zhe Zhuang[1], Shaoqun Qian[1], Jennifer X. Qiao[2], Jin-Quan Yu[1]*

Dehydrogenative transformations of alkyl chains to alkenes through methylene carbon-hydrogen (C–H) activation remain a substantial challenge. We report two classes of pyridine-pyridone ligands that enable divergent dehydrogenation reactions through palladium-catalyzed β-methylene C–H activation of carboxylic acids, leading to the direct syntheses of α,β-unsaturated carboxylic acids or γ-alkylidene butenolides. The directed nature of this pair of reactions allows chemoselective dehydrogenation of carboxylic acids in the presence of other enolizable functionalities such as ketones, providing chemoselectivity that is not possible by means of existing carbonyl desaturation protocols. Product inhibition is overcome through ligand-promoted preferential activation of C(sp^3)–H bonds rather than C(sp^2)–H bonds or a sequence of dehydrogenation and vinyl C–H alkynylation. The dehydrogenation reaction is compatible with molecular oxygen as the terminal oxidant.

Dehydrogenation of aliphatic chains is an important process in both bulk chemical industry and fine chemical synthesis. Various approaches have been developed to meet the synthetic needs of desaturation of carbonyl compounds (*1*). Exploration of syn-eliminations of selenoxide and sulfoxide intermediates has led to useful dehydrogenation methods in synthesis (*2*–*4*). Other organic reagents, such as *N*-*tert*-butyl phenylsulfinimidoyl chloride, hypervalent iodine, and *N*-oxoammonium salts, have also been developed to prepare enones (*5*–*7*). Electrochemically driven desaturation of carbonyl compounds was also recently achieved (*8*). An extensively researched catalytic pathway for dehydrogenation is the formation of Pd(II) enolates from ketones, with subsequent β-hydride elimination affording enone products (*9*–*13*). The Newhouse group has developed a method to desaturate free carboxylic acids through preformation of zinc enediolates, which are subsequently oxidized to α,β-unsaturated acids by Pd(II) complexes, with allyl acetate serving as the stoichiometric oxidant (Fig. 1A) (*14*). However, synthetically useful α-branched carboxylic acids were found to be challenging substrates for the zinc enediolate method. Hence, it is desirable to develop a complementary dehydrogenation approach through methylene C–H activation that can bypass the limitations of enolate chemistry. However, the development of synthetically useful dehydrogenation reactions by means of C–H activation without installing exogeneous directing groups faces two challenges: (i) the difficulty of activating methylene C–H bonds and (ii) product inhibition by, or overreaction of, the olefin products (Fig. 1A).

[1]The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA. [2]Discovery Chemistry, Bristol-Myers Squibb, P.O. Box 4000, Princeton, NJ 08543, USA.
*Corresponding author. Email: yu200@scripps.edu
†These authors contributed equally to this work.

In the past decade, we have focused on ligand-enabled C–H activation reactions directed by native functional groups such as free carboxylic acids, free aliphatic amines, and native amides (*15*). Catalytic β-C–H activation of methyl C–H bonds has led to the development of reactions such as arylation, olefination, acetoxylation, lactonization, and alkynylation. However, methylene C–H activation reactions of acyclic aliphatic substrates directed by innate functionalities such as carboxylic acids remain a substantial challenge (Fig. 1B) (*16*–*18*).

In 2006, Goldman, Brookhart, and colleagues reported a landmark example of alkane dehydrogenation through methylene C–H activation (*19*). In their system, the challenge of product inhibition was addressed by coupling the dehydrogenation with a tandem metathesis and hydrogenation reaction. Although this tandem catalytic system provides a promising strategy for converting low–molecular weight alkanes to high–molecular weight alkanes, so far it has proven difficult to use such a system to develop dehydrogenative transformations with synthetically relevant substrates as limiting reagents for the synthesis of olefins. Recently, dehydrogenation of cyclooctane by a combination of cobalt and tungsten catalysts via a radical pathway has also been observed, albeit in modest yields (*20*). Our early efforts to develop Pd(II)-catalyzed dehydrogenation reactions with substrates as limiting reagents through C–H activation involved the use of oxazoline directing groups with aliphatic acids. However, the synthetic utility of these transformations is limited by the extra steps required to install and remove the directing group (*21*); moreover, the reaction is limited to a single cyclopentanecarboxylic acid substrate. We report a pair of ligand-enabled Pd(II)-catalyzed divergent dehydrogenation reactions. A wide range of aliphatic carboxylic acids can be directly converted into α,β-unsaturated acids or γ-alkylidene butenolides through activation of either methylene or methyl C–H bonds. Product inhibition was overcome by the design of ligands with different bite angles that either prevented vinyl C–H activation of the typically more reactive α,β-unsaturated acids or enabled a tandem vinyl C–H activation and alkynyl bromide coupling, leading to butenolides that mask the carboxylic acid moiety from further reactions (Fig. 1C). Computational studies of the Pd-catalyzed C–H activation step with density functional theory (DFT) showed that the ligand forming a five-membered chelate with Pd is less competent in C(sp^2)–H cleavage compared with the ligand that forms a six-membered chelate, consistent with our experimental findings.

Our efforts toward the development of dehydrogenation reactions began with the preparation of a series of bidentate pyridine-pyridone ligands. On the basis of previous experimental and theoretical studies, 2-pyridones have been established as a powerful class of ligands for promoting Pd(II)-catalyzed C–H activation through ligand participation (*22*–*24*). Computational studies suggest that 2-pyridone is likely to serve as an X-type ligand in a manner similar to that of NH(acetyl) moiety, acting as an internal base directly involved in the C–H bond cleavage step (*23*). Prompted by the success of the acetyl-protected amino quinoline (APAQ) class of ligands for methylene C–H activation directed by the Wasa amide auxiliary (*25*), we began to design a class of bidentate ligands by substitution of the NH(acetyl) moiety with 2-pyridone and investigated their reactivity for methylene C–H activation directed by innate functionalities. Bidentate pyridone ligands **L5** to **L7** posited to coordinate to Pd (II) as six-membered chelates were prepared first since they were shown to be superior in the APAQ ligands (table S1). With hexanoic acid as the model substrate and 1,4-dioxane as the solvent, we observed the formation of 2-hexenoic acid in 15% yield with ligand **L5**. However, attempts at improving the reaction yield using ligands **L5** to **L7** under various conditions proved unsuccessful. In light of these results, we hypothesized that the low yield of the reaction might have resulted from inhibition of the catalyst by the olefin product. Indeed, it was observed that the addition of 15 mol % of the olefin product completely halted the reaction. We surmised that a potential pathway for product inhibition could be the unproductive activation of the β-vinyl C(sp^2)–H bond of 2-hexenoic acid. Indeed, in a hydrogen-deuterium exchange experiment, deuterium incorporation was observed at the vinyl position of the olefin product in the presence of AcOH-d_4 (see supplementary materials for details). A related study indicated that bidentate pyridone ligands forming five-membered chelates with Pd(II) are not reactive

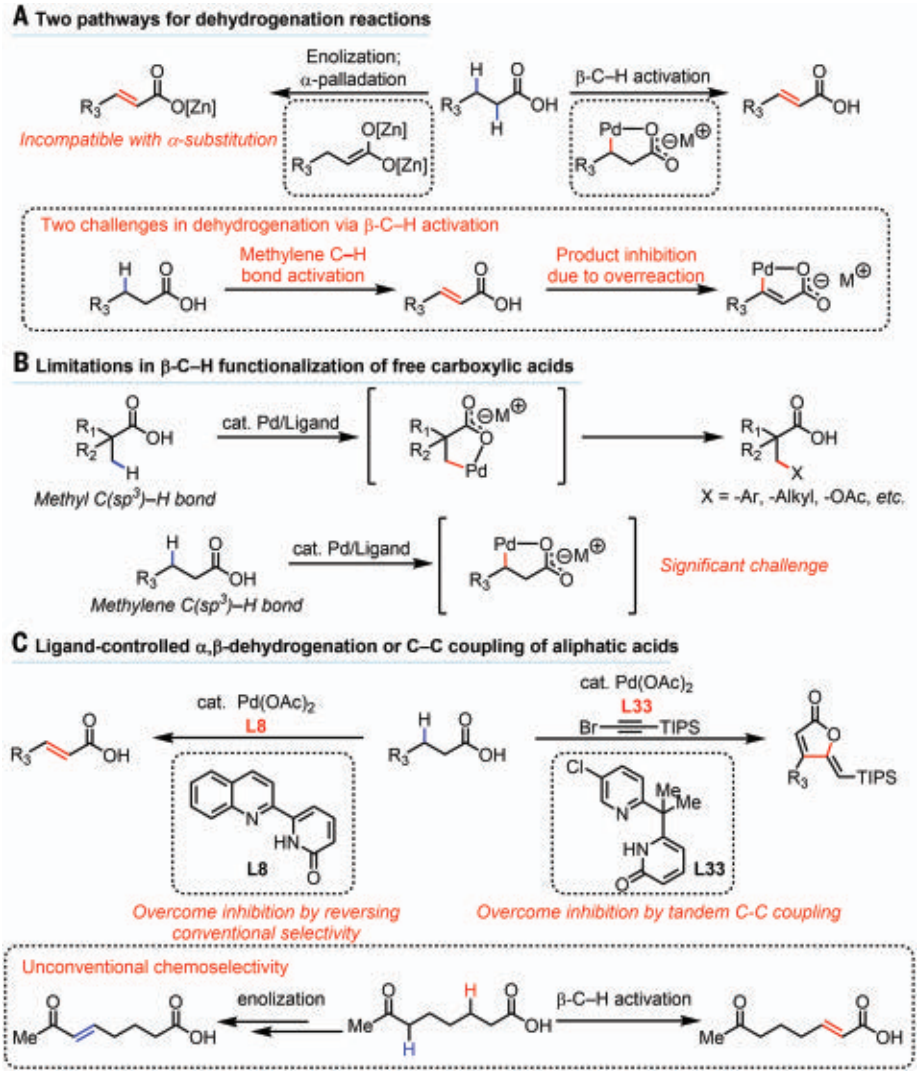


Fig 1. Dehydrogenation through C–H activation. (**A**) Two pathways for dehydrogenation reactions. (**B**) Limitations in β-C–H functionalization of free carboxylic acids. (**C**) Ligand-controlled α,β-dehydrogenation or C–C coupling of aliphatic acids. R_1, R_2, R_3, alkyl or aryl; M, metal; cat, catalyst; Ar, aryl; OAc, acetate; Me (methyl).

for C(sp^2)–H activation of nicotinic acids (*26*); cognizant of this, we examined five-membered chelate analogs of **L5** to **L7** to investigate whether they could promote dehydrogenation reactions through C(sp^3)–H activation while avoiding undesired β-vinyl C–H activation. We found that the yield of the dehydrogenation reaction carried out with ligand **L8** increased considerably to 71%. A deuterium exchange experiment with the olefin product using this five-membered chelating ligand did not incorporate deuterium at the β-vinyl position of the olefin product in the presence of AcOH-d_4. This suggests that activation of the β-vinyl C(sp^2)–H bond of 2-hexenoic acid is unfavorable with ligand **L8**. On the basis of these observations and comparative DFT studies (detailed in the supplementary materials), we propose that although both five- and six-membered chelating ligands are capable of enabling C(sp^3)–H activation, the bite angle of the former (78.7°) disfavors subsequent C(sp^2)–H activation of the olefin products compared with the latter (88.2°). Such a difference in reactivity could be attributed to the increased structural distortion with **L8** in the corresponding C(sp^2)–H cleavage transition state (TS). Substituent effects on the ligands were also surveyed (**L8** to **L19**), revealing that ligands bearing a quinoline or isoquinoline moiety (**L8** and **L17**) were superior, and **L8** was chosen for subsequent investigations. Using *tert*-amyl alcohol as a cosolvent, we were also able to decrease Pd loading from 10 to 4% and retain an 81% isolated yield for the desaturation of hexanoic acid to 2-hexenoic acid.

Subsequently, dehydrogenation of a wide range of free carboxylic acids with β-methylene C–H bonds was carried out under optimized conditions (Fig. 2). Simple linear and branched aliphatic acids, such as butyric acid and 4-methyloctanoic acid, provided the corresponding α,β-unsaturated acids in >75% yield with exclusive generation of the *E*-isomer (**2a** to **2f**). Carboxylic acids bearing ring systems also produced the corresponding α,β-unsaturated acids (**2g** to **2k**) in up to 87% yield. *tert*-butoxycarbonyl (Boc)– and toluenesulfonyl (tosyl)–protected amines and various ethers were all tolerated, giving the corresponding products in good yields (**2l** to **2p**). Phenylpropanoic acids bearing bromo-, chloro-, methylthio-, and Boc-protected amino substituents were successfully desaturated to the corresponding enoic acids (**2q** to **2u**); a substrate containing pyridine was also converted to product in high yield (**2v**). Additionally, phenylvaleric acid underwent smooth dehydrogenation to provide the corresponding α,β-unsaturated acid (**2w**) in good yield.

Additionally, we found that the reaction was effective for the dehydrogenation of α-branched carboxylic acid substrates, offering a complementary strategy to approaches on the basis of enolate oxidation. Dehydrogenation of 2-phenylbutanoic acid and 2-ethylbutanoic acid gave their corresponding dehydrogenated products in 65 and 51% yield, respectively (**2x** and **2y**). α,β-unsaturated acid **2y** generated from 2-ethylbutanoic acid predominantly gave the *E*-isomer. Isobutyric acid, a substrate with only methyl C–H bonds accessible, gave methacrylic acid in 42% yield (**2z**). For carboxylic acids bearing both β-methyl and β-methylene C–H bonds, we observed a slight preference for reactivity at the methylene C–H bonds over the methyl C–H bonds (**2aa**). Cyclopentanecarboxylic acid and cyclohexanecarboxylic acid were also viable substrates (**2ab** to **2ac**). In a complex setting, acetyl-protected lithocholic acid was dehydrogenated in 56% yield (**2ad**). These α,β-unsaturated acids are common building blocks and functional handles in synthetic chemistry that are amenable to a wide range of downstream derivatization, such as conjugate addition, epoxidation, and dihydroxylation.

Notably, the dehydrogenation reaction was also found to be selective for carboxylic acids in the presence of enolizable ketones (**2ae** to **2ag**) and esters (**2ah** to **2ai**), thus providing orthogonal chemoselectivity to enolate-based desaturation reactions (*2*–*14*). An α-quaternary substrate, 1-propylcyclohexane-1-carboxylic acid, was also dehydrogenated and subsequently coupled with an acrylate to give a complex fused lactone (**2aj**). Upon further optimization, this dehydrogenation reaction by means of β-C–H activation could lead to a synthetic disconnection that is not possible using the enolate pathway. Finally, we found that replacement of Ag_2CO_3 by other practical oxidants was also feasible. Several classes of representative substrates afforded moderate

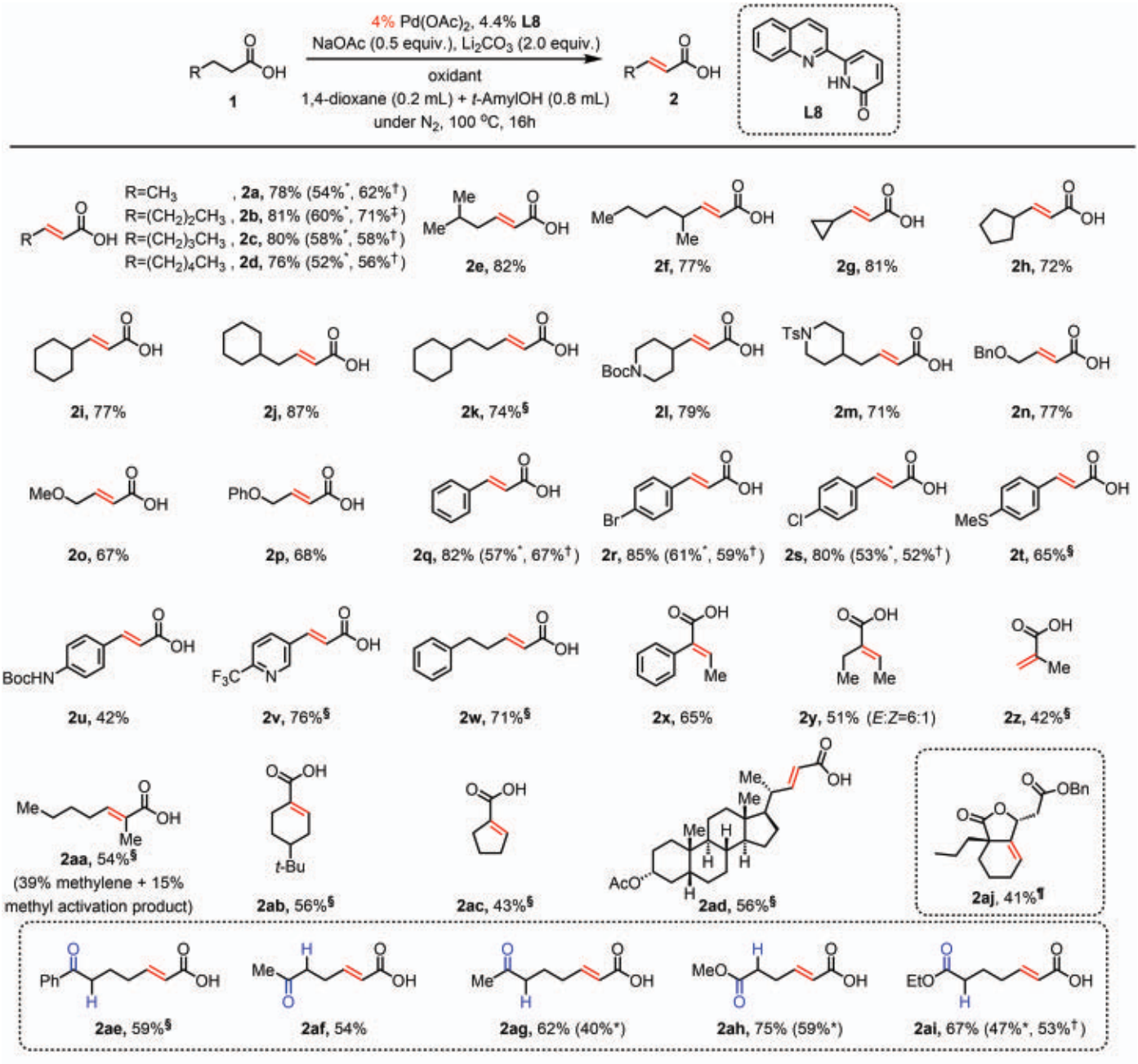


Fig 2. Carboxylic acid scope for the dehydrogenation reaction. Reaction conditions: **1** (0.1 mmol), $Pd(OAc)_2$ (4 mol %), **L8** (4.4 mol %), NaOAc (0.5 equiv), Li_2CO_3 (2.0 equiv), Ag_2CO_3 (2.0 equiv), 1,4-dioxane (0.2 ml), *tert*-amyl alcohol (0.8 mL), under N_2 at 100°C for 16 hours. See supplementary materials for details. *2 equiv of *tert*-butyl hydroperoxide (TBHP) in water were used as the oxidant instead of Ag_2CO_3 at 80°C. †1 atm of O_2 was used instead of Ag_2CO_3; BQ (1,4-benzoquinone, 1.0 equiv), CuBr (0.1 equiv), and dimethylformamide (DMF, 0.8 mL) were added; 8 h. ‡In a pressurized vessel at a scale of 1.0 mmol, 3 atm of O_2 were used instead of Ag_2CO_3; 1,4-benzoquinone (BQ, 1.0 equiv), CuBr (0.1 equiv), and DMF (dimethylformamide, 8.0 mL) were added; 8 h. ¶Benzyl acrylate (2.0 equiv) was added; 8 h §at 110°C. hexafluoro-2-propanol (HFIP 0.8 mL) was used as cosolvent instead of *tert*-amyl alcohol. Bn, benzyl; Et, ethyl; Ph, phenyl; *t*-Bu, *tert*-butyl.

yields with inexpensive *tert*-butyl hydroperoxide (TBHP) oxidant, ranging from 40 to 61% (**2a** to **2d**, **2q** to **2s**, and **2ag** to **2ai**). Using 1 atm of O_2 as the oxidant, up to 67% yield of the dehydrogenation products was obtained with substrates **2a** to **2d**, **2q** to **2s**, and **2ai** (for optimization details, see table S6). When the reaction was scaled up to 1.0 mmol, yield was further improved to 71% (**2b**) by using 3 atm of O_2.

Although the problem of product inhibition could be solved by the use of five-membered chelating ligands such as **L8** that inhibit activation of C(sp^2)–H bonds, we wondered whether it might also be possible to intercept the vinylic C–H palladated intermediate putatively generated by vinyl C(sp^2)–H activation of enoic acids when six-membered chelating ligands are used. If successful, this strategy would realize a tandem process that achieves double functionalization through sequential C(sp^3)–H and then C(sp^2)–H activation. Because the subsequent vinyl C–H functionalization step needs to be compatible with the conditions of

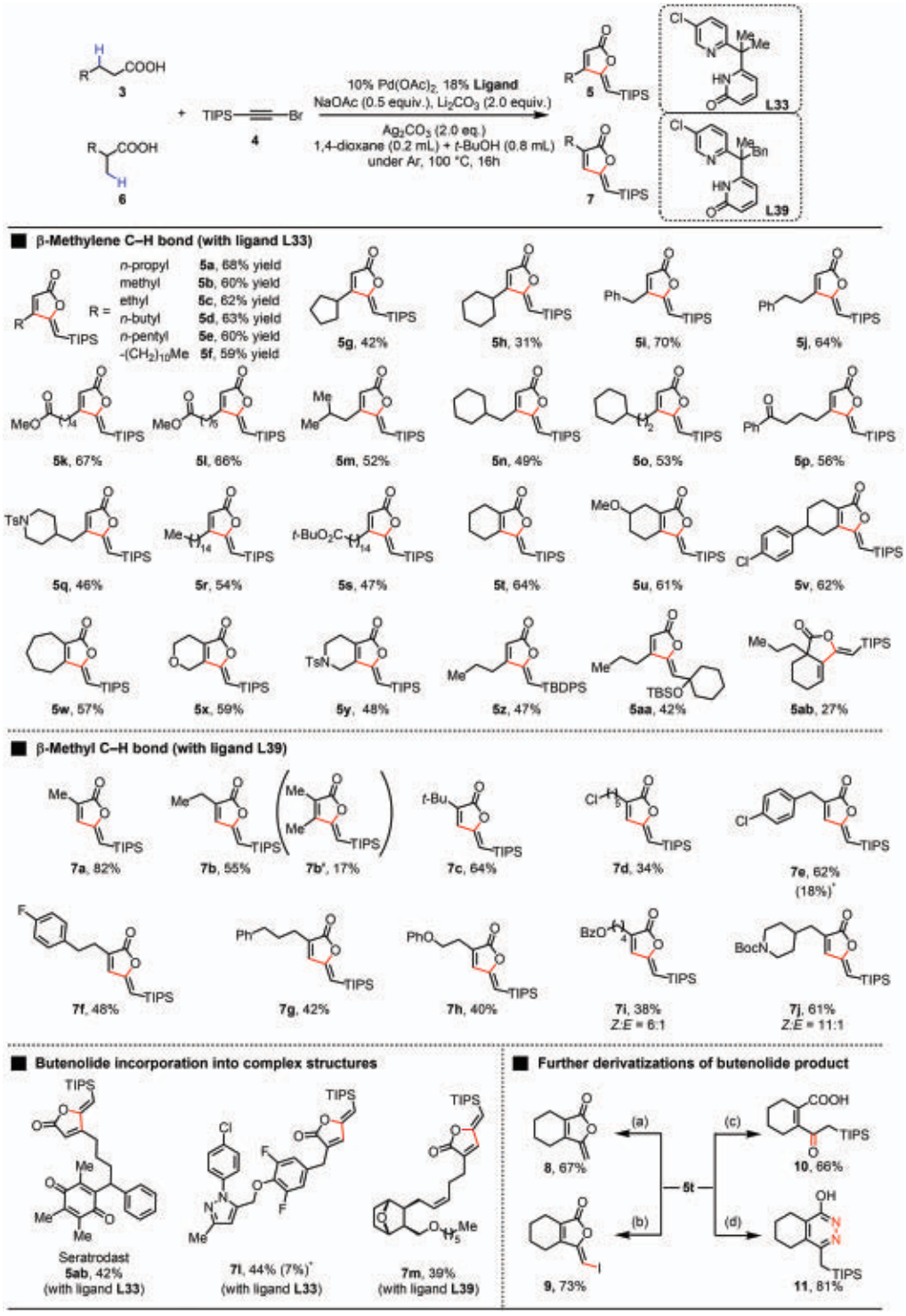


Fig 3. Substrate scope for butenolide formation. Reaction conditions: **3** or **6** (0.1 mmol), **4** (2.0 equiv), $Pd(OAc)_2$ (10 mol %), ligand (18 mol %), Ag_2CO_3 (2.0 equiv), Li_2CO_3 (2.0 equiv), NaOAc (0.5 equiv), 1,4-dioxane (0.2 mL), *t*-BuOH (0.8 mL), under argon at 100°C for 16 h. Isolated yields are reported. TIPS, triisopropylsilyl; TBDPS, *tert*-butyldiphenylsilyl; TBS, *tert*-butyldimethylsilyl. *Methylene activation product observable by ^{1}H NMR. Conditions for further derivatizations of butenolide product: (a) Tetrabutylammonium fluoride (TBAF); (b) (N-iodosuccinimide) NIS, Ag_2CO_3; (c) NaOH; (d) N_2H_4. See supplementary materials for details.

the dehydrogenation reaction, and the coupling partners employed must not interfere with the preceding dehydrogenation step, we began an extensive investigation into potential coupling partners that could be compatible with the dehydrogenation reaction. We discovered that, if well-documented bromoalkynes were used as coupling partners to form complex γ-alkylidene butenolides (*27*), a cascade reaction was possible. The development of synthetic methods for the construction of such scaffolds has been ongoing for decades as a result of the ubiquity of butenolides in natural products and bioactive molecules (*28*–*31*). However, existing methods invariably require multiple synthetic steps stemming from conventional retrosynthetic approaches. The method presented herein, based on an unconventional C–H dehydrogenation-alkynylation-cyclization cascade with aliphatic carboxylic acids, offers a retrosynthetic disconnection for one-step construction of γ-alkylidene butenolides.

A series of bidentate, six-membered chelating pyridine-pyridone ligands (**L5** and **L20** to

L39; table S8) were prepared for the examination of this cascade reaction. We observed that bidentate ligands bearing electron-withdrawing substituents such as **L33** and **L39** were superior, providing the butenolide product **5a** in 68 and 57% yields, respectively. As expected, the use of five-membered chelating ligands such as **L8** only provided α,β-unsaturated acids as products, and extensive modification of reaction conditions in an attempt to coax this ligand to enable the desired reactivity proved futile. This stark contrast in reactivity between the five- and six-membered chelating ligands further highlights how a simple design element—the ligand bite angle—could result in distinct chemical reactivities.

With optimal conditions and effective ligands in hand, we next examined the substrate scope of this cascade reaction (Fig. 3). Simple linear aliphatic carboxylic acids were found to be compatible with the current conditions, providing the β-substituted γ-alkylidene butenolide products in up to 70% yield (**5a** to **5s**). Among them, 3-cyclopentylpropionic acid **3g** and 3-cyclohexanepropionic acid **3h** both afforded their corresponding butenolides in lower yields (**5g** and **5h**), presumably as a result of unfavorable steric hindrance at their β-positions. Carboxylic acid substrates with an ester moiety gave 66 to 67% yield of the β-substituted γ-alkylidene butenolide products (**5k** and **5l**). Fatty acids were also found to be compatible with this transformation, providing the γ-alkylidene butenolide products in up to 54% yields, with the remaining starting materials mostly recovered (**5r** and **5s**). The reaction also proved effective in the formation of fused butenolides, affording products **5t** to **5y** in moderate yields. Notably, the transformation was found to be compatible with substrates of various ring sizes (**5t** and **5w**), substitutions (**5u** and **5v**), and saturated heterocycles (**5x** and **5y**). Various bromoalkynes (**5z** and **5aa**) were briefly evaluated as coupling partners; both gave the corresponding γ-alkylidene butenolide products in synthetically useful yields. α-Quaternary substrate 1-propylcyclohexane-1-carboxylic acid was also compatible, although it gave a lower yield (**5ab**).

Next, branched aliphatic carboxylic acids containing β-methyl C–H bonds were examined. These substrates also proved to be reactive under similar reaction conditions, and ligand **L39** was identified as the optimal ligand for these substrates. With this new set of reaction conditions, isobutyric acid was found to provide the α-substituted γ-alkylidene butenolide product **7a** in 82% yield. For substrates that possessed both β-methyl and β-methylene C–H bonds, preferential butenolide formation at the methyl C–H bond was observed. For example, with 2-methylbutyric acid as the substrate, α-substituted γ-alkylidene butenolide **7b** was formed as the major product, instead of the α,β-disubstituted γ-alkylidene butenolide **7b**. Substrates with alkyl (**7c** and **7d**), aryl (**7e** to **7i**), heteroatoms (**7h** to **7j**), and saturated heterocycle (**7j**) side chains were found to be compatible with the reaction conditions, and provided their corresponding γ-alkylidene butenolides in moderate to high yields. This cascade reaction ensured exclusive monoselectivity in the presence of multiple β-C–H bonds.

Because butenolide natural products are often bioactive, our method for the construction of γ-alkylidene butenolides from aliphatic carboxylic acids presented herein could potentially allow for late-stage introduction of butenolide moieties into complex natural products and drug molecules and thus provide access to hitherto-unknown hybrid molecules with potential biological activities. To illustrate the feasibility of this transformation, the antiasthmatic drug seratrodast was subjected to the standard reaction conditions and the corresponding butenolide hybrid **5ab** was obtained in 42% yield. The success with two more examples (**7l** and **7m**) further demonstrates the compatibility of this reaction with complex molecules.

The γ-alkylidene butenolide **5t** was briefly investigated for further derivatization (Fig. 3). The desilylation of the triisopropylsilyl (TIPS) protecting group was carried out by tetrabutylammonium fluoride (TBAF), affording **8** in 67% yield. Butenolide **5t** readily underwent iododesilylation to afford vinyl iodide product **9** in 73% yield, which could be further derivatized by cross-coupling reactions. Alkaline hydrolysis of **5t** provided a 1,4-dicarbonyl compound **10** in 66% yield. Treatment of **5t** with hydrazine in methanol at room temperature converted the γ-alkylidene butenolide into pyridazine **11** in 81% yield.

In summary, we have developed a pair of Pd-catalyzed dehydrogenation reactions that convert free aliphatic carboxylic acids into either α,β-unsaturated acids or γ-alkylidene butenolides. The design of bidentate pyridine-pyridone ligands with different bite angles (five-membered versus six-membered chelation) is crucial for the observed divergent reactivities.

REFERENCES AND NOTES

1. S. Gnaim, J. C. Vantourout, F. Serpier, P.-G. Echeverria, P. S. Baran, *ACS Catal.* **11**, 883–892 (2021).
2. K. B. Sharpless, R. F. Lauer, A. Y. Teranishi, *J. Am. Chem. Soc.* **95**, 6137–6139 (1973).
3. H. J. Reich, I. L. Reich, J. M. Renga, *J. Am. Chem. Soc.* **95**, 5813–5815 (1973).
4. B. M. Trost, T. N. Salzmann, K. Hiroi, *J. Am. Chem. Soc.* **98**, 4887–4902 (1976).
5. T. Mukaiyama, J.-I. Matsuo, H. Kitagawa, *Chem. Lett.* **29**, 1250–1251 (2000).
6. K. C. Nicolaou, Y. L. Zhong, P. S. Baran, *J. Am. Chem. Soc.* **122**, 7596–7597 (2000).
7. M. Hayashi, M. Shibuya, Y. Iwabuchi, *Org. Lett.* **14**, 154–157 (2012).
8. S. Gnaim *et al.*, *Nat. Chem.* **13**, 367–372 (2021).
9. R. J. Theissen, *J. Org. Chem.* **36**, 752–757 (1971).
10. Y. Ito, T. Hirao, T. Saegusa, *J. Org. Chem.* **43**, 1011–1013 (1978).
11. J.-Q. Yu, H.-C. Wu, E. J. Corey, *Org. Lett.* **7**, 1415–1417 (2005).
12. T. Diao, S. S. Stahl, *J. Am. Chem. Soc.* **133**, 14566–14569 (2011).
13. M. Chen, G. Dong, *J. Am. Chem. Soc.* **139**, 7757–7760 (2017).
14. Y. Zhao, Y. Chen, T. R. Newhouse, *Angew. Chem. Int. Ed.* **56**, 13122–13125 (2017).
15. K. M. Engle, T.-S. Mei, M. Wasa, J.-Q. Yu, *Acc. Chem. Res.* **45**, 788–802 (2012).
16. O. Daugulis, J. Roane, L. D. Tran, *Acc. Chem. Res.* **48**, 1053–1064 (2015).
17. P. X. Shen, L. Hu, Q. Shao, K. Hong, J.-Q. Yu, *J. Am. Chem. Soc.* **140**, 6545–6549 (2018).
18. A. Uttry, S. Mal, M. van Gemmeren, *J. Am. Chem. Soc.* **143**, 10895–10901 (2021).
19. A. S. Goldman *et al.*, *Science* **312**, 257–261 (2006).
20. J. G. West, D. Huang, E. J. Sorensen, *Nat. Commun.* **6**, 10093 (2015).
21. R. Giri, N. Maugel, B. M. Foxman, J.-Q. Yu, *Organometallics* **27**, 1667–1670 (2008).
22. P. Wang *et al.*, *J. Am. Chem. Soc.* **138**, 9269–9276 (2016).
23. P. Wang *et al.*, *Nature* **551**, 489–493 (2017).
24. Y.-Q. Z. Chen *et al.*, *J. Am. Chem. Soc.* **140**, 17884–17894 (2018).
25. G. Chen *et al.*, *Science* **353**, 1023–1027 (2016).
26. Z. Li *et al.*, *Science* **372**, 1452–1457 (2021).
27. F. Ghiringhelli, A. Uttry, K. K. Ghosh, M. van Gemmeren, *Angew. Chem. Int. Ed.* **59**, 23127–23131 (2020).
28. A. Bauer, N. Maulide, *Chem. Sci.* **10**, 9836–9840 (2019).
29. S. J. Gharpure, L. N. Nanda, M. K. Shukla, *Org. Lett.* **16**, 6424–6427 (2014).
30. P. Tuchinda *et al.*, *Phytochemistry* **30**, 2685–2689 (1991).
31. U. A. Pereira *et al.*, *MedChemComm* **7**, 345–352 (2016).

ACKNOWLEDGMENTS

We thank S. Chan for proofreading and S. Chan and D. Strassfeld for helpful suggestions in preparing the manuscript. Z.W. thanks P. Wang for help with ligand synthesis. We also thank the X-ray Crystallography Facility (UCSD) for x-ray crystallography. **Funding:** We acknowledge The Scripps Research Institute, NIH (NIGMS, 2R01GM084019), and Bristol-Myers Squibb for financial support. We acknowledge the Ghadiri Lab and the High Performance Computing Facility at Scripps Research for providing computational resources. **Author contributions:** J.-Q.Y. conceived the concept. Z.W. developed the ligands and the dehydrogenation reaction. L.H. developed the butenolide formation reaction. N.C. carried out computational modeling and analysis. J.X.Q. participated in substrate scope survey and discussions and also provided substrates for late-stage functionalization. J.-Q.Y. directed the project. **Competing interests:** J.-Q.Y., Z.W., and L.H. are inventors on a patent application related to this work (US Patent Application 63/203,241) filed by The Scripps Research Institute. The authors declare no other competing interests. **Data and materials availability:** Metrical parameters for the structures of ligand/Pd complexes (see supplementary materials) are available free of charge from the Cambridge Crystallographic Data Centre under reference numbers CCDC-2062539 and CCDC-2062540. All other data are in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl3939
Materials and Methods
Figs. S1 to S3
Tables S1 to S13
References (*32*–*56*)
X-ray Crystallographic Data
Computational Data and Analysis
NMR Spectra

10 July 2021; accepted 1 September 2021
Published online 11 November 2021
10.1126/science.abl3939

Single-Cell RNA Sequencing
Dolomite Bio announces the release of its RNAdia kit. Together with the Nadia platform, the RNAdia kit offers a high-quality, low-cost, complete solution for single-cell RNA sequencing (scRNA-Seq). At half the cost of its competitors, the new RNAdia kit enables more scientists to access the rapidly expanding field of single-cell research. The launch of RNAdia includes the option to purchase bioinformatic services, in the same way as you would a kit, ensuring easy access to expert data analysis, even if you don't have an in-house bioinformatician. This provides a complete, seamless solution, from cells to data, in a workflow designed to take even less time than before. Nadia takes scRNA-Seq to the next level by using automation and flexibility to generate superior reproducible single-cell data. Encapsulate up to eight samples in parallel, in under 20 min, with over 50,000 single cells captured per cartridge in a run. Adding the Nadia Innovate to the Nadia Instrument transforms it into an open system for the development of novel protocols and applications.
Dolomite Bio
For info: +44-(0)-1763-252-102
www.dolomite-bio.com

Bst DNA/RNA Polymerase
Bst DNA/RNA Polymerase is a mixture of Bst polymerase and extremely thermostable reverse transcriptase (65°C tolerant), which is suitable for isothermal amplification reaction of RNA. It can detect low-sensitivity RNA molecules. This enzyme is recommended in isothermal amplification experiments using RNA as a template. In addition, Bst DNA/RNA Polymerase can also perform isothermal amplification of DNA templates.
Beijing SBS Genetech
For info: +86-(0)-10-82784292
www.sbsgenetech.com/store/products/bst-dna-rna-polymerase

RNA Sequencing Kit
Bio-Rad Laboratories announces the SEQuoia RiboDepletion Kit, which improves assay efficiency by eliminating irrelevant ribosomal RNA (rRNA) fragments from an RNA-Seq library. It is designed for genomics researchers targeting rare transcripts or working with a limited or degraded sample. The kit can retain RNA transcripts that could be lost when using other methods. In addition, the efficient elimination of irrelevant rRNA fragments helps reduce the costs of sequencing RNA samples. Workflow is streamlined to less than 2 h. The kit works with a broad input range and is compatible with most available library prep kits.
Bio-Rad Laboratories
For info: +1-800-424-6723
www.bio-rad.com

Automated In Situ Hybridization System
The new NanoVIP from BioGenex is a diverse, fully automated system for fluorescence in situ hybridization (FISH), in situ hybridization (ISH), and microRNA ISH. Its reliable automation, combined with eXACT temperature modules and liquid-level sensors for accurate liquid handling, ensures robust, reproducible results. The all-in-one system can run 10 different protocols simultaneously. This instrument can simplify a complex 33-step manual FISH protocol and reduce it to three simple steps: load slides, select protocol, and view completed slides. The NanoVIP software is an intuitive, easy-to-use software that creates and stores user-defined protocols. With quality-engineered hardware, unparalleled software capabilities, and one of the smallest footprints on the market, the NanoVIP remains an ideal choice for both clinical and research laboratories.
BioGenex
For info: +1-800-421-4149
biogenex.com/automation/research-instruments/nanovip

COVID-19 Variant Sequencing Service
Zymo Research offers the COVID-19 Variant Sequencing Service. Detecting mutations quickly helps prevent the spread of new viral strain types and can provide an early warning of potential vaccine escape. This end-to-end service includes the necessary collection devices and automated extraction technology, with RNA sequencing completed on Illumina platforms. Zymo Research's DNA/RNA Shield is used for collecting, transporting, and preserving the samples. The preservative inactivates the SARS-CoV-2 virus while stabilizing its RNA, allowing safe room-temperature transport for robust library prep and sequencing. The COVID-19 Variant Sequencing Service workflow consists of sample collection, processing, sequencing, and reporting. COVID-positive samples are collected from test centers and transported to Zymo Research and its affiliates for processing. Automated RNA extraction from the collected samples is performed using the KingFisher Flex robotics platform. Sequencing is then analyzed using the hardware-accelerated DRAGEN COVIDSeq Analysis Pipeline. The final report includes COVID-19 strain identification, genome sequencing coverage, confirmed mutations, consensus genome, and raw sequence files.
Zymo Research
For info: +1-888-882-9682
www.zymoresearch.com/pages/covid19-variant-sequencing-service

Real-Time PCR
The LightCycler 480 System is a high-performance, medium- to high-throughput PCR platform (96- or 384-well plates) that provides various methods for gene detection, gene expression analysis, genetic variation analysis, and array data validation. The benchtop instrument is easily customizable to meet changing user requirements and can be integrated into everyday use as a robotically controlled, automated high-throughput solution. Use basic and advanced methods for gene expression and genotyping based on endpoint analysis or melting curves. Benefit from cutting-edge thermal block and optical technology that leads to excellent well-to-well homogeneity and maximized detection sensitivity. Use a wide range of probe formats (e.g., fluorescence resonance energy transfer or HybProbe probes, SimpleProbe probes, or hydrolysis probes) and fluorescent dyes (e.g., SYBR Green I, ResoLight Dye, FAM, HEX, Cy5, LightCycler Red 610, LightCycler Red 640, and LightCycler CYAN 500). Analyze true raw data without the need for passive reference dyes or normalization plates.
Roche Sequencing and Life Science
For info: +1-800-262-4911
sequencing.roche.com/en-us

Electronically submit your new product description or product literature information! Go to www.science.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

EPFL

Faculty Position in Computational Cancer Research

at the Ecole polytechnique fédérale de Lausanne (EPFL)

The School of Life Sciences of EPFL (Ecole polytechnique fédérale de Lausanne) invites applications for a tenure track assistant professor position in the field of **Computational Cancer Research**.

This search is part of major initiatives in the Lake Geneva region to study the mechanisms of cancer development and progression with an application to therapy, using and developing advanced computational techniques, including machine learning.

The successful candidate will: join the faculty of the Swiss Institute for Experimental Cancer Research (ISREC) http://isrec.epfl.ch; develop an independent internationally prominent research program in the broad domain of computational cancer research and its potential therapeutic applications; commit to excellence in teaching at the undergraduate and graduate levels; and supervise PhD students and postdoctoral fellows. Candidates may work in any computational cancer-related area, including cancer genetics, functional genomics, multi-omics analysis of patient cohorts and genome instability, epigenetic regulation of cancer genotypes, cancer metabolism, computational cancer genetics, tumor evolution, and tumor escape mechanisms.

The successful candidate will be a part of EPFL's cancer research institute (ISREC), and is expected to perform and coordinate highly interactive biomedical research, reaching out and taking advantage of EPFL's interdisciplinary campus (Schools of Basic Sciences, Engineering, and Information and Communication Technologies) and its involvement in the multi-institutional Swiss Cancer Center Leman, which brings together EPFL, the Universities of Lausanne and Geneva, and clinical and translational research components of the University Hospitals of Lausanne and Geneva. The School of Life Sciences currently has five broad strategic aims: Biological data science, structural biology, environmental biology, organoid bioengineering and Neuro-X.

Applications should include a cover letter, a curriculum vitae, a list of publications (annotated to indicate the candidate's contributions) a synopsis of major accomplishments, and a concise statement of future research agenda and teaching interests, along with the contact information of at least three references who are ready to supply their letter upon request.

Applications should be uploaded as PDF files to the recruitment web site:

https://facultyrecruiting.epfl.ch/position/35848194

Formal evaluation of candidates will begin on **January 8, 2022,** and continue until the position is filled.

Enquiries may be sent to:
Prof. Freddy Radtke
Search Committee Chair
E-mail: cancer.research@epfl.ch

For additional information on EPFL, the ISREC institute and the school of life sciences, please consult: www.epfl.ch, sv.epfl.ch

EPFL is an equal opportunity employer and family friendly university. It is committed to increasing the diversity of its faculty. It strongly encourages women to apply.

EPFL

Faculty Position in Cancer Research with focus on Chemical/Bio-engineering

at the Ecole polytechnique fédérale de Lausanne (EPFL)

The School of Life Sciences of EPFL (Ecole polytechnique fédérale de Lausanne) invites applications for a tenure track assistant professor position in the field of **Cancer Chemical/Bio-engineering**.

This search is part of major initiatives in the Lake Geneva region to study the mechanisms of cancer development and progression with an application to therapy, using and developing advanced technologies that bridges the fields of chemistry, biology and engineering to study fundamental processes in cancer and or identify novel therapeutic targets and drugs to treat cancer.

The successful candidate will: join the faculty of the Swiss Institute for Experimental Cancer Research (ISREC) http://isrec.epfl.ch; develop an independent internationally prominent research program in the broad domain of chemical biology and bio-engineering of cancer and its potential therapeutic applications; commit to excellence in teaching at the undergraduate and graduate levels; and supervise PhD students and postdoctoral fellows. Candidates may work in any cancer-related area, including development of chemical probes that interfere with or label growth controlling proteins, engineering of defined functionalized matrices to study patient derived tumoroids, or in computational and structural biology with the aim to transform large numbers of un-druggable human proteins into pharmacological targets, tumor evolution, and tumor escape mechanisms.

The successful candidate will be a part of EPFL's cancer research institute (ISREC), and is expected to perform and coordinate highly interactive biomedical research, reaching out and taking advantage of EPFL's interdisciplinary campus (Schools of Basic Sciences, Engineering, and Information and Communication Technologies) and its involvement in the multi-institutional Swiss Cancer Center Leman, which brings together EPFL, the Universities of Lausanne and Geneva, and clinical and translational research components of the University Hospitals of Lausanne and Geneva. The School of Life Sciences currently has five broad strategic aims: Organoid bioengineering, structural biology, biological data science, environmental biology, and Neuro-X.

Applications should include a cover letter, a curriculum vitae, a list of publications (annotated to indicate the candidate's contributions) a synopsis of major accomplishments, and a concise statement of future research agenda and teaching interests, along with the contact information of at least three references who are ready to supply their letter upon request.

Applications should be uploaded as PDF files to the recruitment web site:

https://facultyrecruiting.epfl.ch/position/35848193

Formal evaluation of candidates will begin on **January 8, 2022,** and continue until the position is filled.

Enquiries may be sent to:
Prof. Freddy Radtke
Search Committee Chair
E-mail: cancer.research@epfl.ch

For additional information on EPFL, the ISREC institute and the school of life sciences, please consult: www.epfl.ch, sv.epfl.ch

EPFL is an equal opportunity employer and family friendly university. It is committed to increasing the diversity of its faculty. It strongly encourages women to apply.

ScienceCareers online @sciencecareers.org

FOX CHASE CANCER CENTER

FACULTY POSITION IN CANCER EPIGENETICS at FOX CHASE CANCER CENTER

The Fox Chase Cancer Center invites applicants for tenure-track faculty positions at the level of **assistant professor** and **associate professor** in the **Cancer Epigenetics Institute** and **Cancer Signaling and Epigenetics Program**. While outstanding candidates working in all areas of epigenetics are invited to apply, we are particularly soliciting applications from candidates that are interested in the following areas:

1. the relationship between higher order genome organization and spatial transcriptomics related to cancer development, progression and/or resistance;
2. the role of modified RNAs and epi-transcriptomics in cancer;
3. the use of computational and systems analyses for chromatin biology;
4. the relationship between immune oncology and epigenetics;
5. the relationship between metabolic pathways and epigenetic control.

Successful applicants will join a highly interactive institution that is committed to bridging basic discovery to clinical application through collaboration. Fox Chase was among the nation's first institutions to receive Comprehensive Cancer Center designation from the NCI in 1974. There are 119 research faculty that generate $63.2 million in funding. The research faculty are surrounded by clinical colleagues that provide care for over 105,000 individuals and support >170 clinical trials, of which >60 are investigator-initiated. Two particular highlights of the center are the superb clinical pipeline that facilitates rapid translation of basic science discoveries into the clinic and the world-class core facilities that are easily accessed.

Applicants interested in joining such an environment should have a Ph.D. and/or M.D. degree with an outstanding record of research productivity. The application should contain the following information - a curriculum vitae, a brief (up to two pages) statement of research interests and future goals and a list of three individuals providing letters of recommendation. Please send the application via email to Johnathan Whetstine (**Epigenetics@fccc.edu**), Director, Cancer Epigenetics Institute, Co-Leader, Cancer Signaling and Epigenetics Program, Fox Chase Cancer Center.

Equal Opportunity Employer

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Confused about your next career move?

Download Free Career Advice Booklets!

ScienceCareers.org/booklets

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers | AAAS In partnership with:

FASEB Federation of American Societies for Experimental Biology
University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

ScienceCareers online @sciencecareers.org

TENURED OR TENURE TRACK FACULTY POSITION

The Institute for Medical Engineering and Science (IMES) http://imes.mit.edu/ at the Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts invites applicants to apply for tenured or tenure track **faculty position(s) in the area of biomedical imaging** to begin **July 1, 2022** or thereafter.

The successful candidate(s) will be appointed to a tenured or tenure track faculty position commensurate on experience at MIT as the Athinoula A. Martinos Professor. The successful candidate is expected to work closely with and have an appointment at the Athinoula A. Martinos Center for Functional and Structural Biomedical Imaging, in particular the facilities housed at Massachusetts General Hospital (MGH).

IMES is pioneering novel research and graduate education paradigms at the convergence of engineering, science, and medicine. IMES also houses the Harvard-MIT Health Science and Technology (HST) program, the Medical Electronic Device Realization Center, the Center for Microbiome Informatics and Therapeutics, and MIT's Clinical Research Center.

Appointments will be made in IMES and the Electrical Engineering and Computer Science (EECS) department at MIT.

Applicants should hold a Ph.D. in engineering, science, or a related field (including MD or MD/PhD) by the beginning of employment. The candidate should have demonstrated excellence in original research and teaching. A minimum of 12 months of teaching experience is required. The successful candidate will be expected to fulfill all teaching, research and service obligations of a faculty member at MIT. Faculty duties include teaching at the undergraduate and graduate levels, advising students, conducting original scholarly research, developing course materials at the graduate and undergraduate levels, and service to MIT and the profession.

Interested candidates should submit application materials electronically at **https://school-of-engineering-faculty-search.mit.edu/imes-senior-search/register.tcl**. Each application should include: curriculum vitae; the names and addresses of three or more references; a strategic statement of research interests; and a statement of teaching interests. Each candidate must arrange for reference letters to be uploaded at **https://imes-senior-search.mit.edu/letters**.

Candidates should provide a statement regarding their views on diversity, inclusion, and belonging, including past and current contributions as well as their vision and plans in these areas.

Questions should be addressed to **imes-search@mit.edu**. To be considered complete, responses must be received by **December 15, 2021**.

With MIT's strong commitment to diversity in engineering education, research, and practice we especially encourage minorities and women to apply.

MIT is an equal employment opportunity employer. All qualified applicants will receive consideration for employment and will not be discriminated against on the basis of race, color, sex, sexual orientation, gender identity, religion, disability, age, genetic information, veteran status, ancestry, or national or ethnic origin.

OPPORTUNITIES IN CHINA

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Faculty Positions in Life and Health Sciences
The Chinese University of Hong Kong, Shenzhen

The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen) invites applications for full-time faculty positions in Life and Health Sciences at ranks of professors/associate professors/assistant professors. Qualified candidates must have an earned doctoral degree in life sciences, postdoctoral training and research and teaching experiences. Applicants for professor and associate professor positions should have an established research program, teaching experience and an excellent publication record. Lecturer and Senior Lecturer positions are also available for candidates with teaching experience in life sciences. Applicants in the following disciplines are preferred.

Bioinformatics. The ideal candidates should demonstrate research interests and teaching potentials in bioinformatics and related fields including data mining, machine learning and mathematical modeling focusing on biological sciences.

Biomedical Engineering. Jointly developed by the School of Life and Health Sciences (LHS) and School of Science and Engineering (SSE), this program is recruiting faculty members with teaching and research interests in biomechanics, tissue engineering, biomaterials, regenerative medicine, biomedical imaging, nanomedicine, biosensors and bio-robotics, and systems and synthetic biology.

Pharmaceutical Science. The applicants are expected to teach and conduct research in pharmaceutics, medicinal chemistry, pharmaceutical analysis, pharmacokinetics, pharmacogenomics, biologics and immunotherapy, and pharmacy administration.

Established in 2014, CUHK-Shenzhen is a research-intensive university that inherits the fine academic traditions of The Chinese University of Hong Kong. The University adopts a tenure-track system for Assistant Professors and above. English is the main language for classroom teaching, and graduates receive degrees of The Chinese University of Hong Kong. Please visit **lhs.cuhk.edu.cn** for additional descriptions of the program. Interested applicants should submit their curriculum vitae, teaching statements and research descriptions to: http://academicrecruit.cuhk.edu.cn/lhs. Applications will be reviewed on a rolling basis until the positions are filled.

Technische Universität Braunschweig

Zertifikat seit 2007 audit familiengerechte hochschule

The **Faculty of Electrical Engineering, Information Technology, Physics,** Institute for Electrical Measurement Science and Fundamental Electrical Engineering offers a

University professorship (m/f/d) for "Cryogenic Electronic Quantum Devices" Salary-Gr. W2

The professorship is expected to focus on physics and engineering of devices and instrumentation for electrical quantum metrology at low temperature in cryogenic environment, e.g. for voltage standard (Josephson effects), current standard (Single electron pumps) and/or conductance standard (quantum Hall effect) and related topics. The quantum metrology activities should include chip design and processing of nanostructures and devices, as well as their analysis and further development by metrological methods. In terms of lecturing, the professorship will be part of the bachelor and master programs of the Department of Electrical Engineering, Information Technology and Physics.

In addition to scientific excellence, an important criterion will be the fit with the activities in the Cluster of Excellence QuantumFrontiers (QF) and the research centre Laboratory for Emerging Nanometrology (LENA).

Vacancy offer:
https://www.tu-braunschweig.de/en/vacancies

Please send your application until **January 15th, 2022** using the **online form** of TU Braunschweig.

By **Pijar Religia**

My weekly assignment

During my early months of graduate school, I struggled with a weekly task: sitting down at my computer and writing an update for my advisers. More often than not, I felt stuck. I didn't know what to write because I didn't think I had done enough work worth sharing. My stress level skyrocketed, and I started to do experiments just for the sake of having something to report. It was the perfect recipe for making no progress. But by the end of my program, I realized the weekly updates didn't need to be packed with data and accomplishments. Instead, they could serve as a tool to refine my thinking and get the feedback I needed. I was glad my advisers required them.

I had moved to Japan from my home country of Indonesia, where I was accustomed to a different research and training culture. When I pursued my master's degree there, my research was much less structured and I only went to my adviser when I had questions or problems. That presented its own set of issues—for example, when my graduation deadline crept up on me, I hadn't yet completed everything I wanted to. Still, it was what I was used to.

In Japan, I was surprised to learn that I was expected to update my advisers on a weekly basis. They didn't tell me what details to report. So, while I was coming up with my research questions and study design, I wrote brief updates about any new ideas or experimental plans I had in mind. But seeing the progress my lab mates were reporting, I soon felt pressure to communicate something more substantive. That spurred me to dive into experimental work faster than I should have.

I started to do experiments each week, usually frantically putting them together during the 2 or 3 days leading up to my weekly email. The experiments worked, but I had a nagging feeling I was getting lucky and that eventually I'd be discovered for being sloppy. During one group meeting, my adviser scolded a lab member for wasting time, energy, and money on a poorly designed experiment. The same could be said for much of my work up to that point, I knew. Something needed to change.

I realized my earliest reports, during the time I'd spent planning my project, could be a model—one that I could build and expand on. I hadn't gone into much detail in those emails because ideas and plans didn't strike me as significant accomplishments. But after stumbling through experiment after experiment, I came to see that there was real value in laying out detailed plans.

"It helped me collect the small stones that I ultimately used to construct the stairs to my Ph.D."

I wasn't sure how my advisers would respond if I reverted to the experimental design stage, but I decided it was worth a try. I spent the weeks that followed reading papers and thinking about the big picture goals of my project. I was nervous when I wrote my first email update listing all the papers I'd read and ideas I'd explored—but lacking concrete data. To my relief, my advisers didn't object. After one report, which focused on papers I'd combed to figure out the right method for one experimental step, one of my advisers reached out to another professor who had relevant expertise to ask for advice. That helped me make a final decision on my method more quickly than I would have on my own.

Expanding my idea of progress helped me feel more productive and gave me the intellectual space to explore topics more deeply. My research became more efficient and I ran into fewer dead ends. I also realized that writing out detailed ideas and plans in the weekly updates provided a valuable opportunity to articulate them for myself, which helped me approach my advisers more confidently and engage in more productive discussions.

Over time, I began to see the updates not as a burden, but as the tool they are intended to be—to gradually build the chapters of my research story, and to spur constructive feedback from mentors along the way.

On my last day in my Ph.D. lab, I did something I never could have foreseen during the first 6 months: I thanked my professors for the lab tradition of weekly progress reports. It helped me collect the small stones I ultimately used to construct the stairs to my Ph.D. ■

Pijar Religia is a visiting researcher at Osaka University. Do you have an interesting career story to share? Send it to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

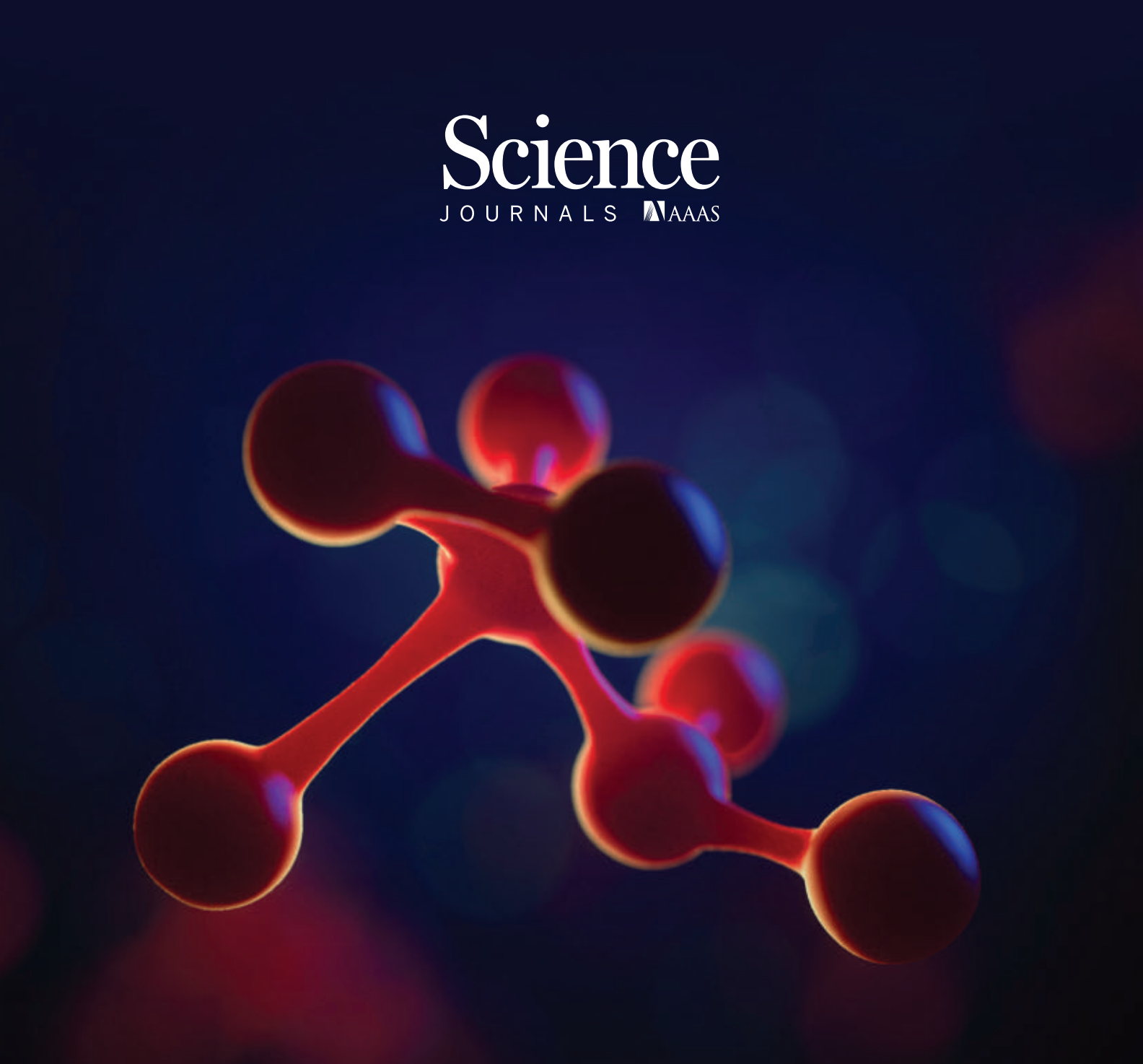
Pushing the Boundaries of Knowledge

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

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**