What's holding up variant-proof coronavirus vaccines? p. 566

Continually pumped time crystals pp. 576 & 670

Promising neurocognitive therapeutic wins *Science* & PINS Prize p. 588

Science

$15
5 AUGUST 2022
science.org

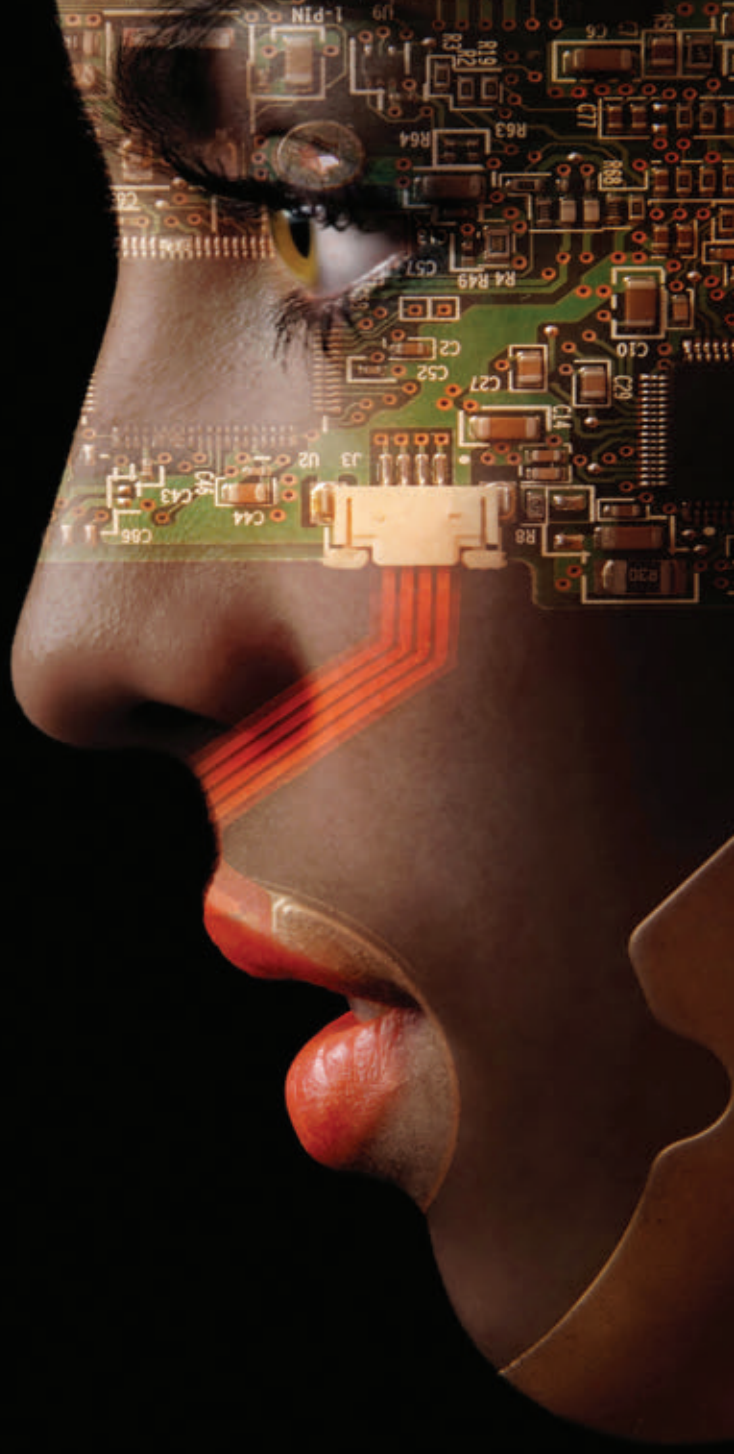
AAAS

SPECIAL ISSUE

GRASS

Essential and overlooked biomes p. 590

Share **Your Robotics Research** with the **World**.



Shaping the future of robotics with high impact research!

As a multidisciplinary online-only journal, *Science Robotics* publishes original, peer-reviewed, research articles that advance the field of robotics. The journal provides a central forum for communication of new ideas, general principles, and original developments in research and applications of robotics for all environments.

Submit your research today. Learn more at: **science.org/journal/scirobotics**

Twitter: @SciRobotics
Facebook: @ScienceRobotics

Science Robotics
AAAS

Seeking life-saving solutions

Dr. Efthymia Iliana Matthaiou

To obtain the most efficient and successful research results, scientists need robust and reliable equipment.

Growing up in rural Greece, nature captivated Dr. Efthymia Iliana Matthaiou. As she says, "I especially like finding ways to answer questions." When her grandfather was diagnosed with cancer and there were limited treatment options, Matthaiou decided to pursue a career in translational research.

In 2020, after completing her postdoctoral training, Matthaiou started an academic scientist position at Stanford University, where she investigates host–pathogen interactions in the airways. She is particularly interested in patients with cystic fibrosis. "Unfortunately, this patient population is prone to infections," she explains.

Like other life science labs, Matthaiou's lab requires a variety of equipment, ranging from an incubator and biosafety cabinet to an ultralow temperature freezer and liquid-nitrogen storage. At the start, though, she needed to set up her laboratory. "We were building up a new lab," she says. "We had zero equipment."

Equipping a lab

An indispensable facet of Matthaiou's research is the collection and safe storage of samples. "We are collecting patient samples on a daily basis," she says. "These are precious human samples that cannot be replaced."

To maintain these samples in stable conditions, Matthaiou chose a Thermo Scientific −80°C TSX Upright Freezer. "We don't want to have temperature fluctuations, we need reliable equipment," she explains.

For samples that need to be stored at lower temperatures, Matthaiou opted for a Thermo Scientific CryoPlus Vapor Phase LN_2 (Liquid Nitrogen) Storage System. "What I love about this is the autofill liquid nitrogen feature," she says. "The vapor phase platform minimizes the chances of sample contamination. Plus, it offers easier access to the samples and may defend researchers from the risks of operating with liquid nitrogen."

Decontamination is essential when culturing infectious agents. For that purpose, Matthaiou selected a Thermo Scientific Forma Steri-Cycle CO_2 Incubator. A key feature that she notes is high-temperature decontamination. As she points out: "The high-temperature sterilization cycle is essential for decontamination and simplifies routine cleaning."

When working with human samples and infectious agents, the safety of the laboratory personnel is a priority, Matthaiou explains. "We're working with airborne pathogens," she says. "We need to work in a safe environment, and BSL-2 [biosafety level 2] biosafety cabinets are essential for our safety." With Matthaiou's choice of Thermo Scientific's 1300 Series Class II, Type A2 Biological Safety Cabinets, she and her colleagues got safety, energy efficiency, and adjustable ergonomics. "We are working long hours under biosafety cabinets, and I want to make sure that my team is working in a safe and comfortable environment," she says. "The Thermo Scientific 1300 series has armrests, and the height is easy to adjust."

As Denzil Vaughn—vice president and general manager, Growth, Protection, and Separation at Thermo Fisher Scientific—says, "We are proud that Dr. Matthaiou selected our products to help further her research into resolving the threat of compromised immunity and pulmonary function."

Clearing pathogens

Having established a well-equipped lab, Matthaiou hopes to discover more about host–pathogen interactions in patients with cystic fibrosis who received a lung transplant. In this patient population, she wants to identify the factors that cause increased risk for infections. "How does this genetic defect and the transplant microenvironment impact the innate immune system?" she asks. "The reasons why these patients cannot clear the pathogens are not fully understood, and these infections drive the rejection of the transplant."

For now, Matthaiou works to understand the basic biology in these patients. "Later on," she says, "I hope to develop therapies for these patients."

Sponsored by

ThermoFisher SCIENTIFIC

PHOTO: © GORODENKOFF/SHUTTERSTOCK.COM

CONTENTS

5 AUGUST 2022
VOLUME 377
ISSUE 6606

SPECIAL SECTION

GRASS

INTRODUCTION

590 The unrecognized value of grass

PERSPECTIVE

592 The history and challenge of grassy biomes *By C. A. E. Strömberg and A. C. Staver*

REVIEWS

594 Ancient grasslands guide ambitious goals in grassland restoration *E. Buisson* et al.

599 Molecular, cellular, and developmental foundations of grass diversity *P. McSteen and E. A. Kellogg*

603 Grassland soil carbon sequestration: Current understanding, challenges, and solutions *Y. Bai and M. F. Cotrufo*

609 The planetary role of seagrass conservation *R. K. F. Unsworth* et al.

ON THE COVER

Marram grass, or beachgrass, grows on and stabilizes coastal sand dunes on Ireland's Dingle Peninsula. Grasses, whether terrestrial or submarine, tend to be undervalued but have influenced the trajectory of human history through their domestication as food staples, as well as natural ecosystems worldwide. If restored and conserved appropriately, grasslands can benefit climate change mitigation efforts. See the special section beginning on page 590. *Photo: Matt Anderson Photography/ Getty Images*

SEE ALSO NEWS STORY p. 568

Neptune seagrass (*Posidonia oceanica*) absorbs carbon dioxide and oxygenates coastal waters.

NEWS

IN BRIEF

560 News at a glance

IN DEPTH

562 Innovation bill will reshape science agencies
CHIPS and Science Act creates NSF tech directorate and boosts applied research *By J. Mervis*

563 Ordinary computer matches Google's quantum computer
High-profile claim of "quantum supremacy" fades *By A. Cho*

564 Ambitious bill leads to 40% cut in emissions, models show
But more action is needed to reach Biden's pledge to halve greenhouse gas emissions by 2030 *By E. Stokstad*

565 Mouse stem cells grown into embryo mimics
Bioreactor lets "embryoids" mature long enough for multiple organs to form *By M. Leslie*

566 Making broader coronavirus vaccines is a struggle
Efforts to protect against future variants or novel coronaviruses face funding constraints and other problems *By J. Cohen*

FEATURES

568 Fiery invasions
Around the world, flammable invasive grasses are increasing the risks of damaging wildfires *By W. Cornwall*
GRASS SECTION p. 590; PODCAST

INSIGHTS

POLICY FORUM

572 Transparency practices at the FDA: A barrier to global health
Data sharing among regulators must be "business as usual"
By Murray M. Lumpkin et al.

PERSPECTIVES

575 Twice the tolerance
A gut microbiota–derived antigen elicits distinct subsets of regulatory T cells to suppress inflammation in mice
By M. K. Shenoy and M. A. Koch
REPORT p. 660

PHOTO: SHANE GROSS

576 Unleashing spontaneity in a time crystal
Ordered patterns reoccur over time in an ultracold atomic gas trapped in light
By L. J. LeBlanc
REPORT p. 670

578 A regenerative niche for stem cells
Production of hyaluronic acid allows regenerative signaling in muscle stem cells after injury *By D. Gabellini*
REPORT p. 666

579 Mitochondria rescue cells from ischemic injury
Activation of a G protein–coupled receptor prevents cardiomyocyte death during ischemia *By S. Cadenas*
RESEARCH ARTICLE p. 621

580 Safe, selective, and scalable carbenes
The synthesis of reactive carbene intermediates is made simpler and safer
By M. S. West and S. A. L. Rousseaux
REPORT p. 649

BOOKS *ET AL.*

582 Why we do what we do
From regenerating sea slugs to self-medicating sheep, a biologist probes the origins and evolution of behavior
By R. Dunn

ILLUSTRATION: DAVIDE BONAZZI/SALZMAN ART

583 Inequality goes viral
Structural factors exacerbate the impact of viruses *By A. Nuriddin*

LETTERS

585 Ozone-reducing urban plants: Choose carefully
By P. Sicard et al.

585 Climate change threatens Pakistan's snow leopards
By U. A. Bhatti et al.

586 Primary forest loss in biodiverse Indian states
By S. Khadakkar

PRIZE ESSAY

588 Synchronizing neural rhythms
Personalized, noninvasive network-based neuromodulation aids impaired cognition
By R. M. G. Reinhart

RESEARCH

IN BRIEF

615 From *Science* and other journals

RESEARCH ARTICLES

618 Coronavirus
Mosaic RBD nanoparticles protect against challenge by diverse sarbecoviruses in animal models
A. A. Cohen et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABQ0839

619 Heart disease
Pathogenic variants damage cell composition and single-cell transcription in cardiomyopathies *D. Reichart* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABO1984

620 Development
The continuum of *Drosophila* embryonic development at single-cell resolution
D. Calderon et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABN5800

621 Medicine
Mitochondrial remodeling and ischemic protection by G protein–coupled receptor 35 agonists *G. A. Wyant* et al.
PERSPECTIVE p. 579

629 Plant science
Meiotic exit in *Arabidopsis* is driven by P-body–mediated inhibition of translation
A. Cairo et al.

635 Hybridization
Selection against admixture and gene regulatory divergence in a long-term primate field study *T. P. Vilgalys* et al.

642 Systems biology
Multimodal perception links cellular state to decision-making in single cells
B. A. Kramer et al.

REPORTS

649 Organic chemistry
Carbene reactivity from alkyl and aryl aldehydes *L. Zhang* et al.
PERSPECTIVE p. 580

654 Ocean oxygen
Changes in North Atlantic deep-water oxygenation across the Middle Pleistocene Transition *N. C. Thomas* et al.

660 Immunology
A conserved Bacteroidetes antigen induces anti-inflammatory intestinal T lymphocytes
D. Bousbaine et al.
PERSPECTIVE p. 575

666 Muscle repair
JMJD3 activated hyaluronan synthesis drives muscle regeneration in an inflammatory environment *K. Nakka* et al.
PERSPECTIVE p. 578

670 Quantum dynamics
Observation of a continuous time crystal *P. Kongkhambut* et al.
PERSPECTIVE p. 576

673 Crystallization
Synchronous assembly of chiral skeletal single-crystalline microvessels
O. Oki et al.

DEPARTMENTS

559 Editorial
A key time for UK–Europe science
By P. Mathieson

682 Working Life
Try, try again *By H. Huang*

Science Staff 558
Science Careers 679

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 © 2022 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): $2212; 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.

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 Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **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), L. Bryan Ray, Seth Thomas Scanlon (UK), H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Yuen Yiu **ASSOCIATE EDITORS** Michael A. Funk, Bianca Lopez, Yury V. Suleymanov, Ekeoma Uzogara **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, Shannon McMahon, Jerry Richardson, Hilary Stewart (UK), Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Maura Byrne, Alexander Kief, Ronmel Navas, Isabel Schnaidt, Qiyam Stewart, Brian White **RESEARCH & DATA ANALYST** Jessica L. 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** Shraddha Chakradhar, 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 **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Zack Savitsky, Viviana Flores, Katherine Irving **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 **GRAPHICS MANAGING EDITOR** Chris Bickel **MULTIMEDIA MANAGING PRODUCER** Kevin McLean **WEB STRATEGY MANAGER** Kara Estelle-Powers **DESIGN EDITOR** Chrystal Smith **DESIGNER** Christina Aycock **INTERACTIVE GRAPHICS EDITOR** Kelly Franklin **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR SCIENTIFIC ILLUSTRATOR** Valerie Altounian **SCIENTIFIC ILLUSTRATORS** Kellie Holoski, Ashley Mastin **SENIOR PODCAST PRODUCER** Sarah Crespi **VIDEO PRODUCER** Meagan Cantwell **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITORS** Charles Borst, Kaitlyn Dolan **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski **INTERN** Khafia Choudhary

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 ANALYSTS** Kurt Ennis, Maggie Clark **FINANCIAL ANALYST** Isacco Fusi **BUSINESS OPERATIONS ADMINISTRATOR** Taylor Fisher **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT & PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR PRODUCTION SPECIALIST** Kristin Wowk **PRODUCTION SPECIALISTS** Kelsey Cartelli, Audrey Diggs **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** Aaron Helmbrecht, Ashley Hylton, Mike Romano, Tori Velasquez, Jenna Voris **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst **PROJECT MANAGER** Melissa Collins

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** Caroline Breul, 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 **SYSTEMS & OPERATIONS ANALYST** Ben Teincuff **FULFILLMENT ANALYST** Aminta Reyes

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** Rie Rambelli (Tokyo), Miyuki Tani (Osaka)

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren **CONTRACT SUPPORT SPECIALIST** Michael Wheeler

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 Susan G. Amara
PRESIDENT Gilda A. Barabino
PRESIDENT-ELECT Keith Yamamoto
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER Sudip Parikh
BOARD Cynthia M. Beall, Ann Bostrom, Janine Austin Clayton, Kaye Husbands Fealing, Maria M. Klawe, Jane Maienschein, Robert B. Millard, Babak Parviz, William D. Provine, Juan S. Ramírez Lugo, Susan M. Rosenberg

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.*
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
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*
Malcolm Brenner, *Baylor Coll. of Med.*
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*
Ib Chorkendorff, *Denmark TU*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Virginia Cornish *Columbia U.*
Carolyn Coyne, *Duke U.*
Roberta Croce, *VU Amsterdam*
Christina Curtis, *Stanford U.*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *UNC*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Frans de Waal, *Emory U.*
Claude Desplan, *NYU*
Sandra Díaz, *U. Nacional de Córdoba*
Samuel Díaz-Muñoz, *UC Davis*
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*
Nader Engheta, *UPenn*
Karen Ersche, *U. of Cambridge*
Beate Escher, *UFZ & U. of Tübingen*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
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.*
Daniel Haber, *Mass. General Hos.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. Coll. London*
Carl-Philipp Heisenberg, *IST Austria*
Janet G. Hering, *Eawag*
Christoph Hess, *U. of Basel & U. of Cambridge*
Heather Hickman, *NIAID, NIH*
Hans Hilgenkamp, *U. of Twente*
Janneke Hille Ris Lambers, *ETH Zürich*
Kai-Uwe Hinrichs, *U. of Bremen*
Deirdre Hollingsworth, *U. of Oxford*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Gwyneth Ingram, *ENS Lyon*
Darrell Irvine, *MIT*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS & U. of Arizona*
Erich Jarvis, *Rockefeller U.*
Peter Jonas, *IST Austria*
Johanna Joyce, *U. de Lausanne*
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*
Laura Lackner, *Northwestern U.*
Gabriel Lander, *Scripps Res.* **(S)**
Mitchell A. Lazar, *UPenn*
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*
Ari Pekka Mähönen, *U. of Helsinki*
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*
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*
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*
Richard Smith, *UNC* **(S)**
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Northwestern U.*
Ira Tabas, *Columbia U.*
Eriko Takano, *U. of Manchester*
Patrick Tan, *Duke-NUS Med. School*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
Maria-Elena Torres Padilla, *Helmholtz Zentrum München*
Kimani Toussaint, *Brown U.*
Barbara Treutlein, *ETH Zürich*
Jason Tylianakis, *U. of Canterbury*
Wim van der Putten, *Netherlands Inst. of Ecology*
Matthew Vander Heiden, *MIT*
Ivo Vankelecom, *KU Leuven*
Judith Varner, *UC San Diego*
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.*
Li Wu, *Tsinghua U.*
Wei Xie, *Tsinghua U.*
Benjamin Youngblood, *St. Jude*
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

A key time for UK–Europe science

Peter Mathieson is the principal of the University of Edinburgh, Edinburgh, UK. principal@ed.ac.uk

The opening line of a recent *Financial Times* article put it best: "Relations between the UK and EU badly need a reset." Although the article was mostly about geopolitics, the disconnect also applies to science and the current uncertainty about whether the UK will remain an associated partner in European Union (EU) research programs such as Horizon Europe. In the post-Brexit era, and with a new UK Prime Minister to be named shortly, the UK and EU should be considering how best to maximize the potential of the numerous brilliant scientists, technicians, academics, and clinicians working in the universities and research institutes of all European countries, including the UK.

An unintended casualty of the UK's withdrawal from the EU was the country's ability to participate fully in the collaborative ecosystem of research and innovation that had evolved during the UK's 47-year membership in the organization. British universities, and the staff and students working and studying in them, have been highly desired partners. These interactions, which could continue if the UK is an associated country, are currently at great risk because of the complexities of the post-Brexit negotiations. Hopefully, after last month's tumult surrounding the resignation of Boris Johnson and the uncertainty about who will be the next Prime Minister, there will be a calm reappraisal of the risks of going it alone for both the UK and the European continent. What could emerge is a renewed effort to find a negotiated solution that allows the UK to continue to maximize its potential as a fully engaged contributor to European research and innovation.

To its credit, the current UK government has worked to promote and enhance the country's own scientific endeavors by increasing current funding and pledging more in the future, and by enhancing visa schemes and other immigration rules to continue to attract and retain talented individuals from home and abroad.

There has also been work on a so-called "Plan B" for the eventuality that the UK fails to associate with Horizon Europe, the EU's key research funding program with a budget of 95.5 billion euros. However, this would be a poor second best: Witness Switzerland's similar approach a few years ago, which left its researchers seriously isolated.

At the same time, the UK government recognizes that some collaborations can pose national security risks, particularly with institutions in countries whose governments the UK disagrees with. This has led to a clamor among a subset of politicians for more legislation to "control" or "manage" the country's universities. I have been involved in some of these discussions and have been heartened by many experts in UK security agencies and in parts of the UK government who recognize the value of these partnerships and of keeping risk mitigation measures proportionate and balanced. The UK could learn from experiences in Australia and the United States where similar recent legislation related to national security has stymied research and innovation. For example, Australian universities expressed concern about their autonomy and about limitation of their abilities to deliver societal benefits.

> **"It's time to untangle science from post-Brexit geopolitics so that European science can thrive."**

The university sector in the UK has warmly welcomed the creation of the Research Collaboration Advice Team, which will provide a single point of contact with UK government and security agencies. I and others will continue to work with them and with the security services to understand risk, disseminate good practices, and provide an early warning system when real dangers are recognized.

A mature, two-way relationship between government and the university sector is in everyone's best interests. A similar maturity should be brought to finding solutions for the challenges posed by securing the UK's association with Horizon Europe and with other EU programs.

Without a "reset" of UK–EU scientific relations, the "brain drain" from the UK—which has already started, with at least 19 researchers funded by the European Research Council recently relocating to EU countries to keep their funding—will become an avalanche. The role of the UK in the cohesion and productivity of European science will be the victim, with serious implications for global science capability. There is an opportunity for the UK government and the European Commission to prevent this now. It's time to untangle science from post-Brexit geopolitics so that European science can thrive.

–Peter Mathieson

10.1126/science.ade1620

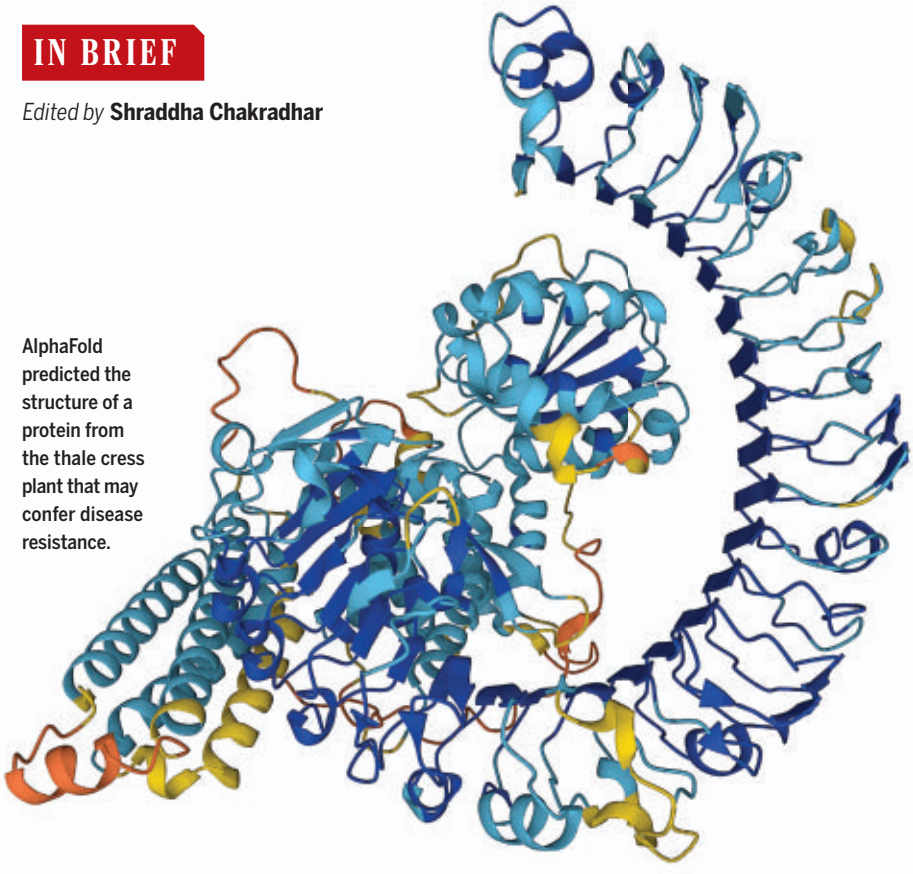
PHOTO: THE UNIVERSITY OF EDINBURGH

NEWS

> “It is embarrassing for the F.D.A. ... to have its employees go to a company that is a leading manufacturer of death.”
>
> **Micah Berman**, a public health law expert at Ohio State University, commenting in *The New York Times* after the chief of the office of science in the Food and Drug Administration’s Center for Tobacco Products moved to a job with tobacco giant Philip Morris International.

IN BRIEF

Edited by **Shraddha Chakradhar**



AlphaFold predicted the structure of a protein from the thale cress plant that may confer disease resistance.

IMAGE: DEEPMIND/EMBL-EBI

ARTIFICIAL INTELLIGENCE

AI reveals structures of nearly all proteins

In a potentially transformative event for drug development and biology studies, the artificial intelligence (AI) company DeepMind last week unveiled the likely structures of nearly all known proteins, more than 200 million in total, from organisms ranging from bacteria to humans. The structural bounty comes from AlphaFold, one of the new AI programs that have cracked the protein-folding problem and learned how to accurately derive 3D shapes of proteins from their amino acid sequences. DeepMind released some 350,000 predicted structures last year and AlphaFold has since continued to churn out new structures, taking about 10 to 20 seconds per protein, according to the company. The latest structures were released into an existing database through a partnership with the European Molecular Biology Laboratory’s European Bioinformatics Institute. “With this new addition of structures illuminating nearly the entire protein universe, we can expect more biological mysteries to be solved each day,” Eric Topol, director of the Scripps Research Translational Institute, tweeted about the achievement.

Boost for heart gene cures

HEART DISEASE | The British Heart Foundation will award £30 million ($36 million) over 5 years to an international team to develop genetic cures for some inherited heart diseases called cardiomyopathies. The group, dubbed CureHeart, won over three others shortlisted by the Big Beat Challenge, a competition launched in 2019 to fund transformative heart disease research. The team aims to use one-time injections of gene-editing tools to precisely correct or silence mutations that cause heart muscle cells to produce too little or a harmful form of a needed protein. These cardiomyopathies affect one in every 250 people, putting them at risk for heart attacks and heart failure; some will need a heart transplant. Within 5 years, CureHeart members in the United States, United Kingdom, and Singapore hope to develop one or more treatments to the point that companies will pick them up for clinical testing.

Racism takes a toll on memory

HEALTH INEQUITIES | Two studies presented this week at the Alzheimer’s Association International Conference in San Diego find people who experienced racism and discrimination have lower memory scores, a symptom of dementia. Systemic inequities such as limited health care access are likely responsible. In an analysis of nearly 1000 middle-aged Black, Latino, and white adults in the United States, researchers found past experiences of individual and structural racism, such as residential segregation, were correlated with poorer episodic memory—the ability to recall personal events in one’s life. Black individuals, who are more likely to have Alzheimer’s and other dementias, were the most affected. In the other study, of almost 450 Asian, Black, Latino, white, and multiracial people ages 90 or older, those who experienced discrimination throughout their lives had lower semantic memory—general world knowledge accumulated over time—compared with those who experienced little to no discrimination.

A World AIDS Day event in Indonesia in December 2021 was held to raise awareness of the disease.

INFECTIOUS DISEASE

Progress toward ending AIDS has stalled

The world's response to the HIV/AIDS pandemic is faltering badly in the face of declines in spending and the COVID-19 pandemic, according to an annual update from the Joint United Nations Programme on HIV/AIDS (UNAIDS) released last week. In a campaign announced in 2015 to "end AIDS as a public health threat" by 2030, UNAIDS set targets for 2025 that the new report finds are far from being met. Last year, 1.5 million people became infected with HIV, 1 million more than the 2025 target. Of the 38.4 million people living with the virus in 2021, 10 million are still not receiving lifesaving antiretroviral drugs, and last year saw the lowest number of new people starting treatment in a decade. Alarmingly, UNAIDS notes, 52% of infected children aren't being treated.

China replaces its CDC chief

INFECTIOUS DISEASE | Virologist George Gao, who helped forge China's response to the COVID-19 pandemic, was replaced as director of the Chinese Center for Disease Control and Prevention (CCDC) on 26 July. Internationally connected and respected, Gao, who led CCDC since 2017, was more outspoken than most Chinese scientists and promoted international cooperation. CCDC has said Gao, 60, was replaced because of his age, but observers suspect the change is part of a wider bureaucratic shake-up to strengthen political control over the agency. In an email to *Science*, Gao wrote it is "hard to say" what comes next in his career but hinted at a continuing interest in public health, noting that emerging diseases and climate change are "two of the most important issues mankind faces." Gao's successor is Shen Hongbing, a public health researcher and former president of Nanjing Medical University.

Mars mission gets revamped

SPACE | NASA and the European Space Agency announced last week they will make greater use of the Mars-based Perseverance rover—with helicopters as a backup—to retrieve samples from the planet, in a mission launching later this decade. Since 2021, Perseverance has been collecting samples of Mars rocks and sealing them in finger-size tubes. The agencies originally planned to send a new rover to gather the tubes for return, but are now betting Perseverance will be fit enough to deliver them in person. When a lander arrives around 2030, it will use an arm to load 31 tubes into a rocket that will launch and rendezvous with an orbiter waiting to return home to Earth. If Perseverance shows signs of faltering before then, operators can order it to drop its cargo. Then two small choppers brought by the lander, resembling Perseverance's hovering sidekick Ingenuity, will collect the dropped tubes. Either way, the most expensive half-kilogram of soil in history—given the $7 billion mission cost—will land in Utah in 2033.

Nepal counts record tiger numbers

CONSERVATION | Wild tiger numbers in Nepal have nearly tripled since 2009, the country announced last week. In 2010, with populations declining because of poaching and habitat loss, countries with tigers pledged to increase the number of wild animals from 3200 to more than 7000 by this year. In December 2021, a conservation group estimated fewer than 5600 across 13 Asian countries. Yet with a wildlife survey counting 355 of the big cats within its borders, Nepal has greatly exceeded its 2022 goal of 250 such animals. Tiger populations are also increasing in China and Thailand. Some of the apparent global increase may be due to improved survey methods and tools such as remote cameras. But in Nepal, the boost is largely thanks to conservationists' focus on habitat restoration, such as replanting forests as corridors for tigers. Nepal has also reduced tiger deaths by helping villagers better defend livestock from the cats with fences and compensating farmers when tigers kill their animals.

BY THE NUMBERS

Senate weighs in on research spending

The U.S. Senate last week released proposed 2023 budgets for key research agencies that are close to levels agreed on in June by the House of Representatives (see below, $ billions). But final numbers likely won't be adopted until after the November elections.

	2022	BIDEN 2023 REQUEST	HOUSE BILLS	SENATE BILLS
National Institutes of Health core	45	50.5	47.5	47
ARPA-Health	1	5	2.7	1
National Science Foundation	8.8	10.5	9.6	10.3
NASA science	7.6	8	7.9	8
Department of Energy science	7.5	7.8	8	8
NIST core labs	0.9	1	1	1
EPA science	0.7	0.9	0.9	0.9

CREDITS: (PHOTO) AGUNG KUNCAHYA B./XINHUA/GETTY IMAGES; (DATA) WHITE HOUSE; CONGRESS

IN DEPTH

U.S. SCIENCE POLICY

Innovation bill will reshape science agencies

CHIPS and Science Act creates NSF tech directorate and boosts applied research

By **Jeffrey Mervis**

A massive bill that Congress completed last week aims high: It envisions a 5-year, $280 billion investment to keep the United States ahead of China in a global competition for technological preeminence. But researchers won't see a surge of new money anytime soon. Instead, the CHIPS and Science Act, which President Joe Biden is expected to sign into law, will result in some of the biggest changes in U.S. innovation policy in more than a decade, with a greater emphasis on applied research and geographic diversity in funding, and closer scrutiny of foreign collaborations.

The 1054-page bill calls for more than doubling the annual budget of the National Science Foundation (NSF)—now $8.8 billion—over 5 years. It would also grow the $7.5 billion Office of Science at the Department of Energy (DOE) by 45% and boost the $850 million research account at the National Institute of Standards and Technology by 50%. But that money is "authorized," not committed, and congressional spending panels must decide each year whether to appropriate the additional dollars.

At the same time, the bill makes significant changes in how those agencies operate—directives that will go into effect immediately. For example, it gives NSF the legislative authority to create a technology directorate that would nurture innovations with commercial potential and social impact, adding to the agency's traditional mission of supporting basic research. The new directorate will focus on both emerging technologies, such as artificial intelligence and quantum information science, and societal challenges such as combating climate change and training a tech-savvy workforce.

Poorer, more rural states will benefit from a new requirement that both NSF and DOE increase the share of research spending going to institutions there to 20%. (NSF is now at roughly 13%, DOE doesn't track the figure.) The bill also directs the Department of Commerce to create regional technology centers designed to accelerate economic growth.

Other sections of the bill address growing concerns that China has been stealing or unfairly benefiting from research done in the United States. In general, these research security provisions tighten oversight of interactions between U.S. and foreign scientists, and with foreign governments.

For example, the legislation prohibits U.S.-based scientists with federal funding from participating in a foreign talent recruitment program sponsored by China and Russia and bans federal employees from participating in such programs from any country. In addition, institutions receiving federal research dollars must now provide research security training to faculty and staff. And NSF has been ordered to set up an independent forum to discuss how to strengthen research security in academic settings.

Although U.S. higher education organizations hailed passage of the CHIPS and Science Act, which appropriates $52 billion over 5 years to help the semiconductor industry, they are disappointed legislators rejected their pleas for at least $10 billion in funding to jump-start the bill's grand vision. (The only immediate increase for research agencies is a 5-year, $200 million appropriation for NSF to boost training in microelectronics.) Advocates fear the authorized funding could be an empty promise.

"We have been here before," says Peter McPherson, president of the 248-member Association of Public and Land-grant Universities, referring to the 2007 America COMPETES Act, which authorized billions of additional funding for NSF and DOE that never materialized. "The CHIPS Act must be step one in a process that ultimately includes Congress delivering the funding that will accomplish the goals of the legislation."

Higher education groups and others were relieved that legislators dropped several contentious research security provisions, including a proposed White House research security office that critics saw as a threat to legitimate research collaborations, a duplication of efforts already underway across the government, and an invitation to target Asian American scientists. But scientists are unhappy that immigration provisions designed to retain high-tech talent were removed. One would have made it easier for

PHOTO: TOM WILLIAMS/CQ-ROLL CALL, INC./GETTY IMAGES

Senate Majority Leader Chuck Schumer (D–NY, center) celebrates Senate passage of the CHIPS and Science Act on 27 July with other Senate backers.

foreign-born scientists to stay after earning an advanced degree from a U.S. institution. Another would have created a new visa category for foreign scientists setting up companies based on their research.

The bill went through numerous iterations—and names—since Senate Majority Leader Chuck Schumer (D–NY) proposed a $100 billion technology directorate at NSF in a November 2019 speech. In May 2020, he joined with Senator Todd Young (R–IN) to introduce the Endless Frontiers Act, named in homage to *Science, the Endless Frontier: A Report to the President on a Program for Postwar Scientific Research*, the 1945 report that led to NSF's creation 5 years later.

In June 2021, a greatly expanded version, rebranded the U.S. Innovation and Competition Act to reflect its increased emphasis on beating China, demanded by many Republicans, passed the Senate 68 to 32. In February, the House of Representatives revived the America COMPETES moniker for legislation adopted on a straight party line vote.

It had become the CHIPS and Science Act by the Senate's final 64-to-33 vote on 27 July, a tally that included 17 Republicans. "This is a Sputnik moment, only instead of Russia it's China, in which America realized that another rival power would get way ahead of us if we didn't pull out all the stops," Schumer said.

Some scientists had questioned Schumer's original vision to beef up NSF by creating a technology directorate focused on applied research that would be much larger than NSF's core programs. They also worried about its proposed quasi-independent status within the agency.

The final legislation dramatically shrinks the new directorate and simply makes it the seventh research directorate at NSF. Still, by 2027 legislators hope it will receive more than one-fifth of NSF's aspirational budget of $18.9 billion. Some academic leaders remain worried that NSF may favor the new directorate over existing programs if NSF's overall budget doesn't grow.

On 28 July, House Democrats withstood a last-ditch effort by Republican leaders to kill the bill, as 24 Republicans contributed to the final winning margin of 243 to 187. "With this legislation, we are ushering in a bold and prosperous future for American science and innovation," said Representative Eddie Bernice Johnson (D–TX), chair of the science committee and a key player in formulating the bill. ■

PHOTO: GOOGLE

COMPUTER SCIENCE

Ordinary computer matches Google's quantum computer

High-profile claim of "quantum supremacy" fades

By **Adrian Cho**

If the quantum computing era dawned 3 years ago, its rising sun may have ducked behind a cloud. In 2019, Google researchers claimed they had passed a milestone known as quantum supremacy when their quantum computer Sycamore performed in 200 seconds an abstruse calculation they said would tie up a supercomputer for 10,000 years. Now, scientists in China have done the computation in a few hours with ordinary processors. A supercomputer, they say, could beat Sycamore outright.

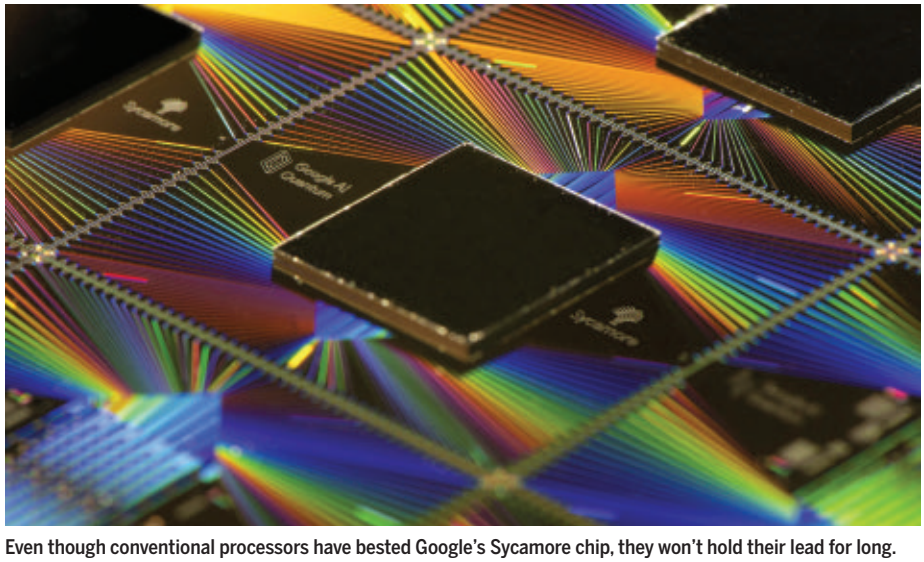
"I think they're right that if they had access to a big enough supercomputer, they could have simulated the … task in a matter of seconds," says Scott Aaronson, a computer scientist at the University of Texas, Austin. The advance takes a bit of the shine off Google's claim, says Greg Kuperberg, a mathematician at the University of California, Davis. "Getting to 300 feet from the summit is less exciting than getting to the summit."

Still, the promise of quantum computing remains undimmed, Kuperberg and others say. And Sergio Boixo, principal scientist for Google Quantum AI, said in an email the Google team knew its edge might not hold for very long. "In our 2019 paper, we said that classical algorithms would improve," he said. But, "we don't think this classical approach can keep up with quantum circuits in 2022 and beyond."

The "problem" Sycamore solved was designed to be hard for a conventional computer but as easy as possible for a quantum computer, which manipulates qubits that can be set to 0, 1, or—thanks to quantum mechanics—any combination of 0 and 1 at the same time. Together, Sycamore's 53 qubits, tiny resonating electrical circuits made of superconducting metal, can encode any number from 0 to 2^{53} (roughly 9 quadrillion)—or even all of them at once.

Starting with all the qubits set to 0, Google researchers applied to single qubits and pairs a random but fixed set of logical operations, or gates, over 20 cycles, then read out the qubits. Crudely speaking, quantum waves representing all possible outputs sloshed among the qubits, and the gates created interference that reinforced some outputs and canceled others. So some should have appeared with greater probability than others. Over millions of trials, a spiky output pattern emerged.

The Google researchers argued that simulating those interference effects would overwhelm even Summit, a supercomputer at Oak Ridge National Laboratory, which has 9216 central processing units and 27,648 faster graphic processing units (GPUs). Researchers with IBM, which developed Summit, quickly countered that if they exploited every bit of



Even though conventional processors have bested Google's Sycamore chip, they won't hold their lead for long.

hard drive available to the computer, it could handle the computation in a few days. Now, Pan Zhang, a statistical physicist at the Institute of Theoretical Physics at the Chinese Academy of Sciences, and colleagues have shown how to beat Sycamore in a paper in press at *Physical Review Letters*.

Following others, Zhang and colleagues recast the problem as a 3D mathematical array called a tensor network. It consisted of 20 layers, one for each cycle of gates, with each layer comprising 53 dots, one for each qubit. Lines connected the dots to represent the gates, with each gate encoded in a tensor—a 2D or 4D grid of complex numbers. Running the simulation then reduced to, essentially, multiplying all the tensors. "The advantage of the tensor network method is we can use many GPUs to do the computations in parallel," Zhang says.

Zhang and colleagues also relied on a key insight: Sycamore's computation was far from exact, so theirs didn't need to be either. Sycamore calculated the distribution of outputs with an estimated fidelity of 0.2%—just enough to distinguish the fingerprintlike spikiness from the noise in the circuitry. So Zhang's team traded accuracy for speed by cutting some lines in its network and eliminating the corresponding gates. Losing just eight lines made the computation 256 times faster while maintaining a fidelity of 0.37%.

The researchers calculated the output pattern for 1 million of the 9 quadrillion possible number strings, relying on an innovation of their own to obtain a truly random, representative set. The computation took 15 hours on 512 GPUs and yielded the telltale spiky output. "It's fair to say that the Google experiment has been simulated on a conventional computer," says Dominik Hangleiter, a quantum computer scientist at the University of Maryland, College Park. On a supercomputer, the computation would take a few dozen seconds, Zhang says—10 billion times faster than the Google team estimated.

The advance underscores the pitfalls of racing a quantum computer against a conventional one, researchers say. "There's an urgent need for better quantum supremacy experiments," Aaronson says. Zhang suggests a more practical approach: "We should find some real-world applications to demonstrate the quantum advantage."

Still, the Google demonstration was not just hype, researchers say. Sycamore required far fewer operations and less power than a supercomputer, Zhang notes. And if Sycamore had slightly higher fidelity, he says, his team's simulation couldn't have kept up. As Hangleiter puts it, "The Google experiment did what it was meant to do, start this race." ■

PHOTO: PAUL HENNESSY/SOPA IMAGES/LIGHTROCKET/GETTY IMAGES

U.S. CLIMATE POLICY

Ambitious bill leads to 40% cut in emissions, models show

But more action is needed to reach Biden's pledge to halve greenhouse gas emissions by 2030

By **Erik Stokstad**

For climate advocates in the United States, the past month felt like a roller coaster. In early July, negotiations in Congress on clean energy legislation of historic proportions collapsed, and the effort seemed doomed. But backroom talks continued and last week key senators suddenly announced an agreement on a $369 billion bill that would provide the most climate funding ever seen in the United States. "It was the best kept secret, potentially, in Washington history," says Leah Stokes, a political scientist at the University of California (UC), Santa Barbara.

The backers—Senate Majority Leader Chuck Schumer (D–NY) and Senator Joe Manchin (D–WV)—who had initially balked at the cost—announced that the draft bill would ensure U.S. carbon dioxide (CO_2) emissions would fall by 40% by 2030, compared with 2005.

Sponsors of the bill, which must still pass the full Senate and the House of Representatives, might be expected to oversell its impact. But energy and climate modelers have now scrutinized its 725 pages and concluded the 40% claim is about on target. They plugged major provisions, including subsidies for renewable energy and tax cuts for electric vehicles, as well as controversial incentives for the fossil fuel industry, into their models. Three models now agree that if the bill's provisions are carried out, U.S. greenhouse gas emissions would fall by perhaps 40% by 2030, although only part of that stems from the bill alone. One model also finds that the renewable energy subsidies will likely create 1.5 million jobs and prevent thousands of premature deaths from air pollution, especially in disadvantaged communities.

"It's a historic step, no doubt about it," says Marshall Shepherd, an atmospheric scientist at the University of Georgia and former head of the American Meteorological Society. "It really does a lot to enhance the transition to a renewable energy economy."

U.S. emissions have been falling by about 1% per year since 2005, when they peaked, largely because of replacing coal power with wind and solar, as well as natural gas, and rising fuel economy in light cars. But this pace is nowhere near fast enough to meet President Joe Biden's goal of a 50% to 52% cut in emissions by 2030 relative to 2005. Officials pledged that dramatic

A proposed bill would subsidize renewable energy sources, such as this solar farm in Florida, but also support some fossil fuel operations.

reduction as the U.S. contribution to the Paris accord's goal of holding global temperature rise to 1.5°C.

Biden's major effort had been the Build Back Better Act, which would have invested $560 billion in cutting greenhouse gases but died in the Senate after Manchin objected. The smaller new bill, called the Inflation Reduction Act of 2022, preserves much of the bang for clean energy, says energy systems expert Jesse Jenkins of Princeton University's Rapid Energy Policy Evaluation and Analysis Toolkit Project, which runs one of the models. "I think [Senate staff] did a miraculous job," he says. In particular, the bill provides subsidies to expand renewable energy and lure consumers to buy electric vehicles, solar panels, and climate-friendly home heat pumps.

To evaluate the climate impacts of the legislation, Jenkins and other modelers simulate the entire U.S. energy system, from the smallest electric vehicles to nuclear plants, and add the proposed policies to see how they impact CO_2 emissions. Scientists also fold in results from other models that focus on factors such as the impact of agricultural policies on two other causes of greenhouse warming: methane emissions from livestock and nitrous oxide released from fertilized fields. Modelers put everything together to forecast emissions trends, says modeler Ben King of the Rhodium Group, an independent research firm.

Just a day after the bill was released, Rhodium posted preliminary estimates on its website. The topline result: a 31% to 44% reduction in greenhouse gas emissions from 2005. Compared with current policies, that's an additional drop of 7 to 9 percentage points. Variables such as the price of natural gas account for much of the uncertainty: If gas prices drop, utilities might favor gas over renewable power, slowing the decline in carbon emissions.

This week, the think tank Energy Innovation narrowed the range, forecasting emissions reductions of 38% to 41%, with 13% to 17% from the bill alone. And the Princeton model estimated about a 42% reduction, with 15% from the bill itself.

All the analyses find the two most important factors driving down emissions are clean electricity tax credits—which the bill provides for at least a decade—and expanded tax credits for both new and used electric vehicles. The subsidies will help utilities install more capacity from wind farms and solar panels and help keep nuclear power plants financially viable as they face competition from cheap natural gas. Previous analyses had also pointed to green electricity generation and transportation as crucial to reducing emissions (*Science*, 27 May, p. 922).

Models can have difficulty predicting human behavior, cautions economist Meredith Fowlie of UC Berkeley. "I wouldn't believe any one projected number, but [key] models agree in a qualitative sense that this is going to bend the trajectory," she says.

Other provisions of the proposed bill could eventually lead to further CO_2 reductions, such as investment in technologies that directly remove carbon from the atmosphere and capture it from fossil fuel plants.

The bill also includes some climate-unfriendly provisions, apparently added at Manchin's request. It requires the federal government to offer several lease sales of offshore oil and gas resources, with more on the table if public lands are opened to renewable energy efforts like wind farms. The leases could boost oil and gas production from federal lands by an extra 50 million tons per year in 2030, according to Energy Innovation. Overall, however, climate wins out, analysts say: For each additional ton of CO_2 from fossil fuels, other provisions of the bill would reduce emissions by 24 tons.

> **"The United States is really going to be a climate leader globally if we can get this bill over the finish line."**
>
> **Leah Stokes**, University of California, Santa Barbara

The bill must still pass the Senate, where Democrats need every possible vote in their party, and then it will go back to the House. Stokes, who advised Democrats on the bill, says she's hopeful the measure will be on Biden's desk by mid-August. "The United States is really going to be a climate leader globally if we can get this bill over the finish line."

The measure won't be enough, however, for the United States to reach its Paris goal of a 50% greenhouse emissions reduction by 2030. For that, more federal regulation and state action will be necessary, King and others say. "It's all hands on deck," says energy and climate modeler John Bistline of the Electric Power Research Institute.

The ultimate—and necessary—goal is cutting U.S. emissions to zero, says Emily Grubert, a civil engineer and environmental sociologist at the University of Notre Dame. "People keep talking about this as the biggest climate investment in a generation. I can only say—I hope not." ■

DEVELOPMENTAL BIOLOGY

Mouse stem cells grown into embryo mimics

Bioreactor lets "embryoids" mature long enough for multiple organs to form

By **Mitch Leslie**

What happens in embryonic development is one of nature's best guarded secrets, unfolding deep in the mother's body. Now, researchers have opened a new window on the process. They've made artificial mouse embryos from stem cells—no sperm or eggs required—and used an innovative bioreactor to nurture their creations for longer than any previous embryo models. The simulated embryos developed anatomy that matched the real thing and "very impressive similarities at the cellular level. The right cells arise at the right time," says stem cell biologist Niels Geijsen of the Leiden University Medical Center, who was not involved in the work.

The feat, reported this week in *Cell*, may allow biologists to delve deeper into developmental mechanisms and better understand what goes wrong in birth defects. And the team's leader, stem cell biologist Jacob Hanna of the Weizmann Institute of Science, says that next, he hopes to do the same with comparable human stem cells.

Researchers have already reprised parts of early development with embryo mimics made from an assortment of mouse or human stem cells, including embryonic stem (ES) cells, which are derived from normal embryos and can form all of a body's tissues. They've mimicked the blastocyst, the simple developmental stage that implants in the uterus, and recreated gastrulation, when embryos become multilayered. These simulated embryos hit a developmental wall, however. Their cells begin to specialize but do not coalesce into organs.

One obstacle has been keeping the ersatz embryos alive for more than a few days. Last year, Hanna and colleagues unveiled a nurturing procedure that allowed them to grow standard mouse embryos outside of the mother's body for a record 11 days. (Typical mouse gestation is about 20 days.)

A key step involves placing the embryos in an incubator outfitted with a Ferris wheel–like device, which rotates the embryos inside bottles of liquid filled with nutrients and growth factors. The setup enables the team to precisely control growth conditions such as oxygen levels.

Those embryos came from fertilized mouse eggs, however. To determine whether the same procedure would allow stem cells to transform into full-fledged embryos, Hanna's team mingled basic mouse ES cells with ES cell lineages genetically altered to spawn tissues outside the embryo that shape and support its growth. After initially rearing the cell congregations on culture plates, the team shifted them to rotating bottles on the fifth day.

By the eighth day, the "embryoids" were very similar to 8.5-day-old natural embryos and boasted a beating heart, distinct head and tail ends, the blocklike segments that become skeletal muscles, a developing brain and spinal cord, and the beginnings of other organs. The researchers also measured gene activity in more than 40,000 embryoid cells, finding all of the expected cell types in the correct locations, Hanna says.

For unknown reasons, the artificial embryos stalled at the eighth day of development. The researchers hope to overcome this barrier and extend development even further. Still, stem cell–derived embryos have an advantage over normal mouse embryos for research because the cells are available in larger numbers and scientists can more easily manipulate them, says stem cell biologist Nicolas Rivron of the Institute of Molecular Biotechnology of the Austrian Academy of Sciences.

The current procedure for making the simulated embryos fails most of the time—less than 1% of the initial cell aggregations form embryo mimics. But, Hanna notes, "The advantage of this technique is that we can make millions of aggregates in one batch."

Achieving the same feat with human ES cells might avoid some of the controversies of research on human embryos. "This is providing an ethical and technical alternative to the use of embryos," Rivron says.

Hanna has co-founded a company that will investigate whether the approach will work with human induced pluripotent stem cells, which are derived from adult cells rather than embryos. Cells and tissues in an embryo release factors that orchestrate the correct development of their neighbors. So growing stem cells into artificial embryos first may provide a better way of producing cell types that can be transplanted to treat human diseases. It is "more physiological," Hanna says. ■

COVID-19

Making broader coronavirus vaccines is a struggle

Efforts to protect against future variants or novel coronaviruses face funding constraints and other problems

By **Jon Cohen**

There's a new call from the White House to develop vaccines that might protect against future SARS-CoV-2 mutants or even unknown coronaviruses. "The vaccines we have are terrific, but we can do better than terrific," Ashish Jha, White House COVID-19 response coordinator, said at a vaccine summit last week that gathered researchers, companies, and government officials. "Ultimately, we need vaccines that can protect us no matter what Mother Nature throws at us."

But no specific funding request to Congress was revealed at the summit, or any concrete plans, so vaccine developers and the public shouldn't expect a second Operation Warp Speed, the U.S. government's multibillion-dollar crash program to develop the first COVID-19 vaccines. And the scientific, logistical, and regulatory hurdles for any next-generation vaccines are higher.

> **"The sense of urgency is completely gone."**
> **Florian Krammer**, Icahn School of Medicine at Mount Sinai

Operation Warp Speed proved it was possible to race from a newly identified virus to safe and effective immunizations in just 11 months—many times faster than ever before. Today, there are a few dozen fledgling efforts to create vaccines that could thwart future SARS-CoV-2 mutants or offer even broader protection, extending to unknown coronaviruses that have yet to jump into humans.

But only a single candidate, developed by the U.S. Army, has made it into a phase 1 clinical trial. "We want to start clinical trials tomorrow, but there are lots of barriers to getting there," says Yale University immunologist Akiko Iwasaki, who has a coronavirus vaccine candidate that's administered as a nasal spray.

For starters, funding for these broader vaccines remains far tighter than in the Warp Speed days: The Coalition for Epidemic Preparedness Innovations (CEPI) has invested a more modest $200 million in 11 efforts run by small companies and academics, and the U.S. National Institute of Allergy and Infectious Diseases (NIAID) has committed just $43 million in four pancoronavirus vaccine programs. Efforts also face a dearth of materials needed to make vaccines, a shortage of nonhuman primates on which to test candidates, and uncertainty about how to assess new products in populations that already have immune response to SARS-CoV-2.

White House summits aside, some scientists contend there's a deeper barrier. "The sense of urgency is completely gone," says Florian Krammer, a virologist at the Icahn School of Medicine at Mount Sinai.

That's understandable, says Moncef Slaoui, the scientific leader of Operation Warp Speed and a veteran vaccine developer. Even though new SARS-CoV-2 variants are eroding the ability of today's COVID-19 vaccines to block infection and prevent symptomatic disease, those vaccines are still preventing severe illness and death. "Current vaccines are effectively able to deal with the pandemic, because the No. 1 priority is mortality and morbidity," Slaoui says. "Pancoronavirus vaccines, whatever definition you use for them, are about preparedness, rather than dealing with the actual pandemic."

The "pan" in potential next generation coronavirus vaccines is often in the eyes of the beholder, as the projects underway have a variety of aims. The most modest, but still ambitious, goal is to avoid racing to create specific boosters for the latest SARS-CoV-2 variant and instead to develop vaccines that can reliably ward off severe disease from any future mutants.

Lawrence Corey of the Fred Hutchinson Cancer Research Center, who co-led the clinical trials network for Operation Warp Speed, wants to go further by developing COVID-19 vaccines that reduce not just severe disease, but the risk of infection and transmission of all SARS-CoV-2 spinoffs. "Do we really want to have 90,000 COVID deaths [in the United States] a year?" Corey asks.

And he calls for more government backing so as to not squander the momentum of the field. "There are plenty of ideas," he says. "What's not forthcoming is the commitment."

The next level of coronavirus vaccines shoots for broad protection against sarbecoviruses, the grouping that includes SARS-CoV-2 and SARS-CoV-1—the cause of the outbreak of severe acute respiratory syndrome 20 years ago—and viral relatives in bats and pangolins that could spark the next human outbreak. A vaccine with even more "breadth" would thwart the beta genus of coronaviruses, which includes sarbecoviruses, merbecoviruses like the one that causes Middle East respiratory syndrome, and two that now trigger the common cold. The ultimate vaccine would work against all four genera in the coronavirus family.

All of the work in this area is "really quite early," says Melanie Saville, who heads vaccine R&D at CEPI. "I would classify this as high-risk, high-reward research."

Kayvon Modjarrad and colleagues at the Walter Reed Army Institute of Research (WRAIR) have the only pancoronavirus vaccine candidate to reach clinical testing so far. Aimed at SARS-CoV-2 variants, it uses the same coronavirus surface protein, spike, as many existing shots but tries to improve the way it is presented to the immune system. The WRAIR candidate contains several copies of spike bound to ferritin, a protein that normally carries iron around the blood. Receptors on the surface of antibody-making B cells can then "cross-link" to these closely arrayed spikes, which theoretically leads to production of more powerful antibodies. In test tube studies, the vaccine "neutralized" a broad range of SARS-CoV-2 variants. WRAIR plans to publish data soon from its phase 1 study, but declined *Science*'s request to interview investigators.

The limited availability of animals to test experimental vaccines has stalled other projects. Andrew Ward, a structural biologist at Scripps Research, says there is "enormous competition" for monkeys and the best mouse model systems. Operation Warp Speed allocated research animals to the most promising vaccine candidates, but now efforts have just become a "free-for-all, not a coordinated Manhattan Project," says Ward, who, as a result, has turned most of his attention from coronavirus to paninfluenza vaccines.

Iwasaki also has had trouble getting monkeys for her studies of nasal sprays, which she contends might be able to broaden protection by stimulating production of mucosal antibodies with twice as many arms to bind viruses as the Y-shaped ones triggered by injection. "If there was a government-organized effort to help us, it would have gone much faster," she says.

NIAID only funds two investigator-initiated grants for pancoronavirus vaccine research. One went to Lbachir BenMohamed, an immunologist at the University of California, Irvine, who has a 5-year, $3.6 million grant and had hoped to launch a clinical trial this year of a pan-sarbecovirus vaccine. But his team, too, has had to wait for access to animal models, which it needs to select the most promising vaccine candidate. He is now looking at 2023—if his team can overcome another challenge.

BenMohamed's team first analyzed sarbecoviruses that have infected humans, camels, bats, minks, and pangolins for shared genetic sequences. They made proteins that reflect the conserved regions and linked them together. Like several existing vaccines, their vaccine encodes these conserved proteins in messenger RNA and relies on the body to turn that code back into protein. But a shortage of the lipid shells needed to enclose the RNA has slowed progress.

The U.S. Army has a phase 1 trial of a pancoronavirus vaccine underway.

Biochemist Pamela Bjorkman of the California Institute of Technology says her team's pan-sarbecovirus vaccine candidate likely won't enter the clinic until perhaps 2024. That team has plucked a critical portion of spike from eight different sarbecoviruses—the so-called receptor-binding domain—and stitched them together into a "mosaic." Her team has faced manufacturing challenges, and only received substantial support recently, with CEPI announcing on 5 July a grant of up to $30 million to take the candidate through phase 1 trials. "This is taking longer than we'd hoped," Bjorkman says.

To prove their worth, pancoronavirus vaccines will have to travel a much rougher road than the first COVID-19 shots. The people in trials of the first COVID-19 vaccines had no specific immunity to SARS-CoV-2, making it straightforward to assess whether the shots provided protection. Today, most everyone has been vaccinated, infected with the virus, or both. Even the lowest hurdle for a pancoronavirus vaccine—proof of protection against all known SARS-CoV-2 variants—will be difficult to establish, predicts Barney Graham, who long worked on pancoronavirus vaccines at NIAID.

Graham, who is now at Morehouse College, also notes that the immune response to any new SARS-CoV-2 vaccine will be skewed by the immune system's memory of the first viral proteins it encountered, whether through vaccination or infection with one of the many variants that have circulated. So assessing a new vaccine's protection may require trials in people who have "no competition in the immune system"—which in many countries would now mean infants. "There are a lot of fundamental biological and immunology questions to answer," Graham says. "It's not going to go as fast as before."

Exactly how these vaccines would be used—given as periodic boosters that top up existing SARS-CoV-2 immunity and add breadth or held in reserve, for when a new coronavirus surfaces—also remains a big question. And developing a panvaccine as insurance is a risky proposition, Graham says. A candidate vetted in clinical trials and sitting on the shelf might only have limited efficacy against, say, a SARS-CoV-3. Graham says if he had to make a choice with limited resources, he'd prefer investing in having "knowledge on the shelf" to quickly make a pathogen-specific vaccine rather than paying to have a panvaccine at the ready that could only "hold the fort until you get the real thing a few months later."

Slaoui says if we "disrupt the economic model" for R&D and take a Warp Speed approach, pancoronavirus vaccine development could be "up to 10 times faster." But he contends the best bang for buck, especially when it comes to thwarting an entirely novel coronavirus, will come from investing heavily in building vaccine manufacturing plants that can quickly make vetted candidates against a new threat. "The day somebody comes up with a panvaccine that actually works, we should celebrate it," Slaoui says. "But we will only know that when another pandemic comes and we try it." ■

PHOTO: MIKE WALTERS/WALTER REED ARMY INSTITUTE OF RESEARCH

FIERY INVASIONS

PHOTO: NTPFES

Around the world, flammable invasive grasses are increasing the risks of damaging wildfires

By **Warren Cornwall**

For decades, eastern Oregon's scablands—rocky patches of open terrain—were a refuge for people fighting wildfires in the surrounding forests. The thin soil and sparse vegetation offered little fuel for the flames, creating an oasis from which firefighters could operate and a barrier that could help halt a fire's spread.

That all changed in 2015. After lightning sparked a fire near a steep-walled canyon, the blaze unexpectedly raced across scablands so quickly that firefighters struggled to catch up. In the end, the Corner Creek Fire scorched more than 11,000 hectares. And Jeff Priest, who has spent more than 2 decades fighting fires in Oregon for the U.S. Forest Service (USFS), realized he had a new problem on his hands: the arrival of an invasive, shin-high grass known as *Ventenata dubia*. The plant created shaggy golden carpets of dry foliage, transforming once fire-resistant scablands into flame-friendly corridors.

"We knew it was coming," Priest says about the annual commonly called wiregrass, which is native to countries sur-

Australian cattle ranchers planted gamba grass from Africa to create forage. It forms dense walls of vegetation that have become a major fire hazard.

rounding the Mediterranean. "But all of a sudden, it was there."

Ventenata's spread into the forests of the northwestern United States is just the latest chapter in a phenomenon reshaping ecosystems—and wildfire—around the globe. In northern Australia, invasive gamba grass from Africa fuels intense blazes that rip through eucalyptus groves. In Brazil, molasses grass from Africa turns vast swaths of the savanna known as the Cerrado into fire-prone grassland. In the western United States, two Old World grasses are creating ecological mayhem: Buffelgrass feeds fires in the Sonoran Desert that torch iconic saguaro cacti, while blaze-tolerant cheatgrass crowds out native sagebrush in the high desert known as the Great Basin.

Even as catastrophic wildfires that roar through towering treetops capture the public's attention, ecologists have been paying increasing attention to this less conspicuous trend: how seemingly modest nonnative grasses are allying with fire to eat away at dry forest and savanna ecosystems.

These invasive grasses can hijack fire to create a self-reinforcing cycle, explains Carla D'Antonio, an ecologist at the University of California, Santa Barbara, who has studied the phenomenon for more than 3 decades in Hawaii and California. Once established, the grasses help fuel blazes that kill and suppress less fire-tolerant native plants, opening up new territory for the invaders to colonize—catalyzing yet more fire. In a short time, land that was once shrubland, savanna, or dry forest is locked into being a grassland. "It's that trigger of grass and fire that sets the system off in some undesirable direction," D'Antonio says.

These grass invasions are now threatening native plants and the animals that rely on them, reshuffling nutrients in the soil and the ability of ecosystems to store planet-warming carbon, and disrupting efforts to use fire to benefit the native flora. And once the invaders take hold, ecologists say it's tough to break their grip. In such places, the future promises to be hot, smoky, and full of grass.

GRASSES AND FIRE have been intertwined since before humans walked the planet. Millions of years ago in southwest Africa and Asia, a dramatic rise in wildfires went hand in hand with the emergence of vast grasslands, researchers say. There and elsewhere, changes in weather patterns—particularly the emergence of a dry season—helped grasses spread. Select species developed a new way to photosynthesize that gave them an advantage in hotter, drier environments.

Evidence suggests some of these grasses evolved to thrive with fire, says Allison Karp, a paleoecologist and postdoctoral researcher at Yale University. In ancient sediments pulled from the Bay of Bengal near India's eastern coast, for example, she found carbon isotopes trapped in ancient plant waxes that indicate grasses became widespread on the subcontinent about 7 million years ago. During the same period, according to the sediments, molecules tied to wildfires increased 10-fold, and the chemical traces suggest grasses played an outsize role in fueling those blazes.

Today, some of the most problematic invasive grasses seem built to burn. They grow and dry quickly, creating abundant fuel each year. Certain species have leaves filled with oily tannins—chemicals that slow the decay of dead leaves, making it easier for them to ignite. One species, molasses grass, is coated with a residue that enables it to catch fire even while green.

Although fire can kill invasive grasses, they often bounce back quickly, enabling them to outcompete charred competitors, including native grasses that didn't evolve with frequent, intense fires. The lack of woody trunks and branches means grass seedlings start to photosynthesize before trees or shrubs put out leaves. Some grasses resprout from rootlike stems that grow underground, insulated from the flames.

Such adaptability has helped many grasses naturally expand their ranges. But in recent times, humans have accelerated that process by scattering grass seeds far from their native habitats, sometimes by accident and sometimes intentionally, to feed livestock, control erosion, and decorate gardens. "The [grass] invasions in the last 100 years or so are just a radical example of a speeded-up process that's been happening over millennia," says Dave Richardson, an ecologist and invasive plant expert at South Africa's Stellenbosch University.

Southern Africa is a disproportionate source of the grasses that have invaded other parts of the world, Richardson has found. Grasses evolved there to take advantage of frequent disturbances, such as fire and grazing by herds of wildlife, making them tough competitors in new habitats. And once they gain a roothold outside Africa, fire often follows. In the United States, when scientists compared fire behavior in areas invaded by fire-prone grasses with nearby uninvaded areas, they found six different grasses were tied to as much as a 150% increase in fire frequency, according to a 2019 report in the *Proceedings of the National Academy of Sciences*.

In northern Australia, the arrival of gamba grass has provided a textbook example of this process. In the 1980s, the Australian government promoted planting the African grass as forage for cattle. But ecologists soon warned of the dangers it posed to the nation's tropical savannas, a mix of sparse grasses and eucalyptus trees that cover one-quarter of the continent. The original ecosystem evolved with frequent, low-level fires, including ones set by Indigenous Aboriginal people to create open savannas that improved hunting and provided habitats for specific plants and animals.

But gamba grass formed dense walls of vegetation, reaching 4 meters tall, that transformed fire behavior. The grass burned four times more intensely than native vegetation during experimental burns, researchers found. The flames ran so high and hot that they had to abandon a common gauge of fire behavior—measuring the highest scorched leaves in a tree—because even the highest leaves in gamba-infested sites were singed.

In the early 2000s, several Australian states reversed course, restricting the use of gamba grass. But in many places it was too late. By then the grass covered more than 15,000 square kilometers. Researchers fear it could ultimately spread through much of the country's 2 million square kilometers of tropical savanna.

OREGON, meanwhile, shows how a single invasive grass species can alter both rangelands and forests. In 2015, reports of *Ventenata*-driven fires reached Becky Kerns, an ecologist at USFS's Pacific Northwest Research Station. Kerns had given little thought to the plant, in part because other invasive grasses were already causing headaches in the region. Cheatgrass, for example, was spreading in parts of Oregon. But the harsh scablands had proved inhospitable to that invader. *Ventenata*, Kerns says, "was really a game changer."

A year after the Corner Creek Fire, Kerns and a group of Oregon State University graduate students began to examine what wiregrass might mean for the region. Their findings were alarming. In scablands overtaken by *Ventenata*, they found oceans of grass that create fuel loads 50 times greater than in areas free of the species. Compared with cheatgrass, wiregrass can colonize cooler, higher elevation locations and take root in thinner soils.

In the process it appears to be crowding out native plants, such as fire-intolerant sagebrush, that support the local wildlife. And it threatens rare endemic plants found in rocky scablands, such as Spalding's catchfly, a federally protected perennial with pale pink, trumpet-shaped flowers.

Unlike some invasive grasses, *Ventenata* doesn't appear to need disturbances such as fire to spread in some parts of eastern Oregon. In study plots on native prairies, the grass advanced just as quickly over unburned land covered with native bunchgrass as it did through plots burned for a long-term experiment. In both cases, in 10 years *Ventenata* spread from approximately 10% of the test plots to nearly 60%.

In forests, by contrast, fires do appear to help spread the grass, Kerns says. An examination of a decade of fires in the region revealed that more severely burned areas were more likely to be invaded. One explanation for that pattern, researchers say, is that although *Ventenata* doesn't fare well in dense shade, it can encroach on forest edges, where it fuels fires that clear parts of the overstory, letting in more sunlight. Computer simulations suggest that cycle could gradually shrink forests, Kerns says, like waves eroding a beach.

That scenario was a revelation, and not a welcome one. Historically, Kerns says, land managers in the western United States have thought of invasive grasses as a problem for wide-open rangelands. But modeling suggests wiregrass could creep into many forests in eastern Oregon, as well as those in dry, higher mountains as far south as Arizona and New Mexico. "*Ventenata* has really challenged a lot of our notions about invasive grasses in the West," she says.

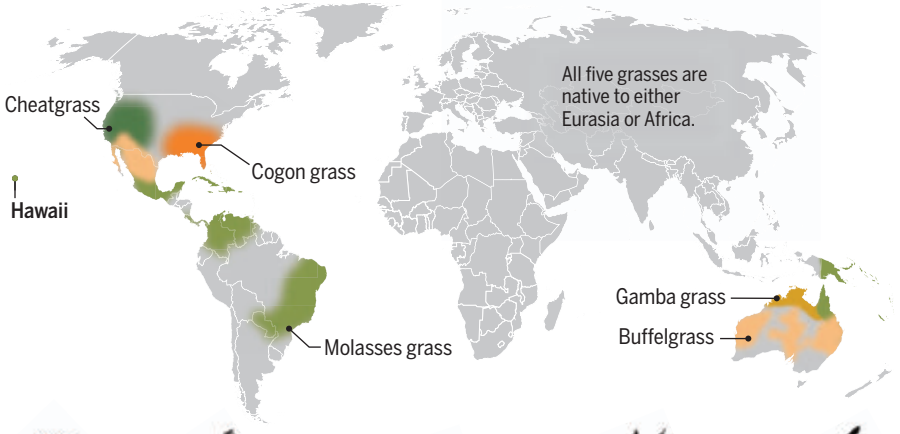
IN MANY LOCATIONS, scientists are just beginning to document the impact of invasive grasses. In Hawaii, however, D'Antonio has been closely observing the interlopers since the early 1990s. And her ringside seat has enabled her to see how the effects can echo through a landscape in complex and surprising ways.

In Hawaii's Volcanoes National Park, for example, she has tracked how a handful of invasive grasses—many introduced by cattle ranchers—have altered environments once rarely touched by fire. Studying places that first burned in 1970, D'Antonio found that in forests dominated by the native ʻŌhiʻa tree, relatively shade-tolerant beardgrass from South America took root first. But if the beardgrass later fuels a fire, she says, then a "second invader, molasses grass, pours in." Native grasses simply can't compete, she adds. Whereas invasive grasses produce large numbers of airborne seeds, for example, a native species called Kawelu, or love grass, produces just a few seeds that drop nearby. Compared with molasses grass, she says, "It's a total wimp."

Molasses grass can also alter soil nutrient cycles in ways that ultimately benefit invasive species, D'Antonio has found. As the grass leaves decomposed, for instance, they led to a bounty of soil nitrogen that helped the grass. After 16 years, however, nitrogen returned to preinvasion levels as the nutrient leached out of the soil. Rather than pave the way for a return of native plants, however, the leaching opened the door to an invasive tree known, appropriately, as the fire tree. That tree, in turn, again increased soil nitrogen levels, which revived the molasses grass. "The situation there gets more and more grim," D'Antonio says.

Invasion hot spots

Fire-friendly grasses have invaded new habitats around the world. Five species are considered among the most problematic grasses, threatening to transform entire ecosystems.



Cheatgrass
Taking over high-desert habitat that's home to endangered sage grouse and sagebrush, a shrub sensitive to frequent, intense fires.

Cogon grass
In the southeastern United States, this grass chokes out native plants and burns at high temperatures, stoking intense fires that could damage forests.

Gamba grass
Planted to feed cattle, it creates dense walls of vegetation up to 4 meters tall. Some Australian states have banned plantings because of fire risks.

Molasses grass
Taking over parts of Hawaii and Brazil's tropical savanna, the Cerrado. Grows in dense mats, and is named after oily substance on leaves.

Buffelgrass
Fuels fires threatening Sonoran desert habitat, including iconic saguaro cacti. Also spreading in central Australia.

CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) EDDMAPS. 2022. UGA CENTER FOR INVASIVE SPECIES AND ECOSYSTEM HEALTH; EDDMAPS. 2014. UGA CENTER FOR INVASIVE SPECIES AND ECOSYSTEM HEALTH; MELANIE BROWN/OSU/CBI; K. S. KUNTH (1829). *REVISION DES GRAMINEES* PART 1D: 163; DEPARTMENT OF ENVIRONMENT, PARKS AND WATER SECURITY; © 2022 CABI; V. M. MARSHALL *ET AL*., *JOURNAL OF ARID ENVIRONMENTS* 78 (2012) 1–12; R. VAN KLINKEN *ET AL*., *ENVIRONMENTAL MANAGEMENT* 2011

Experiments in plots of longleaf pine in Florida suggest a combination of drought and invasive cogon grass can amplify wildfires and kill the trees.

These new grasses could also scramble how much carbon is stored in vegetation and soils. In Hawaii, D'Antonio found that in unburned areas where native ʻŌhiʻa trees remained, the plants held twice the carbon of burned areas taken over by grass—even 25 years after a fire.

Drought could heighten the impact of the grass-fire cycle, suggest studies by ecologist Luke Flory of the University of Florida. He has run field experiments that simulate how drought affects longleaf pine forests invaded by cogon grass from Asia, which can grow in dense, waist-high thickets topped with fluffy seed heads.

Flory's team grew pines on small plots in Florida; many were infested with cogon grass whereas others only held native plants. Some of these miniforests were covered by canopies that limited rain to simulate drought.

After 6 years, trees in the dry plots averaged about 3 meters tall, nearly 1.5 meters shorter than those in wetter plots. The researchers then set fire to all of the plots. In those infested with cogon grass, the fires burned hotter and the flames rose higher than in plots without the grass. And in the dry, grassy plots, the combination of shorter trees and higher flames proved "highly problematic," Flory says, with nearly half of the trees dying within 1 month of being burned. But only 10% of trees in the wetter sites died, regardless of whether cogon grass was present.

Flory is reluctant to sound the alarm too loudly, because there are still plenty of unknowns about how drought, fire, and cogon grass will interact in the wild. But the experimental results so far are worrying, he says. "None of it's good. … It's just how much evidence we have that it's really bad."

SUCH EXPERIMENTS have helped focus attention on a practical matter: What can land managers do about the tenacious invaders once they arrive?

Eradicating the grasses is often not possible once they have reached high densities, researchers say. But preventing them from completely transforming ecosystems might be an option in some places. In Volcanoes National Park, for example, park officials are trying to rehabilitate shrubby or forested habitats that have burned by replanting them with more fire-hardy native species such as māmane, which can grow to 15 meters tall. The māmane could restore lost habitats for birds and insects, while being less vulnerable to grass-fueled fires, says Sierra McDaniel, a botanist who leads the park's natural resource program. "We are accepting these grasses are widespread," she says, and instead focusing on "how do you live with them?"

In Oregon, USFS officials are weighing difficult choices in the fight against wiregrass. On one hand, they want to intentionally burn more forests, in order to clear out brush and saplings, and restore ecosystems that historically evolved with fire. But such controlled burns also risk opening new areas to grass invasions.

With that in mind, Kerns has obtained funding for an experiment that will use herbicides to kill *Ventenata* along forest edges before crews set controlled burns, to see whether that curbs later encroachment by the grass. In the scablands, Priest is trying a more aggressive approach: sending in heavy equipment such as bulldozers to help squelch fires, fearing that native plants will suffer if the open areas burn. But he worries the equipment could leave lasting marks on delicate ecosystems—and that the disturbance could ultimately benefit the grasses. Learning to live in a "*Ventenata* world" is challenging, Priest says, and he doesn't know whether "I actually have the right answer."

Kerns echoes that uncertainty. Trying to cope with invasive grasses poses a conundrum, she says. "We kind of want to go back and restore things, but it's not the same world." ■

In Arizona's Saguaro National Park, workers are trying to curb buffelgrass, which fuels blazes that threaten saguaro cacti.

PHOTOS: (TOP TO BOTTOM) CAMILA GUILLEN/UF/IFAS; NORMA JEAN GARGASZ/ALAMY STOCK PHOTO

INSIGHTS

POLICY FORUM

BIOMEDICAL REGULATION

Transparency practices at the FDA: A barrier to global health

Data sharing among regulators must be "business as usual"

By **Murray M. Lumpkin[1], Margaret A. Hamburg[2], William B. Schultz[3], Joshua M. Sharfstein[4]**

During the COVID-19 pandemic, scientists at the US Food and Drug Administration (FDA) have reviewed large numbers of pandemic-related tests, medications, and vaccines. However, long-standing confidentiality practices have kept FDA from sharing many of these analyses and the data behind them with the regulatory agencies of other nations, especially those in low- and middle-income countries (LMICs). With FDA not

ILLUSTRATION: DAVIDE BONAZZI/SALZMANART

sharing key information, the primary source of dependable COVID-19 product regulatory documentation and information for resource-constrained countries has been the World Health Organization (WHO) in coordination with leading European regulators. These efforts are commendable, but in many cases FDA's assessments will be some of the most sought after and scientifically robust in the world—and should be shared with the widest possible regulatory audience. FDA must demonstrate similar leadership and commitment to global health by reforming its outdated, restrictive practices on information sharing.

Even absent a pandemic, most agencies find the challenge of medical product regulation daunting. Thousands of medical products come onto the global market annually, but few agencies have the capability to assess these products thoroughly. The WHO has estimated that only one-quarter of its member states have agencies with at least a "stable, well-functioning, and integrated regulatory system" (*1*). In most countries, underfunded and understaffed agencies struggle to meet basic regulatory tasks. Even mature agencies find that they sometimes lack the resources they need to meet expectations.

In 2020, a US National Academies of Sciences, Engineering, and Medicine committee highlighted one solution to this challenge: strengthening reliance-based regulatory pathways in which agencies use the extensive reviews and inspections conducted by trusted counterparts to better inform their own regulatory decision-making (*2*). For such an approach to work, timely sharing of complete critical information and relevant documents between agencies is essential. Such access allows a relying agency not only to trust but to understand what a reference agency has decided, and then to use that information to help inform the appropriate decision for its population. Such reliance-based regulatory decision-making is now a 21st-century "best regulatory practice," enshrined in WHO guidelines (*3*).

Unfortunately, FDA entered the current pandemic ill equipped to serve in the role of reference agency to LMICs because of its decades-old confidentiality practices, dating back to a world in which most US pharmaceutical product development, manufacturing, and sales were of US or European manufactured products. In that era, other regulators' actions had negligible impact on product development, authorization, or access in the United States. Yet FDA's strict prohibitions on data sharing have remained in place even as the global pharmaceutical ecosystem has been transformed. Today, products may be manufactured in part or in whole in various countries and shipped widely before distribution. During that journey, the product will often be under the authority of a variety of agencies of varying capabilities. Sharing of critical information garnered along this global manufacturing and transit train is needed for effective oversight of the global supply chain of all products, including those bound for the United States.

Yet until 1993, FDA regulations did not even distinguish between release of information to the general public and document sharing with other regulators (*4*); as a result, no information considered confidential was shared. Then in 1993, FDA took a step toward greater international cooperation. The agency adopted a regulation permitting the agency to share some nonpublic information with foreign regulators, provided that the foreign country makes certain commitments to protect the information from disclosure (*5*). In practice, however, FDA created an extensive process to validate the ability of countries to maintain confidentiality.

Nearly 30 years later, the agency has established confidentiality arrangements with only a few counterpart agencies (*6*), focusing on a small number of countries perceived to be able to provide the most helpful information in return—not on agencies in LMICs. Moreover, in all but one of these arrangements, FDA does not share data classified as "trade secret," a term that the agency has applied more broadly than necessary to include large amounts of manufacturing and biopharmaceutical data (*7*).

The one agreement permitting the sharing of trade secret information, with the European Medicines Agency (EMA), took 25 years and several legislative interventions to operationalize (*8*, *9*). It was made possible by 2012 legislation that permitted the sharing of unredacted reports after FDA verified the "demonstrated ability" of foreign regulators to protect trade secret information (*10*). FDA extended this approach to the United Kingdom's regulatory agency after it left the European Union because it was part of the EMA agreement. FDA has made no other such agreements regarding full inspection reports.

Separately, FDA has established a limited data sharing relationship with WHO, which exemplifies the challenges of current FDA practices with LMICs. FDA sends certain redacted documents to WHO, which then, using its own scarce resources, further distills information into a WHO document, which WHO can share (*11*). But the heavily redacted documents FDA can share make it difficult or impossible for resource-constrained agencies to rely on FDA regulatory activities when making their own regulatory decisions.

During the current pandemic, FDA has issued emergency use authorizations (EUAs) for multiple products. The only

[1]Bill & Melinda Gates Foundation, Seattle, WA, USA. [2]Independent researcher, Washington, DC, USA. [3]Zuckerman Spaeder, Washington, DC, USA. [4]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21209, USA. Email: joshua.sharfstein@jhu.edu

public documents related to these actions are the authorization letters to the companies and the "fact sheets" for health care providers and for patients and caregivers. The full (unredacted) assessment and inspection reports, were they available, would have been enormously valuable to agencies in resource-constrained countries. These agencies are highly interested in what the FDA EUA decision means, what data support it, what dataset was submitted to FDA, where and how the product was manufactured, and why FDA made the decision it made. In some cases, FDA may have been the first or only global regulator inspecting a clinical trials site or manufacturing facility or authorizing a product. Without this critical information, agencies cannot even assure that shipments they receive are the same version as those authorized by FDA (and not, for example, from a manufacturing facility not assessed by FDA).

In contrast to FDA, WHO and the European agencies initiated special procedures to include assessors from several countries, including LMICs, in their deliberations before product authorization (*12*). Such reliance-based pathways reduced the need for every country to conduct time-consuming, redundant, and often poorly informed assessments and inspections. It also created an extra benefit of faster access to markets in LMICs for companies seeking authorization in Europe, compared with the United States.

The success of this WHO and Europe-led system of critical information sharing is exemplified by the rapidity of African COVID-19 vaccine authorizations. As of March 2022, in the 54 African countries, there were 277 initial authorizations of COVID-19 vaccines, which had received emergency use listing by WHO. In 52 countries, multiple such vaccines have been authorized (*13*). These COVID-19 vaccine authorizations were facilitated by using a reliance-based pathway endorsed by the African Union and WHO (*14*). The vast majority are based on EMA's authorization decisions because EMA and WHO worked closely together on the products' assessments and because both EMA and WHO shared—in real time—helpful critical scientific assessment and inspections documents with the African agencies.

Meanwhile, because of its confidentiality practices, FDA was only minimally involved in helpful information sharing with resource-constrained agencies. Had FDA data and analysis been more widely available, many countries could have used them to make rapid, informed decisions for their populations.

Three steps FDA should take for greater global data-sharing

1. **Share copies of assessment and inspections reports for all pandemic-related products.** As a short-term measure during this declared state of emergency, FDA should waive its current practices so that it can share, upon request, with WHO and all other counterpart agencies, whether products have been authorized or not. For the many countries without a confidentiality agreement, FDA should move quickly to accept their statements of authority and commitment for confidentiality. The agency should limit redactions to personally identifiable information and trade secrets as narrowly defined in the statute, which would permit the sharing of much helpful biopharmaceutical and manufacturing data.

   This step can still help with the COVID-19 pandemic response. There are new products and new versions of older products to be assessed; there are pressing questions concerning childhood vaccination, boosters, and variants to be answered; there are changes in manufacturing procedures to be validated to enable greater and more cost-effective product production; and there are new efficacy and safety data to be analyzed.

2. **Develop new practices on data-sharing with WHO and counterpart agencies outside public health emergencies.** This should include a rapid path to accepting assurances from foreign counterparts regarding commercial confidential and trade secret information. These new practices should facilitate reliance-based pathways by counterpart agencies and recognize that these agencies should not be considered part of the "general public" for purposes of confidentiality. FDA should also consider additional reforms to make sharing of fully unredacted reports, especially inspection reports, less onerous.

3. **Publish annual reports of its data-sharing activities with counterpart regulators.** This should include both numbers of requests received and from whom and numbers of requests fulfilled and with whom. These reports should document the consequences both for global health and for oversight of products headed to the US market. They should also assess emerging challenges to data sharing, along with potential solutions.

Successful US leadership in the modern global pharmaceutical ecosystem, as well as global health and safety, depends on the ability of regulators to trust each other and work together. Greater FDA global data-sharing engagement starts with three steps (see the box). Progress will require more resources, capacity building, and will. In the 21st century, sharing data and critical documents among regulators cannot be an afterthought: It must be "business as usual," including for FDA. ■

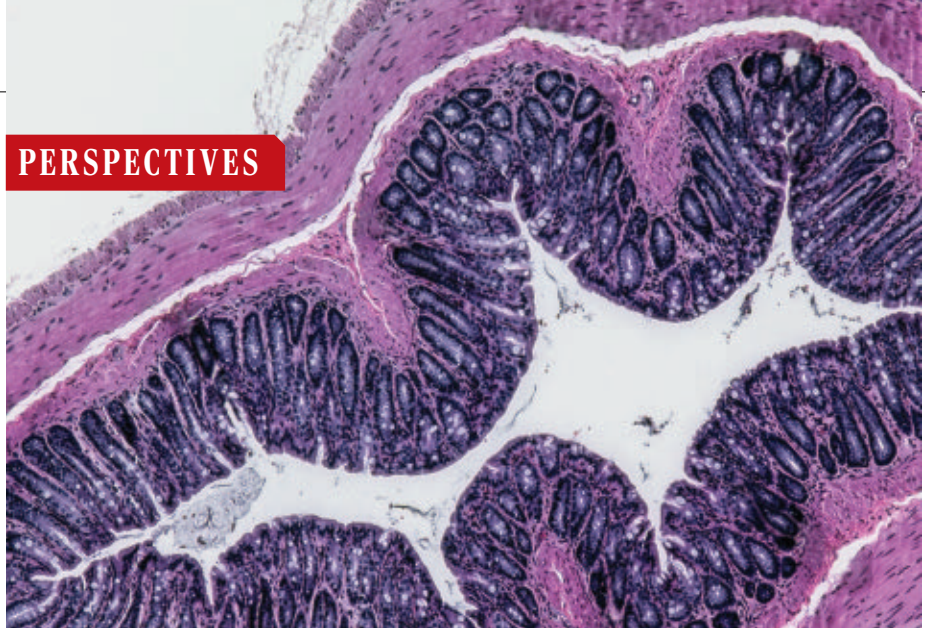
REFERENCES AND NOTES

1. WHO, "List of National Regulatory Authorities (NRAs) operating at maturity level 3 (ML3) and maturity level 4 (ML4)" (WHO, 2022); https://www.who.int/initiatives/who-listed-authority-reg-authorities/maturity-level.
2. US National Academies of Science, Engineering, and Medicine, "Mutual recognition agreements in the regulation of medicines" (2019); https://www.nationalacademies.org/our-work/mutual-recognition-agreements-in-the-regulation-of-medicines.
3. WHO, "TRS 1033 - 55th report of the WHO Expert Committee on Specifications for Pharmaceutical Preparations" (WHO, 2021); https://www.who.int/publications/i/item/55th-report-of-the-who-expert-committee-on-specifications-for-pharmaceutical-preparations.
4. Public information; Communications with Foreign Government Officials. *Fed. Regist.* **58** (223), 61598 ff (19 November 1993).
5. 21 Code of Federal Regulations 20.89.
6. FDA, Confidentiality commitments; https://www.fda.gov/international-programs/international-arrangements/confidentiality-commitments.
7. Trade secret is defined as "any commercially valuable plan, formula [that has] a direct relationship to the production process," with examples including "the type or brand of equipment used in manufacturing, product formulas, product components or ingredients not on the label, specifications that are unique (i. e., not in the US Pharmacopeia), and technical designs" (*8*).
8. FDA, FDA Staff Manual Guide (SMG 2830.3).
9. FDA, Mutual Recognition Agreement (MRA); https://www.fda.gov/international-programs/international-arrangements/mutual-recognition-agreement-mra.
10. The Food and Drug Administration Safety and Innovation Act of 2012, Pub. L. No. 112-144.
11. FDA, "FDA In Brief: FDA announces pilot program with World Health Organization to expedite review of HIV drug applications" (FDA, 2018); https://www.fda.gov/news-events/fda-brief/fda-brief-fda-announces-pilot-program-world-health-organization-expedite-review-hiv-drug.
12. EMA, Questions and answers on the pilot project 'OPEN'. Opening our Procedures at EMA to Non-EU authorities (EMA, 2021); https://www.ema.europa.eu/en/documents/other/questions-answers-pilot-project-open_en.pdf.
13. WHO, AVAREF COVID-19 Africa vaccine dashboard; https://www.afro.who.int/health-topics/immunization/avaref/covid-19-africa-vaccine-dashboard.
14. Africa Centers for Disease Control and Prevention (Africa CDC), "Guidance on emergency expedited regulatory authorisation and access to COVID-19 vaccines in Africa" (Africa CDC, 2021); https://africacdc.org/download/guidance-on-emergency-expedited-regulatory-authorisation-and-access-to-covid-19-vaccines-in-africa.

ACKNOWLEDGMENTS

M.M.L. is a former deputy commissioner for international and special programs of FDA. M.A.H. is a former commissioner of FDA. W.B.S. is a former deputy commissioner for policy of FDA and former general counsel of the US Department of Health and Human Services. J.M.S. is a former principal deputy commissioner of FDA.

10.1126/science.abq4981



Regulatory T cells induced by β-hexosaminidase from Bacteroidetes species in the mouse gut microbiota can suppress colitis, as shown in this hematoxylin and eosin–stained section of a mouse colon.

PHOTO: ANGELINA M. BILATE

IMMUNOLOGY

Twice the tolerance

A gut microbiota–derived antigen elicits distinct subsets of regulatory T cells to suppress inflammation in mice

By **Meera K. Shenoy**[1] and **Meghan A. Koch**[1,2]

Mammalian survival depends on the ability of the immune system to mount sterilizing protection against pathogens while regulating (suppressing) responses toward innocuous antigens such as from food or commensal microbes. Within the intestine, two key helper T cell populations, $CD4^+$ peripheral regulatory T cells (pT_{reg} cells) and $CD4^+CD8\alpha\alpha^+$ intraepithelial lymphocytes (CD4IELs), are largely responsible for restraining aberrant inflammatory responses against self and innocuous foreign antigens (*1*). Although progress has been made in understanding the mechanisms responsible for gut T_{reg} cell development and function, far less is known about the specificity, differentiation, and role of CD4IELs. On page 660 of this issue, Bousbaine *et al.* (*2*) reveal that β-hexosaminadase (β-hex), a conserved antigen expressed by several gut commensal bacterial species, drives CD4IEL differentiation. These β-hex–elicited CD4IELs cooperate with pT_{reg} cells to limit pathology in a mouse model of colitis.

IELs are a tissue-resident immune cell population that resides at the basolateral surface of the intestinal epithelial cell barrier and helps to maintain tissue homeostasis within the intestines. Although there are many types of IELs, CD4IELs are specifically "induced" from conventional $CD4^+$ T cells in the periphery after antigen exposure and tissue cues in gut-associated secondary lymphoid organs. These cells then upregulate $CD8\alpha\alpha$ expression and traffic to the intestinal epithelial barrier, where they can exert both regulatory and cytotoxic functions to reinforce gut health (*3*, *4*).

Recent studies aimed at defining the antigens and tissue signals that drive CD4IEL differentiation have revealed a fundamental role for the gut microbiota. Germ-free mice almost entirely lack CD4IELs; however, upon microbial reconstitution, CD4IELs differentiate and resemble conventional mice in number and phenotype (*4*). Additionally, conventional mice treated with broad-spectrum antibiotics experience a precipitous drop in CD4IELs, which rebound as the microbiota reestablishes itself after antibiotic cessation (*5*). Thus, the microbiota is required for both the differentiation and maintenance of CD4IELs, yet the mechanisms underlying this process remain unclear.

Lactobacillus reuteri, a common member of the mouse intestinal microbiota, metabolizes tryptophan to produce indole derivatives. Binding of these derivatives to the aryl hydrocarbon receptor (AHR) on $CD4^+$ T cells was sufficient to downregulate expression of the transcription factor Thpok, thus driving CD4IEL differentiation. Other studies correlated the abundance of specific commensal species such as *L. reuteri* or *Faecalibacterium prausnitzii* with CD4IEL counts, yet it was unknown whether commensal-derived antigens could directly trigger CD4IEL differentiation or function (*5*, *6*).

Bousbaine *et al.* identified β-hex derived from the commensal *Parabacteroides goldsteinii* as an antigen that is recognized by $CD4^+$ T cells and triggers their differentiation into CD4IELs and pT_{reg} cells in mice. They found that β-hex–specific $CD4^+$ T_{reg} cells and CD4IELs can be readily found in mice colonized with a complex microbiota, and the *P. goldsteinii* β-hex epitope is evolutionarily conserved across several species of the *Bacteroidetes* phylum. Additionally, β-hex from these species stimulated increased frequencies of β-hex–specific CD4IELs, whereas β-hex from species with different epitopes failed to do the same. Thus, the findings of Bousbaine *et al.* indicate that conserved bacterial epitopes derived from abundant, yet distinct, commensal bacteria support the expansion of CD4IELs.

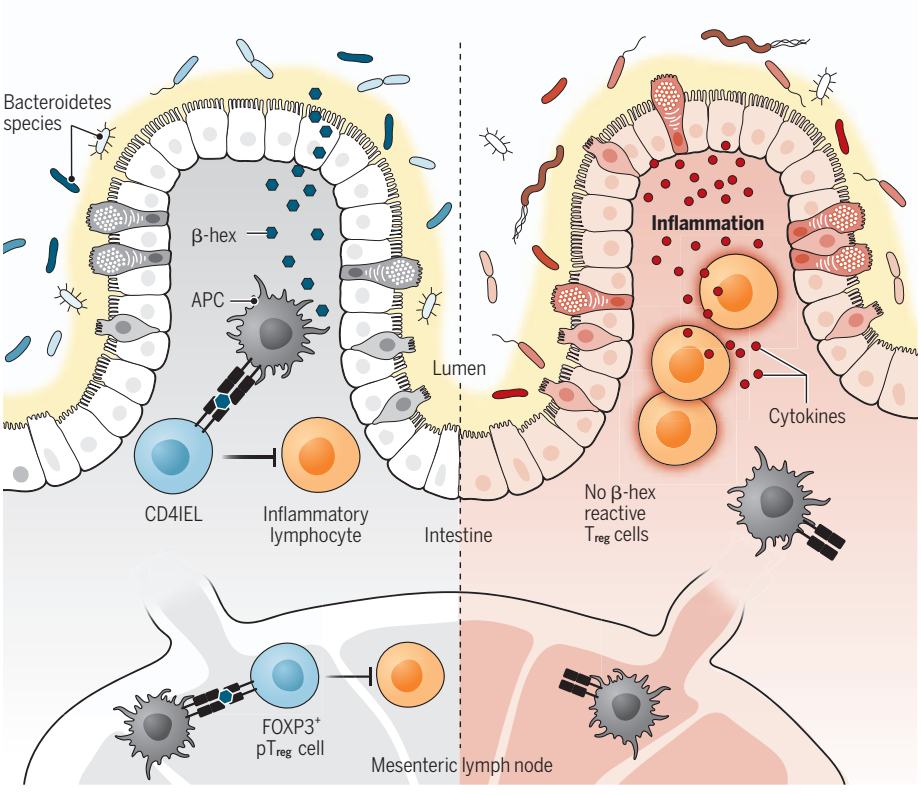
How do β-hex–elicited CD4IELs participate in intestinal homeostasis? Transcriptional analyses of mouse IELs and correlative studies of CD4IELs from patients with HIV or inflammatory bowel disease (IBD) suggested that CD4IELs have a regulatory role in maintaining tissue homeostasis and limiting pathology (*4*, *7*–*9*). Using gene expression analysis, the authors confirmed that β-hex–specific CD4IELs were transcriptionally similar to the total CD4IEL pool. Using an adoptive transfer model of colitis in which naïve $CD4^+$ T cells trigger severe intestinal inflammation when transferred into immune-deficient recipient animals, Bousbaine *et al.* demonstrated that the addition of commensal-specific $CD4^+$ T cells (which differentiate into pT_{reg} cell and CD4IEL populations) could fully rescue colitis pathology and morbidity. Moreover, when mice were given $CD4^+$ T cells that could only differentiate into CD4IELs, half were protected from colitis. These key findings demonstrate that CD4IELs elicited by a conserved commensal antigen can restrain harmful inflammatory responses in the gut (see the figure).

Ultimately, the study of Bousbaine *et al.* addresses an important outstanding question in mucosal immunology regarding how the microbiota contributes to the development and function of CD4IELs. One of the most fascinating findings is that a single commensal antigen can drive the differ-

[1]Division of Basic Sciences, Fred Hutchinson Cancer Center, Seattle, WA, USA. [2]Department of Immunology, University of Washington School of Medicine, Seattle, WA, USA. Email: mkoch@fredhutch.org

A bacterial antigen promotes intestinal homeostasis in mice

A conserved epitope in the Bacteroidetes enzyme β-hexosaminidase (β-hex) is presented by antigen-presenting cells (APCs) to CD4$^+$ cells to drive their differentiation to CD4$^+$CD8αα$^+$ intraepithelial lymphocytes (CD4IELs) in the intestinal tissue and forkhead box P3$^+$ (FOXP3$^+$) peripheral regulatory T cells (pT$_{reg}$ cells) in gut-associated lymphoid organs. These two T$_{reg}$ cell populations suppress intestinal inflammation and promote homeostasis.



entiation of two distinct types of T$_{reg}$ cell: CD8αα$^+$ CD4IELs in the intestinal tissue and forkhead box P3$^+$ (FOXP3$^+$) pT$_{reg}$ cells in the gut-associated lymphoid organs. This finding suggests that there may be numerous factors, both antigen-dependent and -independent, that determine how T$_{reg}$ cells differentiate and how they behave in situ. How β-hex elicits two different regulatory cell populations remains to be revealed. Perhaps the context of antigen presentation (such as location or antigen-presenting cell) dictates β-hex–specific T cell fate. Intriguingly, the microbial antigen in question is conserved across numerous bacterial species. It will be interesting to dissect whether induction of T$_{reg}$ cells confers an advantage to the microbial species that express β-hex. Another open question concerns the specificity of other mouse CD4IELs, as well as CD4IELs in humans: Do they also recognize commensal-derived ligands? Relatedly, if other CD4IELs are indeed specific for the microbiota, it is possible that they may have "sister" pT$_{reg}$ cells with the same specificity.

This work also provides new insight regarding the therapeutic potential of the microbiota. For example, treatment with β-hex or other commensal antigen(s) recognized by CD4IELs may mitigate intestinal inflammation in affected individuals. Notably, only half of the animals were protected from colitis when antigen-specific CD4$^+$ T cells were precluded from becoming FOXP3$^+$ pT$_{reg}$ cells and forced toward CD8αα$^+$ CD4IEL differentiation, highlighting the complexity of immune regulation in the intestine. The discoveries and new tools reported by Bousbaine *et al.* lay the groundwork for testing whether other commensal antigens can elicit CD4IELs, delineating the pathways involved in their differentiation, and manipulating this process to promote intestinal homeostasis. ■

REFERENCES AND NOTES

1. T. Sujino *et al.*, *Science* **352**, 1581 (2016).
2. D. Bousbaine *et al.*, *Science* **377**, 660 (2022).
3. P. J. Morrissey, K. Charrier, D. A. Horovitz, F. A. Fletcher, J. D. Watson, *J. Immunol.* **154**, 2678 (1995).
4. D. Mucida *et al.*, *Nat. Immunol.* **14**, 281 (2013).
5. L. Cervantes-Barragan *et al.*, *Science* **357**, 806 (2017).
6. H. Sokol *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16731 (2008).
7. J. Shires, E. Theodoridis, A. C. Hayday, *Immunity* **15**, 419 (2001).
8. B. S. Boland *et al.*, *Sci. Immunol.* **5**, abb4432 (2020).
9. S. Mehandru *et al.*, *J. Exp. Med.* **200**, 761 (2004).

10.1126/science.add7145

TIME CRYSTAL

Unleashing spontaneity in a time crystal

Ordered patterns reoccur over time in an ultracold atomic gas trapped in light

By **Lindsay J. LeBlanc**

The quest to understand how systems develop order underlies the fundamental understanding of almost all sciences, from the origin of galaxies to the origin of life. Over the past decade, ideas about temporal crystallization have been explored theoretically and experimentally, probing the question of whether systems can develop regular, repeating behavior through time, in the same way that spatial crystals regularly cycle through different configurations. These so-called "time crystal" behaviors have been observed before (*1*–*6*), with their periodic properties usually driven by periodic external forces. On page 670 of this issue, Kongkhambut *et al.* (*7*) demonstrate a time crystal that emerges from a time-independent initial state without the application of a recurring external force, analogous to the formation of spatial crystals. The study provides insight into the nature of dynamical phases in quantum systems and how order can manifest in the presence of dissipative processes.

To help understand the abstract concept of temporal order, it is necessary to first establish what "order" means in space. One important concept to consider when it comes to understanding spatial order is symmetry. For example, a liquid has perfect continuous translational symmetry because its properties are identical at every point in space. However, once the liquid is turned into a crystalline solid—for example, from water to ice—this continuous translational symmetry is lost, and the properties of this solid are the same only when the location of the observer is shifted by multiples of the unit cell length of the crystal. There is no predetermined location at which a specific atom should find itself upon crystallization; an anchoring lattice site is spontaneously and randomly selected as the crystal freezes in place. This sponta-

Department of Physics, University of Alberta, Edmonton, AB, Canada. Email: lindsay.leblanc@ualberta.ca

GRAPHIC: K. HOLOSKI/*SCIENCE*

neity in selecting a starting point for the emergent order is a defining characteristic in phase transitions.

In the spirit of understanding systems through their symmetries, physicists have wondered whether the same concept exists in time (*8*, *9*)—whether there could exist "time crystals," physical systems with stable, periodic behaviors that break time-translation symmetry. This notion is at odds with conventional quantum mechanics, which presumes that the natural state of a system does not depend on time. Although studies have shown that time crystals are not possible in most "natural" situations, they are possible for systems that are driven and/or dissipative—when energy is added to or lost by the system. These "time crystallization" behaviors have been demonstrated, for example, in trapped ions (*1*), nitrogen vacancy centers (*2*), nuclear magnetic resonance (*3*), superfluid helium (*4*), cold atoms (*5*), and superconducting qubits (*6*). All of these experiments were performed in periodically driven systems, in which the periodicity emerged at twice the period of the driving force, changing the original symmetry to a different symmetry. Although these pioneering results lend insight into emergent behavior in driven systems, the timing of the recurrent behavior is a simple multiple of the external driver's timing. Because the temporal pattern relies on an external driver, this raises the question of whether the nomenclature "time crystallization" is really the best way to describe this phenomenon.

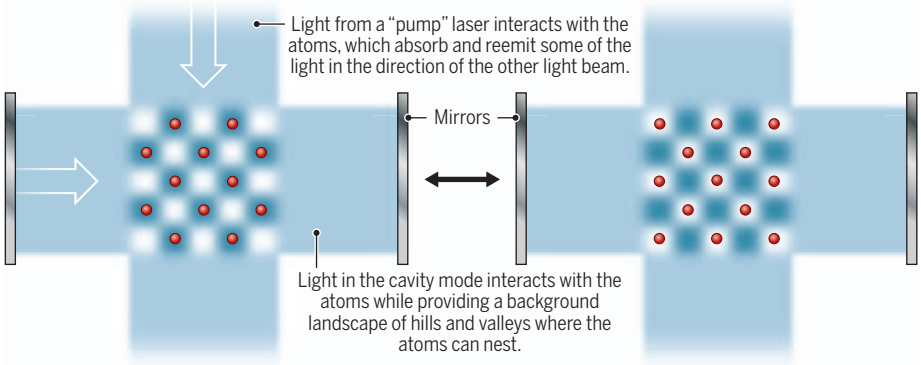
Kongkhambut *et al.* demonstrate a kind of time crystallization that breaks continuous time translation symmetry without the need of a recurring external force. Using a Bose-Einstein condensate of ultracold rubidium atoms inside an optical cavity, the resonance frequency of the cavity was engineered to be about the same as that of the atoms, while a finely tuned standing-wave laser was used as a pump to interact with the atoms (see the figure). Through a phenomenon known as Dicke superradiance, the atoms absorb the photons from the laser and then reemit the light into the cavity (*10*, *11*). In turn, the light emitted into the cavity is reabsorbed by the atoms again and mediates an effective interaction across all atoms in the cavity, repeatedly. This phenomenon prompts the emergence of a spatial order among the atoms, breaking the continuous spatial translation symmetry of the atoms—a Bose-Einstein condensate—into a two-dimensional periodic lattice and creating a spatial crystal.

Here, Kongkhambut *et al.* make a small but crucial addition to the setup: by adding a "blue-detuned" light (*12*), whose energy is higher relative to that of the atoms. This change makes two possible structural configurations possible, in which both would compete to lower the overall energy of the system. Bathing in the blue light, the atoms prefer to remain in the low-intensity regions of the laser field. This reduces the probability of them emitting light into the cavity, which then lowers the strength of the standing-wave potential that keeps them in the first configuration. Free from this trap, the atoms would then move into regions where they are more likely to emit light into the cavity in the second configuration. However, in this location, the atoms begin to interact more strongly with the pump light, which reestablishes the standing-wave lattice and shifts atoms back to their original positions, and a recurring cycle is born.

As a first demonstration, the set of conditions over which this phenomenon arises is narrow because of the delicate balance between energy scales. And like the periodically driven systems, there is a native time scale built into this system in the round-trip timing of light reflecting between the mirrors of the cavity. Moving forward, the limits of these systems are sure to be pushed to better understand the nature of dynamical phases, such as exploring the transitions between normal and crystalline states (*13*), and in adding experimental probes such as quantum gas microscopy to gain insight into the correspondence between spatial and temporal order. The study of dynamical many-body quantum systems remains an important area of investigation whose outcomes may range from the practical applications of time crystallization in precision metrology to answering questions about the fundamental nature of time in quantum mechanics. ■

A shape-shifting lattice of atoms

Atoms inside a cavity choose their locations according to the lowest energy configuration they can find. Kongkhambut *et al.* designed conditions in which two competing configurations are equally preferred and the system spontaneously begins oscillations between them, creating a "time crystal."



In regions of high intensity (the "hills"), the atoms emit more light into the cavity mode than they absorb, making the configuration energetically unfavorable over time.

After moving to the "valleys" with low laser intensity, emission from the atoms into the cavity is reduced, and the atoms eventually return to the hills where the interaction energy is lower.

GRAPHIC: K. HOLOSKI/*SCIENCE*

The final ingredient that makes this a time crystal is a short delay between the atoms' shift in position and the light being emitted to shift them back, which determines the rhythm of the process. A key feature in this system that allows this behavior to be observed is that a small portion of the light leaks out of the cavity during each cycle and varies in amplitude with the same time dependence as that of the atom positions in the cavity. This makes possible the direct measurement of the spontaneity of the symmetry breaking, revealed as the randomly distributed phases of oscillation measured over repeated instances of this experiment. Together with an additional check of the rigidity of this phase in the presence of noise, Kongkhambut *et al.* lay claim to the discovery of a continuous time crystal—one that spontaneously begins its recurring behavior at any point in time, in contrast to discrete time crystals that begin oscillating only at specific, predetermined times.

REFERENCES AND NOTES

1. J. Zhang *et al.*, *Nature* **543**, 217 (2017).
2. S. Choi *et al.*, *Nature* **543**, 221 (2017).
3. J. Rovny, R. L. Blum, S. E. Barrett, *Phys. Rev. Lett.* **120**, 180603 (2018).
4. S. Autti, V. B. Eltsov, G. E. Volovik, *Phys. Rev. Lett.* **120**, 215301 (2018).
5. J. Smits, L. Liao, H. T. C. Stoof, P. van der Straten, *Phys. Rev. Lett.* **121**, 185301 (2018).
6. X. Mi *et al.*, *Nature* **601**, 531 (2022).
7. P. Kongkhambut *et al.*, *Science* **377**, 670 (2022).
8. F. Wilczek, *Phys. Rev. Lett.* **109**, 160401 (2012).
9. A. Shapere, F. Wilczek, *Phys. Rev. Lett.* **109**, 160402 (2012).
10. K. Baumann, C. Guerlin, F. Brennecke, T. Esslinger, *Nature* **464**, 1301 (2010).
11. N. Dogra *et al.*, *Science* **366**, 1496 (2019).
12. F. Piazza, H. Ritsch, *Phys. Rev. Lett.* **115**, 163601 (2015).
13. D. Dreon *et al.*, arXiv:2112.11502 [cond-mat.quant-gas] (2021).

10.1126/science.add2015

REGENERATION

A regenerative niche for stem cells

Production of hyaluronic acid allows regenerative signaling in muscle stem cells after injury

By **Davide Gabellini**

Skeletal muscle is the largest tissue in the human body and has a substantial capacity to recover its structure and function after acute damage. This regenerative ability is due to muscle stem cells (MuSCs), which are normally in a dormant state called quiescence (*1*). After injury, signals from dying muscle fibers and various tissue-resident and infiltrating cell types, including inflammatory cells, lead to MuSC activation and promotion of their muscle regeneration activity. Although integration of the different signals present in the injury microenvironment (niche) is key to muscle repair and reconstitution of the quiescent stem cell pool (*2*), how MuSCs adapt to the altered niche is not completely known. On page 666 of this issue, Nakka *et al.* (*3*) show that MuSCs activate the production of the extracellular matrix protein hyaluronic acid (HA), which is required to overcome inhibitory inflammation signaling from the injury niche, exit quiescence, and initiate muscle repair.

Cell fate transitions induced by muscle damage are associated with important changes in the epigenetic landscape of MuSCs (*4*). Nakka *et al.* investigated the role of Jumonji domain–containing protein 3 (JMJD3) and ubiquitously transcribed TPR protein on the X chromosome (UTX), two enzymes responsible for removal of the inhibitory modification trimethylated histone H3 lysine 27. By generating mice conditionally ablated for either enzyme in MuSCs, they found that the two proteins have distinct roles in muscle regeneration. JMJD3 is mainly involved in the early steps of MuSC activation, whereas UTX plays an important role in subsequent MuSC proliferation and differentiation. Among the specific JMJD3 targets, the authors identified that the gene hyaluronan synthase 2 (*Has2*), which encodes a key enzyme in HA biosynthesis, is important for MuSC activation. Accordingly, JMJD3 ablation in MuSCs in mice results in a reduction of HA deposition in the injured niche. Notably, MuSC activation and muscle regeneration defects observed upon JMJD3 MuSC-specific deletion can be phenocopied by treatment of wild-type mice with the HA synthesis inhibitor 4-methylumbelliferone. Also, treatment with HA can partially rescue defects from JMJD3 MuSC-specific deletion. Collectively, these data indicate that the JMJD3-HAS2-HA pathway plays a key role in muscle repair (see the figure). In the future it will be important to determine the upstream signaling pathway(s) that are responsible for JMJD3 activation upon muscle injury. Also, understanding whether other stem cells use the same molecular switch to adapt to inflammatory conditions and allow tissue repair could help identify potential therapeutic targets.

HA is not only a major component of the extracellular matrix, which maintains tissue structure and rigidity, but is also important for regulating cell-cell interactions and modulating the local concentration of soluble factors in the microenvironment (*5*). This is particularly relevant in the context of muscle regeneration, in which MuSCs and various inflammatory cells reciprocally coordinate the repair process (*6*). Upon muscle injury, MuSCs stimulate tissue-resident macrophages to attract circulating inflammatory cells. These cells are initially needed to remove dead cells and make room for new muscle (*6*). They also produce a number of cytokines, including interferon-γ (IFN-γ) and interleukin-6 (IL-6), which exert muscle proregenerative effects by stimulating MuSC proliferation (*7*, *8*). Notably, Nakka *et al.* demonstrated that IFN-γ and IL-6 hinder MuSC activation and that HA is needed to overcome this inhibition, allowing MuSCs to be activated and contribute to muscle regeneration.

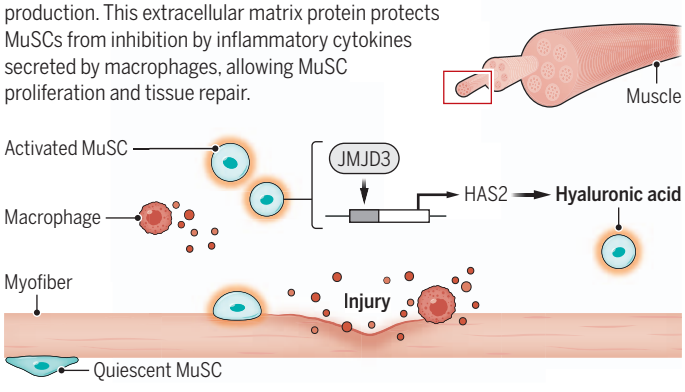
Nakka *et al.* also showed that HA is not intrinsically required for MuSCs to exit quiescence. HA is needed only when MuSCs are exposed to an inflammatory microenvironment. Indeed, MuSCs that are unable to produce HA display an apparently intact activation in response to injury in the absence of inflammation. This raises the exciting possibility, deserving of future investigation, that HA is not needed just to shield MuSCs from the negative influence of inflammatory cytokines but that HA and its receptors could be signals in the injury niche that promote regeneration.

Later during the muscle regeneration process, inflammatory cells modify their gene expression and metabolic programs and stimulate MuSC fusion and differentiation to build new muscle tissue. These processes are required for resolution of inflammation, tissue restoration, and the return to homeostasis (*6*). It would be interesting to determine whether HA produced by MuSCs contributes to integrating signals and regulating cell-cell interactions during this phase of muscle regeneration.

Previous in vitro studies found that, while promoting muscle cell proliferation, HA inhibits their differentiation (*9*). Nakka *et al.* showed that HA levels are very low in the quiescent MuSC niche and are dynamically regulated during regeneration, peaking at the time of MuSC activation. Besides synthesis by hyaluronan synthases, HA levels are controlled by degradation, which is

Hyaluronic acid protects muscle stem cells

In the absence of muscle damage, muscle stem cells (MuSCs) are in a dormant, quiescent state. After muscle injury, MuSCs up-regulate the expression of Jumonji domain–containing protein 3 (JMJD3), which derepresses the hyaluronan synthase 2 (*Has2*) promoter; the resulting HAS2 protein increases hyaluronic acid production. This extracellular matrix protein protects MuSCs from inhibition by inflammatory cytokines secreted by macrophages, allowing MuSC proliferation and tissue repair.



GRAPHIC: A. MASTIN/*SCIENCE*

Division of Genetics and Cell Biology, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) San Raffaele Scientific Institute, Milano, Italy. Email: gabellini.davide@hsr.it

carried out by several hyaluronidases that are expressed in a tissue-specific manner (*10*). Whether HA degradation contributes to MuSC return to quiescence upon muscle damage repair is a relevant aspect to be investigated in the future.

Aberrant HA accumulation has been reported in cellular models of facioscapulohumeral muscular dystrophy (FSHD), and preventing HA buildup reduces pathological symptoms (*11*). Hence, it would be interesting to investigate whether increased or persistent HA production contributes to FSHD by interfering with MuSC homeostasis and repair. Another condition associated with altered HA levels is aging, in which HA in skin connective tissue and in synovial joint fluid is reduced. For this reason, HA creams and injections are used to reduce wrinkles or for arthritis treatments. Whether an altered HA production by MuSCs contributes to the abnormal deposition of extracellular matrix components and the reduced muscle tissue regenerative potential in aging (*12*) remains to be investigated. Overall, the discovery of the JMJD3-HAS2-HA pathway and its central role in MuSC activation upon muscle injury by Nakka *et al.* opens many new avenues for both understanding damage responses and developing therapies for muscle diseases. ■

"...Muscle stem cells activate the production of...hyaluronic acid...to overcome inhibitory inflammation signaling... and initiate muscle repair."

REFERENCES AND NOTES

1. H. Yin, F. Price, M. A. Rudnicki, *Physiol. Rev.* **93**, 23 (2013).
2. M. Loreti, A. Sacco, *NPJ Regen. Med.* **7**, 16 (2022).
3. K. Nakka *et al.*, *Science* **377**, 666 (2022).
4. D. Cicciarello, L. Schaeffer, I. Scionti, *Front. Cell Dev. Biol.* **10**, 917771 (2022).
5. F. Gattazzo, A. Urciuolo, P. Bonaldo, *Biochim. Biophys. Acta, Gen. Subj.* **1840**, 2506 (2014).
6. B. Chazaud, *Trends Immunol.* **41**, 481 (2020).
7. M. Cheng, M. H. Nguyen, G. Fantuzzi, T. J. Koh, *Am. J. Physiol. Cell Physiol.* **294**, C1183 (2008).
8. C. Zhang *et al.*, *J. Biol. Chem.* **288**, 1489 (2013).
9. M. J. Kujawa, K. Tepperman, *Dev. Biol.* **99**, 277 (1983).
10. A. Lierova *et al.*, *Pharmaceutics* **14**, 838 (2022).
11. A. M. DeSimone, J. Leszyk, K. Wagner, C. P. Emerson Jr., *Sci. Adv.* **5**, aaw7099 (2019).
12. P. Sousa-Victor, L. García-Prat, P. Muñoz-Cánoves, *Nat. Rev. Mol. Cell Biol.* **23**, 204 (2022).

ACKNOWLEDGMENTS

D.G. is supported by the Italian Ministry of Health (RF-2018–12366631), the European Joint Programme on Rare Diseases (EJPRD20–191), the Italian Association for Cancer Research (IG 2017-ID.19919), and the National Institutes of Health (1R21CA249378–01).

10.1126/science.add6804

METABOLISM

Mitochondria rescue cells from ischemic injury

Activation of a G protein–coupled receptor prevents cardiomyocyte death during ischemia

By **Susana Cadenas**[1,2]

Ischemic diseases, such as myocardial infarction and stroke, are major causes of death worldwide. To date, therapies aimed at preventing ischemic damage or restoring functional tissue have not been successful. Kynurenic acid (KynA), a metabolite of tryptophan metabolism, is tissue protective in cardiac, cerebral, renal, and retinal models of ischemia, but the mechanism of this protection was unknown. On page 621 of this issue, Wyant *et al.* (*1*) report that KynA binds G protein–coupled receptor 35 (GPR35) to activate downstream signaling and, once internalized, indirectly interacts with adenosine triphosphate (ATP) synthase inhibitory factor 1 (IF1) to block ATP depletion during ischemia by promoting the dimerization and inactivation of ATP synthase. This cardioprotective mechanism could be further explored for preventing or treating ischemic injury.

GPR35 expression is up-regulated in the myocardium in human heart failure (*2*), and its expression increases during hypoxia under the control of hypoxia-inducible factor 1 (HIF-1) (*3*), which directs multiple transcriptional programs in response to reduced oxygen availability. Studies investigating the tissue-protective role of GPR35 during ischemia have yielded conflicting results. Using a variety of approaches, Wyant *et al.* provide compelling evidence that, by activating GPR35, KynA is cardioprotective against ischemia-reperfusion (IR) injury. They show that KynA is unable to prevent IR injury in mice lacking GPR35 and in human induced pluripotent stem cell–derived cardiomyocytes expressing a mutant version of the receptor. The cardioprotective effect of KynA is mimicked by several GPR35 agonists in mice.

HIF transcription factors are heterodimers consisting of an oxygen-labile α subunit that is stabilized under hypoxia and a constitutively expressed β subunit. Oxygen-dependent hydroxylation of HIFα by HIF prolyl hydroxylases (PHDs), which are α-ketoglutarate (αKG)–dependent dioxygenases, promotes its ubiquitination and proteasomal degradation (*4*). HIF activation improves ischemia-related outcomes through several mechanisms, the main one being the restoration of oxygen homeostasis mediated by reprogramming metabolism and inducing angiogenesis (blood vessel formation). HIF is also important in ischemic preconditioning (*5*), a phenomenon whereby brief periods of nonlethal ischemia protect against a subsequent prolonged ischemic insult; inhibiting the HIF pathway impairs different types of ischemic preconditioning (*6*, *7*).

A clinically practical approach to improving the outcomes of ischemic disease is remote ischemic preconditioning (RIPC), in which transient ischemia to a nonvital organ protects distant organs. The molecular mediators of RIPC have been sought for many years in the hope that they might be useful to treat cardiovascular disease. The study of Wyant *et al.* expands on the authors' previous work showing that KynA affords cardioprotection in a mouse model of RIPC (*8*). They showed that inhibition of PHD2 leads to HIF activation, which protects mice against cardiac IR injury. Systemic PHD2 loss or specifically that in skeletal muscle increases circulating αKG, driving hepatic production and secretion of KynA, which is necessary and sufficient to mediate cardiac ischemic protection.

More than 95% of the ATP in cardiomyocytes is produced by the inner mitochondrial membrane ATP synthase during oxidative phosphorylation. In normally respiring mitochondria, ATP synthase produces ATP from adenosine diphosphate (ADP) and phosphate using the proton motive force (PMF) generated by electron transfer along the respiratory chain. During ischemia, the PMF is dissipated, and ATP synthase reverses direction and hydrolyzes ATP, contributing to cell death (*9*). This hydrolytic activity is limited by IF1, thus conserving cellular energy. Wyant *et al.* propose that, once activated, GPR35 translocates to the outer mitochondrial membrane, where it indirectly interacts with IF1. Internalized GPR35 inhibits mitochondrial adenylate cyclase [and therefore cyclic adenosine monophosphate (cAMP) production] and consequently protein kinase

[1]Centro de Biología Molecular Severo Ochoa (CBMSO) (CSIC/UAM), Madrid, Spain. [2]Instituto de Investigación Sanitaria del Hospital de la Princesa (IIS-IP), Madrid, Spain. Email: scadenas@cbm.csic.es

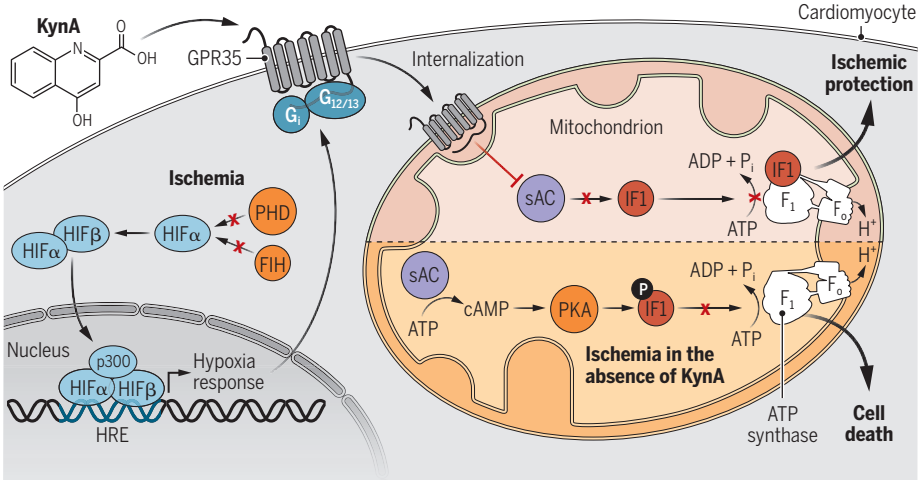
A (PKA), which favors dephosphorylated IF1, the active form of the inhibitor (*10*). They show that IF1 promotes ATP synthase dimerization and prevents ATP loss during ischemia (see the figure).

The binding of IF1 to ATP synthase also affects the mitochondrial permeability transition pore (mPTP), a high-conductance nonspecific channel that is formed in the inner mitochondrial membrane under oxidative stress, leading to depolarization and cessation of ATP synthesis, with catastrophic consequences for energy conservation. The mPTP contains ATP synthase dimers (*11*), and IF1 binding may prevent mPTP opening in the low-pH conditions of ischemia (*12*). Indeed, low pH favors IF1 binding to the enzyme (*13*). Additionally, IF1 is up-regulated in hypertrophied hearts, promoting the formation of inactive ATP synthase tetramers, which triggers mitochondrial reactive oxygen species (ROS) production, stabilizing HIF-1α and activating a metabolic switch that stimulates glycolysis (*14*). It is possible that the inhibition of ATP synthase by IF1 during ischemia also initiates ROS signaling, which may affect cardiomyocyte metabolism.

The study of Wyant *et al.* supports further examination of KynA and, more broadly, GPR35 agonists, including synthetic compounds with greater potency than KynA, for preventing or treating ischemic injury. Their findings also provide a molecular explanation for the tissue-protective effects of kynurenine monooxygenase inhibitors, which promote accumulation of KynA and thus have cardioprotective potential. Similarly, preclinical studies with IF1 mimetics have generated promising results in models of cardiac ischemia (*9*). GPR35 activation by KynA stimulates lipid metabolism and thermogenic and anti-inflammatory gene expression, increasing energy expenditure in adipose tissue (*15*). Accordingly, can KynA induce a transcriptional program that affects cardiomyocyte metabolism? Elucidating the pathways orchestrating ischemic protection and the development of therapeutics that modulate them has important clinical implications for a patient population with historically poor outcomes. ■

Protecting against cardiac ischemia

Upon ischemia, G protein–coupled receptor 35 (GPR35) expression is induced by hypoxia-inducible factor (HIF), which then binds kynurenic acid (KynA). GPR35 is internalized to the outer mitochondrial membrane, where it inhibits mitochondrial adenylate cyclase (sAC), which in turn prevents phosphorylation (P) of ATP synthase inhibitory factor 1 (IF1) by protein kinase A (PKA). Dephosphorylated IF1 leads to ATP synthase dimerization, preventing ATP hydrolysis and promoting energy conservation.



ADP, adenosine diphosphate; ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate; FIH, factor inhibiting HIF; HRE, hypoxia response element; PHD, prolyl hydroxylase; P_i, inorganic phosphate.

REFERENCES AND NOTES

1. G. A. Wyant *et al.*, *Science* **377**, 621 (2022).
2. K.-D. Min *et al.*, *Biochem. Biophys. Res. Commun.* **393**, 55 (2010).
3. V.-P. Ronkainen *et al.*, *Cardiovasc. Res.* **101**, 69 (2014).
4. P. Jaakkola *et al.*, *Science* **292**, 468 (2001).
5. S. Cadenas, *Free Radic. Biol. Med.* **117**, 76 (2018).
6. Z. Cai *et al.*, *Cardiovasc. Res.* **77**, 463 (2008).
7. Z. Cai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17462 (2013).
8. B. A. Olenchock *et al.*, *Cell* **164**, 884 (2016).
9. G. J. Grover *et al.*, *Am. J. Physiol. Heart Circ. Physiol.* **287**, H1747 (2004).
10. J. García-Bermúdez *et al.*, *Cell Rep.* **12**, 2143 (2015).
11. A. Urbani *et al.*, *Nat. Commun.* **10**, 4341 (2019).
12. P. Bernardi *et al.*, *Physiol. Rev.* **95**, 1111 (2015).
13. V. Boreikaite *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10354 (2019).
14. B. Zhou *et al.*, *J. Clin. Invest.* **132**, e155333 (2022).
15. L. Z. Agudelo *et al.*, *Cell Metab.* **27**, 378 (2018).

ACKNOWLEDGMENTS

S.C.'s laboratory is funded by Instituto de Salud Carlos III grant no. PI19/01030 (Spanish Ministry of Science and Innovation) and by institutional grants from the Banco de Santander and Fundación Ramón Areces to the CBMSO.

10.1126/science.add4629

CARBENES

Safe, selective, and scalable carbenes

The synthesis of reactive carbene intermediates is made simpler and safer

By **Michael S. West** *and* **Sophie A. L. Rousseaux**

Across chemistry, new strategies to access transient and reactive intermediates are of interest for further research and synthetic applications. In organic chemistry, carbenes are one such intermediate. These include any neutral carbon molecule with two unshared valence electrons (*1*, *2*). Carbenes can participate in a variety of reactions such as dimerization, formation of three-membered rings, and bond insertions. However, accessing these valuable intermediates has traditionally required the use of precursor molecules that are unstable or are limited to specific types of reactions (*3*, *4*). For example, nonstabilized alkyl-substituted carbenes are highly reactive intermediates, which limits their use in synthetic transformations (*3*). On page 649 of this issue, Zhang *et al.* (*5*) report a means to access both stabilized and nonstabilized carbenes under safe and simple reaction conditions.

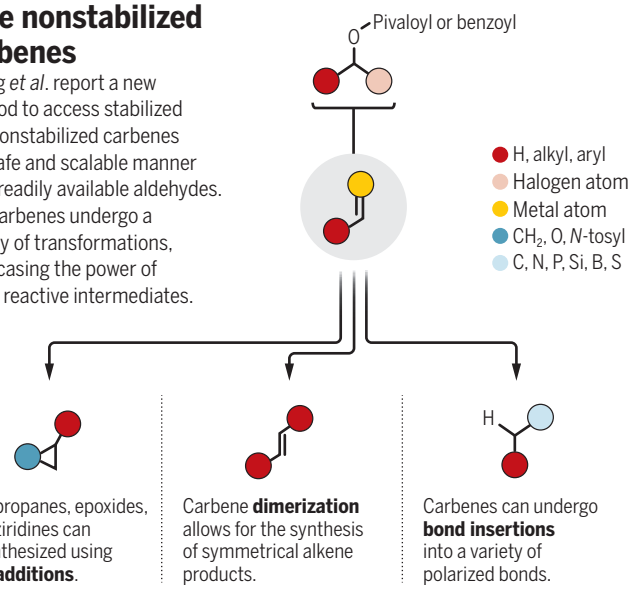
A common strategy to generate carbenes is through the use of diazoalkanes, which are reactive carbon-based molecules that contain a diazo functional group (i.e., have two linked nitrogen atoms) that releases nitrogen upon carbene formation. These can undergo a variety of different reactions when paired with a metal catalyst (*6*). The diazoalkane starting materials can be substituted with both electron-accepting and electron-donating groups, which can modulate the electronic properties—and, subsequently, the reactivity profile—of the generated carbene. In all cases, the release of nitrogen during the reaction leads to important safety concerns when handling these thermally unstable molecules, because large volumes of colorless and odor-

Davenport Research Laboratories, Department of Chemistry, University of Toronto, Toronto, ON, Canada. Email: sophie.rousseaux@utoronto.ca

GRAPHIC: K. HOLOSKI/*SCIENCE*

Safe nonstabilized carbenes

Zhang *et al.* report a new method to access stabilized and nonstabilized carbenes in a safe and scalable manner from readily available aldehydes. The carbenes undergo a variety of transformations, showcasing the power of these reactive intermediates.



less nitrogen can displace oxygen and cause suffocation to unsuspecting bystanders (*7*). As a result, their use in industrial-scale reactions is severely limited (*8*). Another common method to generate metal carbenoids (i.e., reactive intermediates with carbene-like characteristics) is the reaction of metals, such as zinc, with *gem*-dihaloalkanes, which are carbon-based molecules that contain two halogen substituents on the same carbon atom (*9*). Although this method is safer, the reactivity of the resulting zinc carbenoids is limited to cyclopropane formation (i.e., cyclopropanation). The preparation of zinc carbenoids can also suffer from detrimental side reactions if the starting material is an alkyl-substituted *gem*-dihaloalkane (*10*, *11*), which further limits their application.

Owing to these limitations, there is a need for methods to make carbenes from stable precursors in a safe manner. Driven by their prior work in developing methods to access reactive intermediates (*12*, *13*), Zhang *et al.* demonstrate the formation of aryl- and alkyl-substituted carbenes by starting with stable and scalable acetate-containing organohalide precursors, which themselves can be derived from simple aldehydes (*12*). By pairing these carbenes with appropriate metal catalysts, their reactivity can be controlled for a variety of synthetic transformations (see the figure). This generalized approach avoids the outstanding safety concerns of diazoalkanes and the stability and selectivity limitations of *gem*-dihaloalkanes.

Diazoalkanes can decompose near 100°C, where the energy produced upon decomposition can be more than 200 kJ/mol (*10*). This thermal instability makes scale-up reactions with these reagents very dangerous. With aldehyde-derived carbene precursors, Zhang *et al.* observed stability at >300°C and that less than 30 kJ/mol of energy was produced upon decomposition; the output of nitrogen was also avoided. As a result, a cyclopropanation reaction using these carbene reagents can be scaled up 1000-fold without incident (*10*). This method can also be used to generate an unsubstituted carbene precursor from formaldehyde in kilogram quantities, thereby avoiding the use of the notoriously dangerous diazomethane (*14*, *15*).

Besides addressing safety and stability concerns, selectivity and versatility in terms of reactivity are also important considerations in the development of carbene precursors. Zhang *et al.* show that the carbene precursors are compatible in more than 10 different types of reactions. Using inexpensive base metal catalysts, such as copper(I) chloride (CuCl) and cobalt(II) chloride ($CoCl_2$), alkene products were selectively produced through carbene dimerization. This worked well in the presence of different functional groups. Notably, nonstabilized alkyl-substituted carbenes, which are typically too reactive to achieve good yields in synthetic reactions, were also compatible under the reaction conditions with no problematic side reactions observed (*12*). Tuning carbene reactivity with other catalysts, such as rhodium(II) acetate [$Rh_2(OAc)_4$] and iron(II) chloride ($FeCl_2$), enabled even more chemical reactivity, resulting in cyclopropanation reactions of both electron-rich and electron-deficient alkenes with the use of a sulfide cocatalyst. Other three-membered rings such as epoxides and aziridines can also be generated from the corresponding aldehydes and imines under this diazo-free, catalytic strategy. In addition to dimerization and cyclopropanation reactions, another common reaction pathway for carbenes is insertion into polarized heteroatom-hydrogen bonds. Both stabilized aryl- and nonstabilized alkyl-carbenes generated from these precursors succeeded in this reactivity as well, undergoing efficient insertion into a wide variety of polarized bonds—for example, nitrogen-hydrogen, phosphorus-hydrogen, silicon-hydrogen, boron-hydrogen, and sulfur-hydrogen.

"The straightforward synthesis, stability, scalability, and selectivity of the carbene precursors reported by Zhang *et al.* will provide chemists with a better means for achieving interesting carbene applications in synthesis."

The straightforward synthesis, stability, scalability, and selectivity of the carbene precursors reported by Zhang *et al.* will provide chemists with a better means for achieving interesting carbene applications in synthesis. The authors have provided a straightforward method to generate these valuable stabilized and nonstabilized reactive intermediates from aldehydes by eliminating long-standing safety and reactivity challenges. Academic and industrial chemists who may have previously been hesitant to explore carbene chemistry because of the associated hazards and challenging synthesis can now try their hand with these intermediates, which are generated from simple building blocks. ■

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES AND NOTES

1. P. de Frémont, N. Marion, S. P. Nolan, *Coord. Chem. Rev.* **253**, 862 (2009).
2. V. H. Gessner, *Chem. Commun.* **52**, 12011 (2016).
3. D. Zhu, L. Chen, H. Fan, Q. Yao, S. Zhu, *Chem. Soc. Rev.* **49**, 908 (2020).
4. H. M. Davies, R. E. J. Beckwith, *Chem. Rev.* **103**, 2861 (2003).
5. L. Zhang, B. M. DeMuynck, A. N. Paneque, J. E. Rutherford, D. A. Nagib, *Science* **377**, 649 (2022).
6. A. G. H. Wee, *Curr. Org. Synth.* **3**, 499 (2006).
7. J. H. Simpson *et al.*, in *Managing Hazardous Reactions and Compounds in Process Chemistry*, J. A. Pesti, A. F. Abdel-Magid, Eds. (American Chemical Society, 2014), chap. 9, pp. 235–244.
8. S. P. Green *et al.*, *Org. Process Res. Dev.* **24**, 67 (2020).
9. M. Pasco, N. Gilboa, T. Mejuch, I. Marek, *Organometallics* **32**, 942 (2013).
10. W. B. Motherwell, *J. Organomet. Chem.* **624**, 41 (2001).
11. J. Werth, C. Uyeda, *Angew. Chem. Int. Ed.* **57**, 13902 (2018).
12. L. Wang, J. M. Lear, S. M. Rafferty, S. C. Fosu, D. A. Nagib, *Science* **362**, 225 (2018).
13. S. M. Rafferty, J. E. Rutherford, L. Zhang, L. Wang, D. A. Nagib, *J. Am. Chem. Soc.* **143**, 5622 (2021).
14. T. J. de Boer, H. J. Backer, *Org. Synth.* **36**, 16 (1956).
15. B. Morandi, E. M. Carreira, *Science* **335**, 1471 (2012).

ACKNOWLEDGMENTS

The authors acknowledge the University of Toronto and the Natural Sciences and Engineering Research Council of Canada (PGS-D scholarship for M.S.W.) for supporting this work.

10.1126/science.abq8253

BOOKS *et al.*

EVOLUTIONARY BIOLOGY

Why we do what we do

From regenerating sea slugs to self-medicating sheep, a biologist probes the origins and evolution of behavior

By **Rob Dunn**

The spider-tailed horned viper's prey have helped shape the evolution of the snake's remarkably lifelike lure.

PHOTO: F. DESCHANDOL AND P. SABINE/MINDEN PICTURES

Many of us have become estranged from the rest of the living world. We live inside and, in those rare moments outside, are as likely to be annoyed by other species as inspired. Just this week, one of my neighbors asked how to kill the chipmunks in his yard, another how to get rid of rabbits, and still another wanted "a solution" for the robins that were "flinging mulch into his driveway in their quest for worms." Even when wild animals show themselves, we seem ill-prepared to pay attention.

In her new book, *Dancing Cockatoos and the Dead Man Test: How Behavior Evolves and Why It Matters*, Marlene Zuk documents the behavior of the animals around us on our behalf. She watches and writes with a sense of wonder, curiosity, and the abiding recognition that our own human lives only make sense in light of the behavior of other species.

In this framing, Zuk implicitly advances what to some will be a radical idea: that humans are just another animal species. We may be unusual, and hence "special," in some of our behaviors, but so too, she argues, is the sea slug that abandons its body when attacked by parasites only to grow a new one from its disembodied head. To Zuk, an animal behaviorist, the world is full of species with individual stories, species that teach about life's general rules and exceptions, and species from which we can and should learn.

Are behaviors governed by nature or nurture? Zuk comes down in favor of a third perspective in the book's opening chapter. Genes, she argues, influence behavior, but how they do so depends on the environment. Similarly, the environment influences behavior, but how it does so depends on the genes. Here, as in every subsequent chapter, Zuk makes her point through the stories of a cavalcade of one species after another (including, eventually, humans). Then, she introduces exceptions and caveats, critiques, and counterarguments. One has the feeling of being inside a mind that is constantly weighing every perspective, ever aware that the work of making sense of the world is far from done.

Each chapter considers a general concept or a question. The second chapter explores how behavior evolves and the third how behavior is inherited. The fourth documents the surprisingly complex story of the evolution of the behavior of dogs. Later chapters describe the evolution of the behavior of domestic animals more generally, the evolution of human language, mental health among animals (including humans), big brains, sex, sex roles, and disease.

Dancing Cockatoos and the Dead Man Test
Marlene Zuk
Norton, 2022. 352 pp.

In the book's timely final chapter, Zuk describes how the behaviors of animals can be influenced by disease. Some fungi, for example, take over the bodies of ants and cause them to climb high onto trees where the fungi can more readily disperse. Jewel wasps, meanwhile, inject venom into cockroaches that makes them easy to control and coax home (as food for their babies).

The book also considers those cases in which animals' behaviors help them to avoid disease. Chimpanzees self-medicate by eating plants that help to kill their intestinal parasites, as do goats and sheep. Some populations of house sparrows bring cigarette butts into their nests to kill ticks. Ants gather antimicrobial resins and incorporate them into their mounds.

After exploring these examples, Zuk reminds us of very similar phenomena in humans that are neither fully genetic nor fully environmental in origin. Like ants, we can be manipulated by microbes. Humans who received the flu vaccine, for example, have been shown to be much more likely to be social (without realizing it) in the 48 hours after the shot than they had been beforehand, behavior that could potentially benefit a live virus. Like chimpanzees, humans use plants as medicines. Like many animals, humans employ social distancing in the presence of parasites. And like one African ant species, humans use a mix of techniques to wash pathogens off their bodies to reduce the risk of infection.

In the end, Zuk's lovely book feels like a cabinet of curiosities whose details remind us to pay attention to the behaviors around us every day. Don't try to kill the chipmunk, the rabbit, or the robin. Watch them. Learn from them, and in doing so, learn also about yourself. Our story, however important it feels to us, is just one among millions. ■

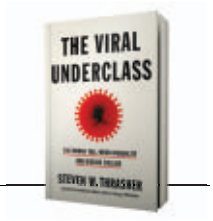
The reviewer is a professor in the Department of Applied Ecology, North Carolina State University, Raleigh, NC, USA, where he is also the senior vice provost for University Interdisciplinary Programs, and is the coauthor of *Delicious: The Evolution of Flavor and How It Made Us Human* (Princeton Univ. Press, 2021).
Email: rrdunn@ncsu.edu

10.1126/science.add5306

Inequality goes viral

Structural factors exacerbate the impact of viruses

The Viral Underclass: The Human Toll When Inequality and Disease Collide
Steven W. Thrasher
Celadon Books, 2022. 352 pp.



By **Ayah Nuriddin**

Michael Johnson was an undergraduate student and athlete at Lindenwood University near St. Louis, Missouri, when he was arrested for "recklessly" transmitting HIV to five men in 2013. Throughout his trial, Johnson was labeled as predatory, hypersexualized, and a vector of disease. His identity as a gay Black man was put on trial as well. Johnson refused to plead guilty, arguing that he had notified his partners of his HIV status and had therefore not committed a crime. He was convicted and sentenced to 30 years in prison, but his conviction was eventually overturned and his sentence vacated because it was found that the prosecution had withheld evidence. He was released in 2019.

Johnson, according to Steven Thrasher, professor of journalism at Northwestern University, is a member of the "viral underclass"—those individuals who bear the disproportionate impact of virus transmission, morbidity, and mortality because of structural inequality. Borrowing the term from HIV activist Sean Strub, in his new book, *The Viral Underclass*, Thrasher argues that viruses amplify existing power structures and exacerbate the impact of viral infection on people of color, the poor, the disabled, and those who identify as LGBTQ+.

The viral underclass is a global phenomenon, but, according to Thrasher, it is deeply connected to "the particular cruelties of the American empire." Thrasher's beautifully written account illustrates the complex and textured relationship between disease and inequality, building on decades of scholarship and activism from Black, brown, queer, and disabled people who understood that the disproportionate burden of disease is the result of the structural nature of health inequality.

The book focuses primarily on two viruses—HIV and COVID-19—identifying 12 interconnected social vectors that lead to unequal virus transmission: racism, individualized shame, capitalism, the law, austerity, borders, the liberal carceral state, unequal prophylaxis, ableism, speciesism, the myth of white immunity, and collective punishment. Each vector receives its own chapter in which Thrasher relates the personal stories of individuals who have experienced the burden of systemic viral inequality, providing a lens into the different and devastating ways that it produces the viral underclass and the beautiful and heartbreaking ways that individuals and communities survive, resist, and find joy despite marginalization.

A virus can have different social consequences, depending on which vectors it intersects with. The book's first chapter, "Mandingo: Racism," opens with the murder of George Floyd and then focuses on Michael Johnson's story. "From Athens to Appalachia: Austerity" centers on the life and brutal murder of Greek drag queen and HIV activist Zak Kostopoulos amid austerity measures and subsequent police violence in Greece and shows readers how austerity measures exacerbate suffering for those disproportionately affected by viral inequality. "Borderlands: Borders" describes the vulnerability of being trans and undocumented through the life of trans HIV activist Lorena Borjas, considered "the mother of the trans Latinx community" in Queens, New York, who ultimately died alone of COVID-19. Together, these and other stories are woven into a rich tapestry that narrates the impact of inequality on human experience.

Volunteers conduct COVID-19 outreach in Rochester, New York, in 2020.

PHOTO: LINDSAY DEDARIO/REUTERS

In Thrasher's analysis, viruses often serve as a proxy for disease as a category. The book would have been enhanced, however, if he had drilled down into the particularities of specific viruses—polio, influenza, human papillomavirus, or monkeypox, for example—and how each reveals or obscures different dimensions of inequality. Likewise, including other agents of disease might have further enriched the book's arguments. A microbial underclass or an infectious underclass could expand a theory of the viral underclass to illustrate the unequal and disproportionate impacts of conditions such as tuberculosis, malaria, or asthma.

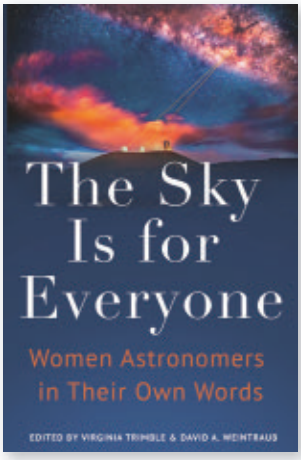
Thrasher's theory of the viral underclass creates opportunities to imagine new possibilities for a better world, or, as he writes, it "provides a map for a kind of worldwide liberation." Identifying and addressing the vectors that produce the viral underclass would not only improve unequal virus transmission, it "would improve life on earth for nearly everyone."

Dismantling structural viral inequality requires recognition that the vectors that lead to it are deeply interconnected. Solutions will require us to imagine our destinies as similarly intertwined. ■

The reviewer is at the Society of Fellows and the Department of African American Studies, Princeton University, Princeton, NJ 08544, USA. Email: an4984@princeton.edu

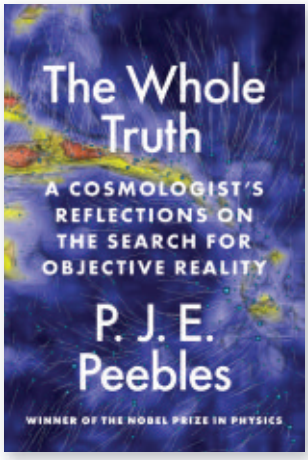
10.1126/science.add5428

New from Princeton



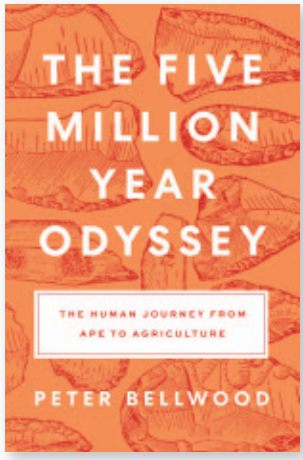
"An inspiring anthology of writings by trailblazing female astronomers from 1960 to today."

—Simon Ings, *New Scientist*



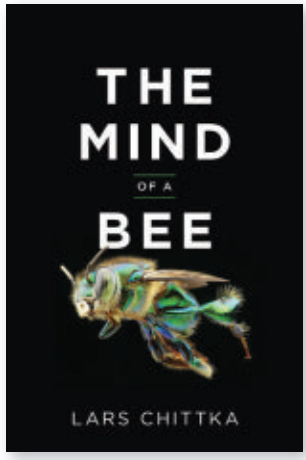
"An engaging and inviting account of the history of the physical sciences."

—Grace O'Hanlon, *Library Journal*



"An entertaining introduction to deep human history."

—Rebecca Wragg Sykes, author of *Kindred*



"The knowledge on offer here is as entertaining as it is edifying. Readers won't look at bees the same way again."

—*Publishers Weekly*

PRINCETON UNIVERSITY PRESS

AN ESTATE GIFT TO THE American Association for the Advancement of Science

1848 SOCIETY

AAAS

AAAS is 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.

You can help us.

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 – 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.

If you plan 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

For more information or to let us know you are including AAAS in your future giving plans, contact the AAAS Office of Philanthropy and Strategic Partnerships at **1-800-215-1969** or **philanthropy@aaas.org.**

Additional details online at **www.aaas.org/1848Society.**

LETTERS

Mature common beech trees are among the species that can improve urban quality.

Edited by **Jennifer Sills**

Ozone-reducing urban plants: Choose carefully

Ozone (O_3) pollution is a threat to human health, vegetation, biodiversity, and climate (*1*). Millions of urban citizens face O_3 levels above the World Health Organization standards, leading to nearly 150,000 deaths worldwide in 2019 (*2*). By 2050, 70% of the world population will reside in cities (*3*). Policies are urgently needed to reduce O_3 levels and prevent further deaths. Urban plants are one promising strategy (*4*).

Plants can reduce air pollution by adsorbing pollutants on their surfaces or absorbing them through their leaves or needles (*5*). Green urban infrastructure can mitigate O_3 pollution, and mature urban trees show higher O_3-removal capacity than, for example, green roofs, shrubs, or green walls (*6*). However, not all trees are the same—to maximize benefits, tree species should be selected based on their O_3-removal capacity. Top O_3-reducing tree species include common beech, small- and large-leaved lime, London plane, sycamore maple, Norway maple, tulip tree, horse chestnut, and turkey oak (*6*, *7*). Conversely, some species form more O_3 than they remove, such as blue gum, Italian and pubescent oak, spruce, hazel pine, weeping willow, and common myrtle (*6*, *7*). Generally, broadleaf tree species remove more O_3 than conifers (*8*). Worldwide urban greening programs should make sure to select species that improve air quality.

To maximize green O_3 removal, urban strategies should also consider environmental conditions such as meteorology, soil, and air quality and select the species best adapted to the region. Every region should find O_3-reducing species that are also long lived and low maintenance, with the greatest total leaf area possible. Although planting trees will not be able to counterbalance all anthropogenic O_3 pollution, choosing the right species can maximize the benefits of this valuable air quality strategy.

Pierre Sicard[1]*, Evgenios Agathokleous[2], Alessandra De Marco[3], Elena Paoletti[4]
[1]ARGANS, Sophia Antipolis, France. [2]School of Applied Meteorology, Nanjing University of Information Science and Technology, Nanjing, China. [3]National Agency for New Technologies, Energy, and Sustainable Economic Development, Rome, Italy. [4]Research Institute on Terrestrial Ecosystems–National Research Council, Sesto Fiorentino, Italy.
*Corresponding author. Email: psicard@argans.eu

REFERENCES AND NOTES

1. E. Agathokleous *et al.*, *Sci. Adv.* **6**, eabc1176 (2020).
2. D. A. Malashock *et al.*, *Environ. Res. Lett.* **17**, 054023 (2022).
3. United Nations, "The world's cities in 2016" (2018).
4. R. Pace *et al.*, *Int. J. Biometeorol.* **65**, 277 (2021)
5. J. M. Wedow *et al.*, *Trends Biochem. Sci.* **46**, 992 (2021).
6. P. Sicard *et al.*, *Environ. Pollut.* **243**, 163 (2018).
7. Y. Ren *et al.*, *Environ. Pollut.* **230**, 849 (2017).
8. A. N. Muresan *et al.*, *Ecol. Indic.* **135**, 108554 (2022).

10.1126/science.add9734

Climate change threatens Pakistan's snow leopards

The snow leopard (*Panthera uncia*) population currently spans the mountainous regions of 12 countries, including more than 80,000 square kilometers in northern Pakistan (*1*). As a result of human encroachment and hunting, snow leopards are classified as Vulnerable (*2*). Climate change is exacerbating the threats the snow leopard already faces as well as transforming their environment in ways that make survival more difficult. Pakistan must take urgent action to protect this important species.

Since the partition of India in 1947, the human population in the snow leopard's range in Pakistan has quadrupled, and livestock numbers have increased by 40 to 60% (*3*). Habitat loss driven by human activities and economic growth has fragmented the snow leopard's habitat. Overgrazing has caused severe degradation of grasslands, and an overabundance of livestock has encroached on the habitat of wild ungulates, decreasing the prey available to snow leopards (*4*).

Growing human populations and livestock in the snow leopard's habitat have led to increased illegal trafficking, poaching, and human-wildlife conflict. On average, one snow leopard per day has been lost to poaching during the past decade (*5*). Snow leopards are also frequently

PHOTO: DAVE PRESSLAND/MINDEN PICTURES

killed in reprisal for attacks on local herders' livestock (*6*).

In addition to these harmful trends, global warming is causing the forest line in mountainous regions to move to higher altitudes, compressing the suitable habitat available to snow leopards. Changing climate has also allowed leopards and other low-altitude carnivores to migrate to higher altitudes, increasing the snow leopard's competition (*7*). Furthermore, reduced glacier cover will lead to increased drought risk, changes to local flora and fauna, and fewer food alternatives for species like the snow leopard (*8*).

The snow leopard is a top predator and an indicator of the overall health of its high-altitude habitat (*9*). To protect the species, Pakistan must mitigate the effects of climate change and human settlements (*10*). Priorities for climate-informed conservation should include passing legislation to prevent poaching and retaliatory killings. Predator-proof livestock cages could decrease the chances of human-wildlife conflict. Helping people in the snow leopard's range adapt to the effects of climate change will also benefit wildlife, as will continuing efforts to minimize habitat degradation. Training and education could help tourists and the public avoid the species' habitat. Researchers should learn more about the species' biology, especially its genetics and the diseases to which it is susceptible. To monitor progress, Pakistan should track the number of snow leopards and look for trends in population size and distribution.

Uzair Aslam Bhatti[1,3], Mir Muhammad Nizamani[2], Huang Mengxing[1,3]*
[1]School of Information and Communication Engineering, Hainan University, Haikou 570100, China. [2]Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou 570228, China. [3]State Key Laboratory of Marine Resource Utilization in the South China Sea, Hainan University, Haikou 570100, China.
*Corresponding author.
Email: huangmx09@163.com

REFERENCES AND NOTES

1. W. Rashid *et al.*, *Sustainability* **13**, 13229 (2021).
2. R. K. Sharma, Y. V. Bhatnagar, C. Mishra, *Biol. Conserv.* **190**, 8 (2015).
3. T. U. Khan *et al.*, *Sustainability* **13**, 3590 (2021).
4. K. R. Suryawanshi *et al.*, *R. Soc. Open Sci.* **4**, 170026 (2017).
5. Snow Leopard Trust, The Threats (2022); https://snowleopard.org/the-threats/.
6. G. Mohammad *et al.*, in *Snow Leopards*, (Academic Press, 2016), pp. 179–195.
7. J. D. Farrington, J. Li, in *Snow Leopards*, T. McCarthy, D. Mallon, P. J. Nyhus. Eds. (Academic Press, 2016), pp. 85–95.
8. T. M. McCarthy, G. Chapron, Eds., "Snow leopard survival strategy," (International Snow Leopard Trust and Snow Leopard Network, 2014).
9. J. Li, L. Xiao, Z. Lu, *Sci. China Life Sci.* **59**, 637 (2016).
10. E. Distefano, "Human-wildlife conflict worldwide: Collection of case studies, analysis of management strategies and good practices" (Food and Agricultural Organization of the United Nations, Sustainable Agriculture and Rural Development Initiative, Rome, 2005).

10.1126/science.add9065

Primary forest loss in biodiverse Indian states

Primary forests—old-growth forests that have remained undisturbed by humankind—comprise one-third of the world's forests. These dense, wild forests are important habitats for unique species (*1*) and provide a variety of ecosystem services, including global biosphere-atmosphere CO_2 exchange (*2*). Such forests are irreplaceable in terms of biodiversity value and ecosystem services, and replantation to compensate for such forest loss is inadequate (*3*). Yet loss of primary forests continues all over the world, and governments have failed to protect them (*4*). For example, in India, primary forest is being lost in biodiversity hotspots, which should be environmental priorities. India's government must acknowledge the value of primary forests and prioritize their conservation by committing to holistic conservation strategies.

India is home to four global biodiversity hotspots, including the Himalayas and the Indo-Burma region. These areas, with high altitudes and unique climates, support a variety of ecosystems and harbor distinct species diversity and endemism (*5*, *6*). However, nine states in the Himalayas and Indo-Burma regions accounted for 89% of total primary forest loss in India between 2015 and 2021 (*7*). Between 2002 and 2021, India saw a 3.6% reduction in the total area covered by humid primary forest (*7*).

The ongoing primary forest loss will have detrimental effects on global environment and biodiversity. In these high-altitude areas, the rate of regional warming has exceeded the rate of global warming (*8*). The massive forest loss, climate extremes, increased temperatures, and altered rainfall pattern in these biodiversity hotspots intensify competition for species survival, leading to both species extinction and changes in species distribution (*9*, *10*). Deforestation in the Himalayas alone is expected to cause the extinction of a quarter of endemic species by 2100, including 366 vascular plant taxa and 35 vertebrate taxa (*11*).

Scientists, activists, and the public must persuade local and national governments to pass legislation and implement holistic strategies that halt primary forest loss. Sustainable forest conservation strategies include community conserved forest management, in which the residing communities (mostly native and tribal peoples) are given rights to the forest, and strict prohibition of forest land diversion for nonforestry purposes, including infrastructure. It is time to secure the long-term integrity of these irreplaceable primary forests.

Suvarna Khadakkar
Society for Ecology, Evolution and Development, Wardha 442001, India.
Email: suvarnask17@gmail.com

REFERENCES AND NOTES

1. J. Barlow *et al.*, *Proc. Natl. Acad. Sci.* **104**, 18555 (2007).
2. C. Cleveland *et al.*, *Ecol. Lett.* **14**, 1313 (2011).
3. L. Gibson *et al.*, *Nature* **478**, 378 (2011).
4. *Nat. Ecol. Evol.* **6**, 653 (2022).

NEXTGEN VOICES: SUBMIT NOW

When internships disappoint

Add your voice to *Science*! In this NextGen Voices survey, a reader asks for your advice! Have you been in this situation or one like it? Do you have any tips that you would like to share? Become a NextGen Voices peer mentor by contributing your thoughts!

Dear NextGen Voices,

I am an undergraduate student trying to get exposure to research in a lab. I found an internship, but the graduate students and even postdocs seem to resent the idea of an undergraduate in the lab. Some have said that I take up valuable time and resources that should be devoted to more experienced lab members. I am often excluded from even informal conversations and activities. This unwelcoming culture took me by surprise and has decreased my motivation to pursue science. What should I do to improve my current situation, and how can I learn from this experience when making future career decisions?

Sincerely,
Unwelcome Undergrad

To submit, go to www.science.org/nextgen-voices
Deadline for submissions is 19 August. A selection of the best responses will be published in the 7 October issue of *Science*. Anonymous submissions will not be considered.

5. R. A. Mittermeier *et al.*, in *Biodiversity Hotspots* (Springer, 2011), pp. 3–22.
6. G. Ellepola *et al.*, *Commun. Biol.* **5**, 347 (2022).
7. Global Forest Watch, (World Resources Institute, 2014); www.globalforestwatch.org. To search for primary forest lost data, click on "Dashboard" on the upper bar. Then type "India" in "Select country" dropdown. In the "PRIMARY FOREST LOSS IN INDIA" graph (which notes that 3.6% of primary forest cover was lost between 2002 and 2021), click on the "Download" icon in the upper right corner of the graph. In the spreadsheet titled "treecover_loss_in_primary_forests_2001_tropics_only__ha," the "umd_tree_cover_loss__ha" column shows the annual primary forest loss in India, totaling 161,000 ha from 2015 to 2021. To find primary forest loss data for each Himalayan/Indo-Burma state, click on the "Select a region" dropdown below "India" on the site and type the state name. For each state, click the "Download" icon in the "PRIMARY FOREST LOSS IN [state name]" graph. In the spreadsheet titled "treecover_loss_in_primary_forests_2001_tropics_only__ha," the "umd_tree_cover_loss__ha" column shows the annual primary forest loss in the state. The sum of the forest loss, rounded to the nearest hectare, between 2015 and 2021 in each available Himalayan/Indo-Burma state is 18 (Uttarakhand), 87 (Sikkim), 23,179 (Meghalaya), 23,430 (Assam), 69 (Tripura), 9033 (Mizoram), 23,992 (Manipur), 13,380 (Nagaland), and 50,567 (Arunachal Pradesh), for a total of 143,755, representing 89.28% of the total forest loss in India during that time.
8. P. Chaudhari, K. S. Bawa, *Biol. Lett.* **7**, 767 (2011).
9. M. B. Ashcroft, *J. Biogeogr.* **37**, 1407 (2010).
10. V. S. Chitale *et al.*, *PLOS One* **9**, 12 (2014).
11. M. K. Pandit *et al.*, *Biodiver. Conserv.* **16,** 1 (2007).

10.1126/science.add8815

TECHNICAL COMMENT ABSTRACTS

Comment on "Pushing the frontiers of density functionals by solving the fractional electron problem"

Igor S. Gerasimov *et al*.

Kirkpatrick *et al*. (Reports, 9 December 2021, p. 1385) trained a neural network–based DFT functional, DM21, on fractional-charge (FC) and fractional-spin (FS) systems, and they claim that it has outstanding accuracy for chemical systems exhibiting strong correlation. Here, we show that the ability of DM21 to generalize the behavior of such systems does not follow from the published results and requires revisiting.

Full text: dx.doi.org/10.1126/science.abq3385

Response to Comment on "Pushing the frontiers of density functionals by solving the fractional electron problem"

James Kirkpatrick *et al*.

Gerasimov *et al*. claim that the ability of DM21 to respect fractional charge (FC) and fractional spin (FS) conditions outside of the training set has not been demonstrated in our paper. This is based on (i) asserting that the training set has a ~50% overlap with our bond-breaking benchmark (BBB) and (ii) questioning the validity and accuracy of our other generalization examples. We disagree with their analysis and believe that the points raised are either incorrect or not relevant to the main conclusions of the paper and to the assessment of general quality of DM21.

Full text: dx.doi.org/10.1126/science.abq4282

Where Science Gets Social.

AAAS.ORG/COMMUNITY

AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

PRIZE ESSAY

GRAND PRIZE WINNER

Robert M. G. Reinhart

Robert Reinhart received his PhD from Vanderbilt University, forwent postdoctoral training, and accepted an appointment as assistant professor at Boston University in 2016, where he is currently director of the Cognitive and Clinical Neuroscience Laboratory and on faculty in the Departments of Psychological & Brain Science and Biomedical Engineering. His research combines electrophysiology, noninvasive neuromodulation, and behavior to understand the large-scale mechanisms of human cognition in health and disease and to develop drug-free, neuroscience therapeutics to help people with brain disorders.
www.science.org/doi/10.1126/science.add4834

FINALIST

Davide Folloni

Davide Folloni received his undergraduate degree from the University of Parma, a Master's degree in clinical psychology from the University of Padua, and a Master's degree and PhD in neuroscience from the University of Oxford. His research focuses on the development of multimodal approaches that combine neuromodulation, electrophysiological, and imaging techniques to understand the neural mechanisms underlying learning and decision-making, with the primary aim being to describe how neuronal oscillations guide our adaptation to the world.
www.science.org/doi/10.1126/science.add4836

NEUROMODULATION

Synchronizing neural rhythms

Personalized, noninvasive network-based neuromodulation aids impaired cognition

By **Robert M. G. Reinhart**[1,2]

Cognitive brain disorders are among the most disabling health states in the world. Existing pharmacological, surgical, and behavioral therapeutic approaches for impaired cognition are limited by heterogeneous treatment outcomes, slow symptom resolution, and accompanying risks and side effects. Consequently, there is an urgent need to develop innovative and personalized therapeutic interventions that are capable of providing rapid and sustainable improvements with minimal side effects. Noninvasive neuromodulation technology is an emerging class of tools that offer such translational potential for neurocognitive disorders. Tools such as high-definition transcranial alternating current stimulation (HD-tACS) offer unprecedented control in modulating rhythmic activity in cortical regions that are implicated in neurocognitive dysfunction (*1*). Our latest research in precision or personalized HD-tACS shows promise at noninvasively manipulating neural population dynamics and improving human cognition and adaptive behavior in a selective and long-lasting fashion (*2*–*5*).

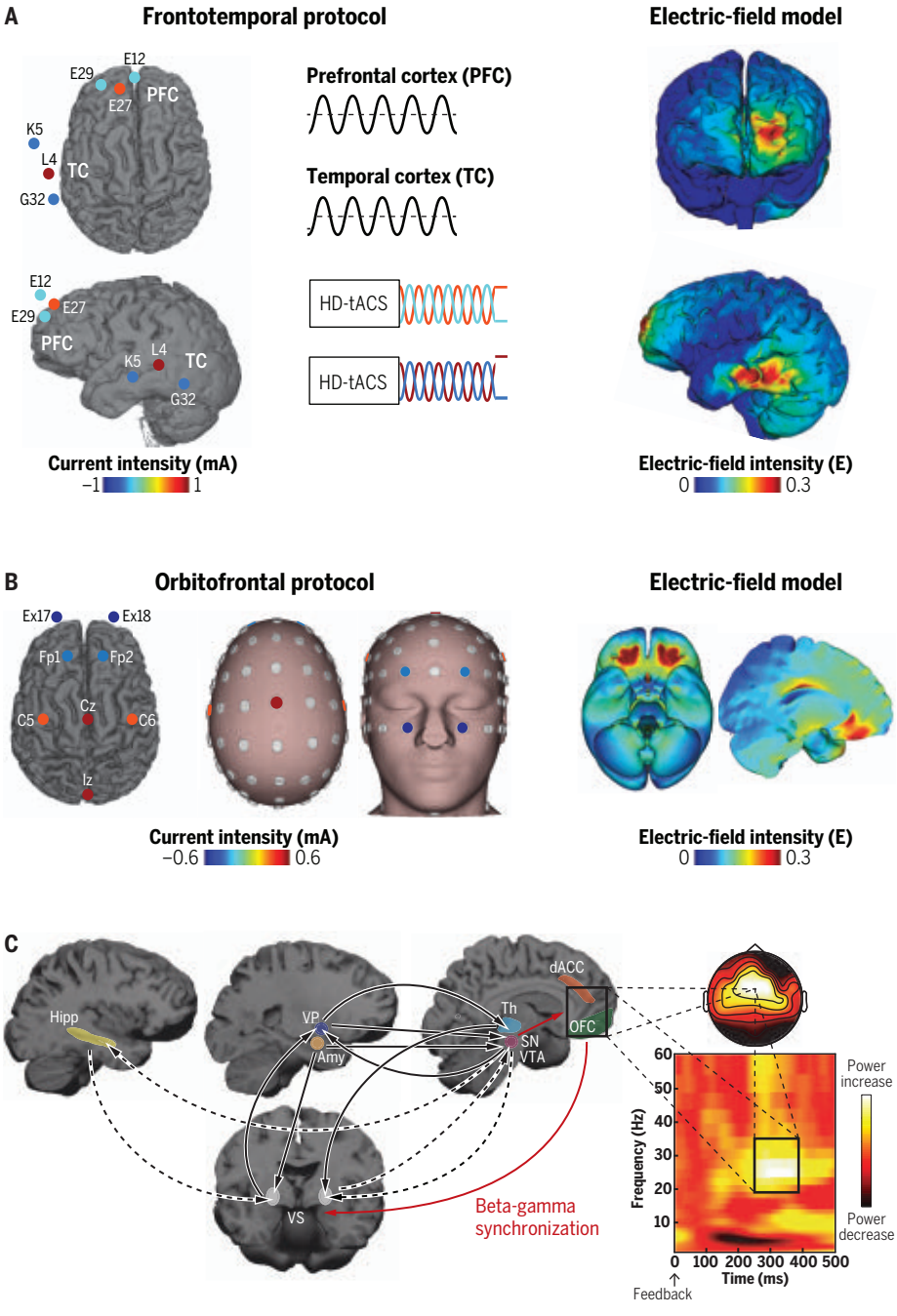
Our neuromodulation design is grounded in fundamental neuroscience research that views cognition as arising from synchronous electrophysiological rhythms across multiple spatiotemporal scales (*2*, *4*–*6*). Neural rhythms derive from inhibitory and excitatory postsynaptic currents and are evident as cyclic changes in the voltages of local field potentials, electrocorticograms, electroencephalograms, and magnetoencephalograms. Functionally, synchronized rhythms sculpt neurophysiological dynamics to facilitate precise yet flexible communication necessary for goal-directed action and cognition (*1*, *6*). This synchronization is achieved by gating information transmission during windows of high excitability through transient, phase-coordinated local neuronal spiking within and across neuronal networks. By coordinating spike timing, synchronization increases the likelihood of inducing spike timing–dependent plasticity, thereby promoting flexible cognitive function. As a result, neural synchronization is recognized as a fundamental neurophysiological process that is necessary for human cognition.

The synchrony of neural rhythms within or between cortical regions can be changed noninvasively using tACS. This technique delivers low-intensity alternating currents on the scalp using a predetermined topographical arrangement of electrodes to induce alternating electric field gradients in a target brain region. These alternating electric fields manipulate spike timing through phase alignment of neuronal activity, without necessarily affecting the spiking rate of the neuronal population (*7*). The entrainment of neuronal activity can lead to the induction of neuroplasticity, which allows any neural and cognitive changes to persist beyond the duration of neuromodulation. Moreover, simultaneous application of two (or more) tACS currents targeted to different brain regions can be used to modulate phase synchronization among them (*3*). Thus, tACS can be used to noninvasively manipulate synchronized rhythmic activity within and between cortical networks.

Advances in tACS technology and protocol design offer expanded functional and anatomical targeting over conventional tACS. Multichannel HD-tACS designs, which involve substantially smaller electrodes arranged in a center-surround configuration, have been shown to produce more focal stimulation in cortical targets (*8*). The spatial arrangement of these electrodes on the scalp is guided by computational models of electric fields, which further improves the focality of stimulation (*8*). Enhanced focality minimizes the likelihood of producing unwanted effects in other anatomical structures such as the retina or the extracranial tissue. Recently, considerable attention has been devoted to the personalization of HD-tACS (*9*). Here, key properties of individual neurophysiology such as anatomical variability or intrinsic rhythmic frequency of a neural circuit are considered. Using a combination of these developments, we have demonstrated effective, functionally specific, bidirectional, and long-lasting control over network synchronization patterns that support cognitive function (*2*–*5*, *10*, *11*). The integration of these advances into the development of personalized, highly fo-

[1]Cognitive and Clinical Neuroscience Laboratory, Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA. [2]Department of Biomedical Engineering, Boston University, Boston, MA, USA. Email: rmgr@bu.edu

PHOTOS: (TOP TO BOTTOM) CYDNEY M. SCOTT; VERONICA SZAREJKO



Multichannel high-definition frontotemporal (A) and orbitofrontal (B) neuromodulation protocols and electric-field models for high-definition transcranial alternating current stimulation (HD-tACS) are shown. The colored circles indicate electrodes. Network integrating reward and learning circuitry showing the reward feedback beta-gamma effect that is hypothesized to rise from the orbitofrontal cortex (OFC) and synchronize with the ventral striatum (VS) via beta-gamma rhythms is presented in (C). Amy, amygdala; dACC, dorsal anterior cingulate cortex; Hipp, hippocampus; SN, substantia nigra; Th, thalamus; VP, ventral pallidum; VTA, ventral tegmental area.

cal designs offers a promising opportunity to better steer the plasticity mechanisms of human cognition.

We recently discovered that synchronization-dependent neural coding schemes underlie poorer memory function in people aged 60 to 76 years and developed advanced neuromodulation protocols that target these motifs for memory enhancement (see the online figure, top). Before neuromodulation, these individuals showed poorer working memory performance compared with younger adults (*2*). These impairments were found to be associated with reduced theta-gamma phase-amplitude coupling (PAC) in the temporal cortex (*2*). PAC is a well-studied neural coding motif that occurs when the amplitude of a high-frequency rhythm synchronizes with the phase of a low-frequency rhythm. This form of synchronization facilitates the integration of information across spatiotemporal scales within a nested cortical network (*6*, *12*). We found that local PAC deficits in the temporal cortex arose because of deficient prefrontal control marked by reduced theta-phase synchronization between the frontotemporal areas. Phase synchronization—when two or more rhythmic neuronal signals tend to cycle with consistent relative phase—is another leading neural coding motif for coordinating spatiotemporal neuronal activity (*1*, *6*, *12*). These synchronization schemes thus serve as potential targets for neuromodulation to improve memory function.

Guided by electric field modeling, we developed a personalized HD-tACS protocol to rescue theta-phase synchronization in the frontotemporal cortex. The frequency of synchronization was individually determined for each participant to maximize the likelihood of entrainment. Simultaneous in-phase entrainment of both frontal and temporal regions at personalized theta frequencies induced in this manner restored intrinsic frontotemporal theta-phase synchronization, recovered the deficient theta-gamma PAC in the temporal cortex (see the online figure, top), and improved working memory performance in older adults (*2*). Even though neuromodulation was performed for ~25 min, improvements in memory function were sustained for at least 50 min, suggesting that the protocol produced neuroplastic changes outlasting the modulation period (*2*). Moreover, an additional experiment in younger adults with antiphase synchronization of frontotemporal regions demonstrated that memory performance can even be down-regulated. This finding suggests that cognitive function can be bidirectionally manipulated using phase-dependent interregional synchronization. This property of precision neuromodulation may be useful in pathologies where overactive memory processes need to be regulated, such as in posttraumatic stress disorder.

Our precision neuromodulation approach identified that it was essential to perform HD-tACS using personalized theta frequencies. By contrast, control experiments with a fixed theta frequency for all participants did not produce any improvements in memory function in older adults. Thus, advances in noninvasive neuromodulation that leverage the spatial and spectral parameters of individual neurophysiology offer a promising opportunity to effectively synchronize large-scale brain rhythms and rapidly improve memory function in older people. Such developments are especially valuable considering the rapidly aging global popu-

GRAPHIC: C. SMITH/*SCIENCE* BASED ON REINHART *ET AL.* (*2*) AND GROVER *ET AL.* (*4*)

lation and its associated personal, social, health care, and economic costs.

Current theories in biological psychiatry on the nature of compulsivity, including obsessive-compulsive disorder (OCD), view symptoms as outcomes of dysregulated habits and atypical reward processing due to abnormalities in cortico-basal ganglia networks (*13*, *14*). In parallel, fundamental neuroscience research has identified a neural signature in the form of medial-frontal beta-gamma rhythms, presumed to arise from the orbitofrontal cortex (OFC) during reward processing (see the online figure, bottom) (*15*). Combining these insights, we proposed that beta-gamma rhythms may constitute the neural code underlying orbitofrontal-striatal interactions that give rise to abnormal reward processing and OCD symptoms. To test this theory, we devised a personalized model-guided HD-tACS protocol for targeting individual beta-gamma rhythms of the OFC (see the online figure, middle) and demonstrated rapid, reversible, frequency-specific modulation of reward-guided choice behavior and learning in healthy young adults (*4*). Next, by repeatedly modulating personalized OFC beta-gamma rhythms over 5 days, we effectively reduced obsessive-compulsive behaviors in a nonclinical population. The rapid reduction in obsessive-compulsive behaviors—including hoarding, ordering, and checking—lasted for at least 3 months (*4*), and the largest improvements were experienced in people with more severe symptoms. These findings bode well for extending this personalized neuroscience intervention to people with clinical OCD and other compulsivity disorders, such as behavioral addiction (e.g., gambling, internet), eating disorders, substance use or abuse, and Tourette syndrome. More broadly, because the OFC is increasingly recognized to play a central role in the pathophysiology of mood, anxiety, psychosis, and other major categories of psychiatric disorders (*14*), the noninvasive procedure we developed for selectively modulating OFC beta-gamma rhythms could lay the basis for future nonpharmacological therapeutics that are applicable to a wide range of psychiatric illnesses.

The fields of fundamental and clinical neuroscience have made extraordinary advances in understanding the dynamic structure of the neuronal network activity that underlies cognitive function and dysfunction. Leveraging these insights has allowed us to develop neuromodulation protocols, personalized to individual neurophysiology, that can selectively augment components of rhythmic cortical networks and improve cognitive function and adaptive behavior in a rapid and sustainable fashion. Although it is challenging to predict the future, we are optimistic that personalization rooted in the neuroscience of network dynamics will rise to the forefront of next-generation noninvasive neuromodulation and pave the way toward future use of precision electroceuticals in neurology and psychiatry. ■

REFERENCES AND NOTES

1. P. Fries, *Neuron* **88**, 220 (2015).
2. R. M. G. Reinhart, J. A. Nguyen, *Nat. Neurosci.* **22**, 820 (2019).
3. S. Grover, J. A. Nguyen, R. M. G. Reinhart, *Annu. Rev. Med.* **72**, 29 (2021).
4. S. Grover, J. A. Nguyen, V. Viswanathan, R. M. G. Reinhart, *Nat. Med.* **27**, 232 (2021).
5. R. M. G. Reinhart, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11542 (2017).
6. M. Siegel, T. H. Donner, A. K. Engel, *Nat. Rev. Neurosci.* **13**, 121 (2012).
7. L. Johnson *et al.*, *Sci. Adv.* **6**, eaaz2747 (2020).
8. R. M. G. Reinhart, J. D. Cosman, K. Fukuda, G. F. Woodman, *Atten. Percept. Psychophys.* **79**, 3 (2017).
9. F. H. Kasten, K. Duecker, M. C. Maack, A. Meiser, C. S. Herrmann, *Nat. Commun.* **10**, 5427 (2019).
10. J. Nguyen, Y. Deng, R. M. G. Reinhart, *Brain Stimul.* **11**, 723 (2018).
11. Y. Deng, R. M. Reinhart, I. Choi, B. G. Shinn-Cunningham, *eLife* **8**, e51184 (2019).
12. G. Buzsáki, C. A. Anastassiou, C. Koch, *Nat. Rev. Neurosci.* **13**, 407 (2012).
13. T. W. Robbins, M. M. Vaghi, P. Banca, *Neuron* **102**, 27 (2019).
14. L. A. Gunaydin, A. C. Kreitzer, *Annu. Rev. Physiol.* **78**, 327 (2016).
15. J. Marco-Pallarés, T. F. Münte, A. Rodríguez-Fornells, *Neurosci. Biobehav. Rev.* **49**, 1 (2015).

ACKNOWLEDGMENTS

The research described in this essay was funded by grants from the National Institutes of Health (R01-MH-114877 and R01-AG-063775) and a generous gift from an individual philanthropist.

10.1126/science.add4834

PRIZE ESSAY

FINALIST

Davide Folloni

Davide Folloni received his undergraduate degree from the University of Parma, a Master's degree in clinical psychology from the University of Padua, and a Master's degree and PhD in neuroscience from the University of Oxford. His research focuses on the development of multimodal approaches that combine neuromodulation, electrophysiological, and imaging techniques to understand the neural mechanisms underlying learning and decision-making, with the primary aim being to describe how neuronal oscillations guide our adaptation to the world.

www.science.org/

NEUROMODULATION

Ultrasound neuromodulation of the deep brain

Noninvasive, reversible stimulation of neural circuits can regulate behavior

By **Davide Folloni**

Understanding the relationship between the brain and behavior is one of the main goals of neuroscience and is crucial for the development of rehabilitative technologies. When it comes to humans, however, there is still no tool able to modulate activity in deep areas of the brain with millimetric resolution and in a noninvasive and reversible fashion. Beyond neurorehabilitation, the understanding of the functional role of a brain area can be established only through causal inference—that is, by manipulating its activity and observing associated behavioral changes.

Current reversible neuromodulation methods have limited spatial resolution and do not reach regions deep in the brain (*1*). For this reason, my colleagues and I decided to implement an alternative approach, called focused transcranial ultrasound stimulation (TUS) that relies on the strong interactions between low-intensity sound waves and brain tissue to transiently modulate neural activity noninvasively, reversibly, and with high spatial precision. Our work shows that TUS, compared with magnetic or electric fields, can directly focus sound waves to small regions deep in the brain without affecting the overlying cortex and with results previously only achievable using surgically implanted electrodes and other invasive procedures (*2*).

Activity in each brain area is modulated by the network of regions with which that area is interconnected. Such connections determine which regions influence an area and, in turn, which regions are influenced by that area. In an initial set of experiments (*2*), we developed a multimodal approach combining TUS with whole-brain resting-state functional magnetic resonance imaging (fMRI) in macaque monkeys to selectively manipulate neural activity in a subcortical area (the amygdala) and in a deep cortical region [the anterior cingulate cortex (ACC)] (see the figure, upper left). Whole-brain fMRI enabled the recording of poststimulation activity in both the stimulated area and interconnected neural circuits.

By using a 40-s offline TUS protocol, with a wave frequency set to the 250-kHz resonance frequency and 30-ms bursts of ultrasound generated every 100 ms, we demonstrated for the first time in primates (which are characterized by a thick skull bone, similar to that in humans) that TUS can be used transcranially to modulate activity in both the amygdala and ACC for up to 2 hours with high spatial specificity and without affecting the overlying cortex (*2*). TUS targeted to the amygdala resulted in focal changes in the activity coupling of this region with a set of areas compared with a no-stimulation (i.e., sham) condition. However, this modulation was not present if we targeted TUS to an active control area, such as the ACC. The same focal effects—now centered on the neuromodulation of the ACC—were observed when we sonicated this latter region (*2*). We replicated these effects successfully in other cortical regions (*3*). The duration of these effects and their reversibility may present therapeutic opportunities in patients with dementia, motor-related pathologies, and psychiatric and neurological conditions.

Transcranial ultrasound modulation of deep brain regions could be groundbreaking in health care, but only if it is able to translate into a behavioral change. We therefore continued our research by applying TUS to awake-behaving monkeys trained to perform a decision-making task in which they had to learn the value of choice options (*4*). The animals' choice strategies were typically supported by activity in the same ACC region that had been targeted previously (*2*) and the hippocampus. The ACC specifically tracked the value of all potential counterfactual choices during a trial, and its activity predicted the translation of this value into a change in future better choices. TUS targeted offline to the ACC led to a substantial alteration in this behavior (*4*). In subsequent experiments (*5*, *6*), we reproduced the selective neuromodulatory effects of

Icahn School of Medicine at Mount Sinai, New York, NY, USA.
Email: davide.folloni@mssm.edu

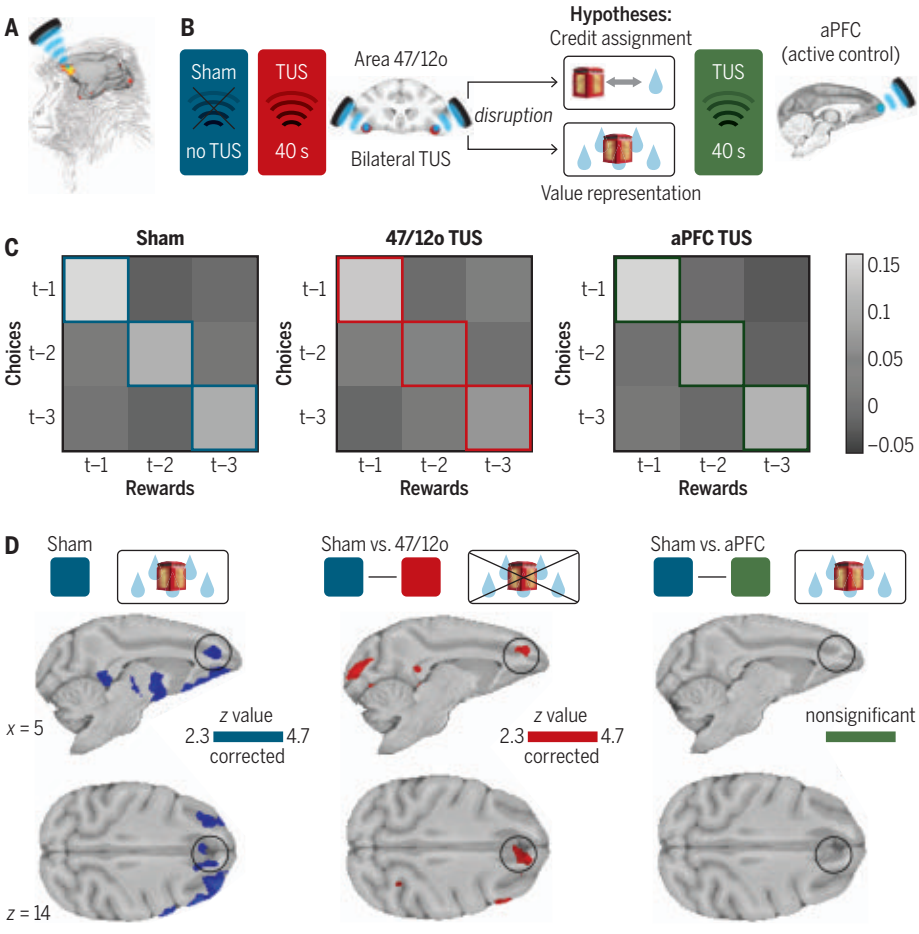
PHOTO: VERONICA SZAREJKO

TUS on behavior, using it to modulate activity in the basal forebrain and medial prefrontal cortex.

A brain area's role is partly a function of the interactions that it has with other brain regions. Because our previous results show that each area's interactions can be altered in a substantial way (*2*–*6*), it suggests that by temporarily modulating activity within a deep brain area using TUS, this area's ability to contribute to a cognitive process can be disrupted. Although we had learned about the neural (*2*) and behavioral (*4*) effects of TUS separately, we still did not know whether deep brain TUS caused simultaneous changes in both activity—at the target as well as within its interconnected network—and behavior.

To answer this, we applied TUS to macaque monkeys performing a decision-making task while simultaneously recording their neural activity using an MRI scanner (*7*). The animals were required to flexibly learn contingent associations between choice options and outcomes and use these associations to guide subsequent behavior—a process called credit assignment [suggested to be encoded in the ventral prefrontal cortex (vPFC), more specifically in area 47/12o] (*8*, *9*, *10*). To causally test the role of 47/12o in credit assignment, TUS was applied to the animals' 47/12o areas immediately before they entered the MRI scanner to perform the task. The resultant changes in behavior and associated neural activity were compared with a sham condition and with a third active-control condition, where TUS was instead applied to the adjacent anterior prefrontal cortex (aPFC) (see the figure, upper right). By exploiting the transient and noninvasive properties of our TUS protocol, we were able to interleave sham and TUS sessions within each animal and counterbalance their order across all animals, thereby controlling for any potential order effect.

Primates, including humans and macaques, learn how to make choices by identifying over time the pattern of outcomes that follow a specific action. In our daily life, it is essential to know whether the beneficial consequences that follow our choices are simply a feature of the current environment we find ourselves in or whether these outcomes are genuinely caused by the choices we made. Subject animals were generally good at learning from previous outcomes, and this behavior was associated with activity in the vPFC, including area 47/12o. Focal TUS modulation of 47/12o activity, however, disrupted such learning and the animals' ability to integrate the observed positive outcomes with the concomitant choices recently taken (*7*).



(A) Focused transcranial ultrasound stimulation (TUS) applied to the primate anterior cingulate cortex (ACC) in a macaque. (B) Experimental TUS conditions: sham, where no stimulation was applied (blue); TUS targeting brain area 47/12o (red); and an active control condition (green) in which TUS was applied to the anterior prefrontal cortex (aPFC). (C) These plots indicate the influence of past reward history (*x* axis) and choice history (*y* axis) on which choice is taken next during the three conditions. The brighter (i.e., lighter gray) the main diagonal, the stronger the influence of the conjoint history of choice and reward on future behavior. Labels t–1, t–2, and t–3 refer, respectively, to the last occasion, the previous occasion, and the occasion before that on which a given choice option was encountered. In sham, choices are influenced by the conjoint history of choice and reward. (D) In the sham condition, activity in several frontal cortical areas, including the ACC, reflected values of choices that could be taken. This value signal in the ACC was significantly and selectively reduced after 47/12o TUS but not after aPFC TUS.

We next used a complementary approach to examine how the extended history of choice-reward contingencies experienced would influence which choice a macaque would make in the future (*8*, *10*). We combined TUS and fMRI with computational models to show that two separate subfields of cortex encode these two features of rewards: Area 47/12o activity reflects the process of credit assignment, whereas the adjacent anterior insula tracks a general signal reflecting the value of the environment. TUS targeted to 47/12o (see the figure, middle, red) selectively changed the animals' ability to learn choice-reward contingencies compared with sham (see the figure, middle, blue). This impairment was absent if TUS was applied to the nearby control aPFC (see the figure, middle, green) (*7*).

Credit assignment encoded in area 47/12o is pivotal in our everyday behavior because it determines the value assigned to every choice we make. Ultrasound-based disruption of this area not only affected credit assignment but also resulted in activity changes in an extended cortical circuit, including a subregion of the ACC where we had previously shown that choice values are held and compared during decision-making (see the figure, middle, red) (*4*). Notably, despite the result that TUS applied to 47/12o modulated functionally dependent signals in these distant components of monosynaptically interconnected neural circuits (see the figure, bottom), such alteration of 47/12o left functionally independent activity (e.g., signals in the immediately adjacent anterior insula that kept tracking the general value of the environment regardless of the choice taken) intact (*7*). Although these

GRAPHIC: K. FRANKLIN/*SCIENCE* BASED ON K. FOLLONI *ET AL.* (*7*)

studies were done in nonhuman primates, they suggest the possibility of safely translating TUS to humans, especially in those brain circuits that are evolutionarily preserved across species (*11*).

Over the past few years, we have demonstrated that the noninvasive properties of TUS and its high spatial resolution have the potential to substantially enhance our causal understanding of how the human brain works as well as our development of therapeutic protocols targeting abnormal activity in deep brain regions. TUS's ability to provide manipulation techniques previously only available in rodent models is likely to critically affect future human brain investigations.

The possibility of using a short and offline protocol makes TUS a promising cutting-edge treatment for psychiatric and neurological patients, especially when their symptomatology prevents the application of longer or invasive treatments. Finally, the possibility of reaching reversible modulation of neural activity with millimetric resolution even in subcortical brain areas may finally provide surgeons with a presurgical tool to identify, in each of their patients, which portions of gray or white matter are more likely to lead to an improvement of symptoms after implantation of deep brain stimulation electrodes. ■

REFERENCES AND NOTES

1. R. Polanía, M. A. Nitsche, C. C. Ruff, *Nat. Neurosci.* **21**, 174 (2018).
2. D. Folloni *et al.*, *Neuron* **101**, 1109 (2019).
3. L. Verhagen *et al.*, *eLife* **8**, e40541 (2019).
4. E. F. Fouragnan *et al.*, *Nat. Neurosci.* **22**, 797 (2019).
5. N. Khalighinejad *et al.*, *Neuron* **105**, 370 (2020).
6. A. Bongioanni *et al.*, *Nature* **591**, 270 (2021).
7. D. Folloni *et al.*, *Sci. Adv.* **7**, eabg7700 (2021).
8. M. E. Walton, T. E. J. Behrens, M. J. Buckley, P. H. Rudebeck, M. F. S. Rushworth, *Neuron* **65**, 927 (2010).
9. B. K. H. Chau *et al.*, *Neuron* **87**, 1106 (2015).
10. P. H. Rudebeck, R. C. Saunders, D. A. Lundgren, E. A. Murray, *Neuron* **95**, 1208 (2017).
11. D. Folloni *et al.*, *eLife* **8**, e47175 (2019).

ACKNOWLEDGMENTS

The work detailed here was conducted at the Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, Oxford, UK.

10.1126/science.add4836

SPECIAL SECTION

PERSPECTIVE

The history and challenge of grassy biomes p. 592

REVIEWS

Ancient grasslands guide ambitious goals in grassland restoration p. 594

Molecular, cellular, and developmental foundations of grass diversity p. 599

Grassland soil carbon sequestration: Current understanding, challenges, and solutions p. 603

The planetary role of seagrass conservation p. 609

RELATED ITEMS

NEWS STORY p. 568 PODCAST

THE UNRECOGNIZED VALUE OF GRASS

By **Bianca Lopez, Pamela J. Hines,** *and* **Caroline Ash**

Grasses are highly diverse, yet only six or seven grass species provide most of the calories that humans consume. Domestication of grasses as crops began some 10,000 years ago and continues today to optimize the genetic basis of traits useful for crop cultivation. Techniques to maximize yields and utility of staple grain crops still dominate modern agriculture. In addition to cultivated fields and pastures, grassy ecosystems (both Poaceae and Alismatales) cover large swaths of the planet, forming terrestrial grasslands and submarine meadows. Grasslands create and stabilize fertile soil; store carbon; generate oxygen; and provide animal habitat, building materials, and food. Even so, these species and systems are often undervalued. Land-use conversion and climate change pose threats, as do climate change mitigation efforts that prioritize carbon stored in trees over that stored in grasslands. Nevertheless, grasses could offer solutions to many of our societal challenges, if only we would fully recognize their diversity and value.

Meadows of Neptune seagrass (*Posidonia oceanica*) were once widespread throughout the Mediterranean but are threatened by climate change and human activities.

PHOTO: SHANE GROSS/MINDEN PICTURES

PERSPECTIVE

The history and challenge of grassy biomes

Grassy biomes are >20 million years old but are undervalued and under threat today

By **Caroline A. E. Strömberg**[1] and **A. Carla Staver**[2,3]

Grassy biomes—from the steppes of Mongolia to the savannas of Tanzania—are predicted to be the ecosystems hardest hit by the ongoing climate and land use crises. The history of humans has been profoundly intertwined with grassy biomes. *Homo* evolved in the savannas 2 million years ago (Ma), and agricultural societies arose through the domestication of grasses, such as wheat and barley, 10,000 years ago. These grass crops, as well as corn and rice, remain dominant staple foods globally (*1*). Livestock production also centers in areas that were once (and sometimes still are) native grasslands. Grassy biomes harbor distinct and diverse sets of plants and animals that have adapted to these environments through millions of years of evolution (*2*). As the biodiversity and economic prominence of grassy biomes are increasingly being recognized, there is a demand for better understanding of their past and present function to inform policy and management.

Grassy biomes are biogeographically widespread, accounting for >25% of all land on Earth, including 35% of the tropics and subtropics. The emergence of grassy systems during the Cenozoic (the past 66 million years) was complex, shaped by climate, soils, fire, and herbivory in ways that are not fully understood (see the figure). Clarifying these mechanisms will be key for managing the fate of grassy biomes under ongoing and future environmental changes that are driven by human activities.

Grasses, defined as plant species in the family Poaceae, originated by the Late Cretaceous (100 Ma) (*3*) but did not become ecologically dominant until >70 million years later, in the later Cenozoic. This exceptionally long lag has prompted evolutionary biologists and paleontologists to search for the drivers that allowed grass to reach its current global prominence. Today, most grasses are associated with open-canopy habitats, owing to several traits acquired relatively early in Poaceae evolution (100 to 60 Ma) (*1*, *3*). For example, grasses may have quickly evolved a rapid life cycle and persistent buds, permitting quick regrowth after drought, frost, or disturbances such as fire and grazing. Starting by 55 Ma, several groups of grasses evolved so-called C_4 photosynthesis (as opposed to C_3 photosynthesis), which allows them to prosper in hot and dry areas (*1*). In colder climates, C_3 open-habitat grasses developed the tolerance needed to survive frosts by 30 Ma (*4*). However, although the evolutionary traits suited to open habitats appeared earlier, open-habitat grasses remained ecologically rare until later in the Cenozoic.

Once grasses started spreading across the globe, their takeover was asynchronous and followed continent-specific trajectories. For instance, grassy habitats appeared in North America by 25 Ma but not until 7 Ma in Australia (*5*, *6*). However, the first subtropical grassy biomes were unlike anything observable there today, featuring C_3 open-habitat grasses that today are found in colder regions (*6*). It was not until several million years later that tropical open-habitat C_4 grasses expanded to form grasslands and savannas at low to mid-latitudes (*5*, *7*), roughly coincident with the spread of frost-tolerant grasses at higher latitudes.

Grassy biomes thus emerged during the Cenozoic at different times in different places and, at least in part, for different reasons. Studies in modern grassy biomes suggest that aridity and rainfall seasonality, as well as fire and herbivory, could all favor grasses over trees (*2*), with even larger benefits at lower atmospheric CO_2 concentrations. The fossil record shows that many of these conditions did occur in the late Cenozoic. By 34 Ma, atmospheric CO_2 levels had dropped, and the globe underwent a period of cooling. In many areas, altered atmospheric circulation and mountain uplift (e.g., of the Tibetan Plateau) resulted in aridification or seasonal drought, and fossil evidence indicates increased fire activity near the end of the Cenozoic (*5*). Further, large grassland-type mammal herbivores (e.g., bovids) diversified during the mid- to late Cenozoic (*8*).

Asynchrony in the emergence of grasses on different continents suggests that, although global factors such as low-CO_2 conditions may have spurred the diversification and expansion of open-habitat and especially C_4 grasses (*7*), changes in CO_2 were typically not enough to allow grasses to dominate. A rapidly expanding geochemical and paleontological tool kit has allowed for more detailed insights. Studies have shown that regional changes in climate and fire interacted with existing vegetation to influence trajectories of emerging grass dominance, with divergence across continents. For example, the earliest North American C_3 grassy habitats replaced forests as seasonal drought developed (*6*), and in Australia, C_4 grasses favored by pronounced aridification overtook fire-adapted eucalypt woodlands that had existed there for tens of millions of years before (*5*). By contrast, in South Asia and southwest Africa, more frequent and intense wildfires promoted replacement of fire-sensitive vegetation with grasses (*9*), suggesting a substantial regional, if not global, role for fire.

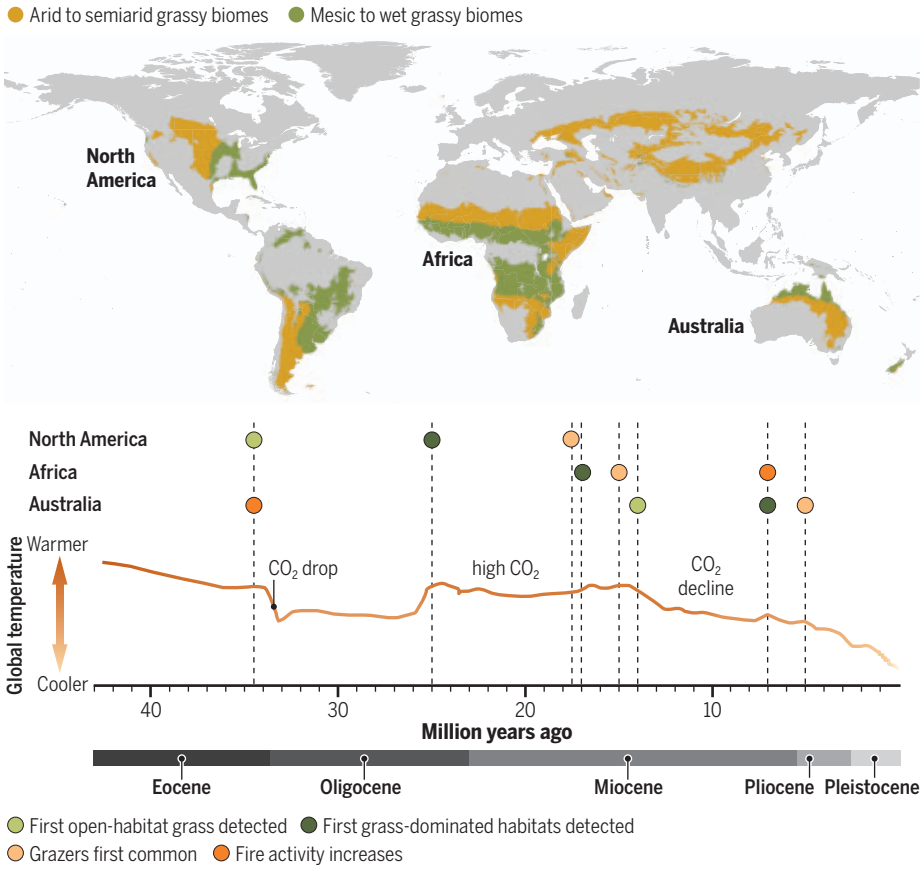
In addition to environmental conditions, herbivores may also have directly contributed to the spread of grassy vegetation, although the mechanisms are not yet understood. Defense strategies against herbivores by savanna trees, such as growing spines or thorns, evolved concurrently with the spread of grasses and the diversification of bovids in Africa (~17 Ma) but long before fire activity increased (*8*). This suggests that, at least in Africa, herbivores structured grassy biomes before fire did. However, just how important animals were in shaping the evolution of grassy vegetation remains untested and will require adapting methods of estimating past herbivore intensity (such as studying fungal spores in fossilized dung) for Miocene and older samples.

Since they first appeared, grassy biomes have continued to shift in extent, structure, and composition, prompted by advancing and retreating ice sheets during the global Ice Age (2.6 Ma onward). Today, they are widely distributed on every continent except Antarctica, with a range in part associated with aridity and rainfall seasonality. Some 60% of grassy ecosystems receive <750 mm of annual rainfall, most with a dry season that shapes plant physiology. This provides a rationale for the argument that aridity drove late Cenozoic grassland expansion. However, 40% of grassy ecosystems extend into higher-rainfall regions with >750 mm of annual rainfall that can support forests. These moderately wet, or "mesic" grassy ecosystems are biogeographically distinct from semiarid ones, but both are evolutionarily ancient (*1*). Yet, whereas

[1]Department of Biology, Burke Museum of Natural History and Culture, University of Washington, Seattle, WA, USA. [2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. [3]Yale Institute for Biospheric Studies, Yale University, New Haven, CT, USA. Email: caestrom@uw.edu; carla.staver@yale.edu

The history and legacy of grassy biomes

Grassy biomes exist in a wide range of climates, from cold to hot and arid to wet (top). Although changing environmental conditions through time have shaped their past and present distribution, disturbance regimes (fire, herbivory) and vegetation histories also shaped their evolution and current and future function (bottom).



GRAPHIC: K. FRANKLIN/*SCIENCE*; (DATA) KARP *ET AL.* (5); KUKLA *ET AL.* (6); CHARLES-DOMINIQUE *ET AL.* (8); HOETZEL *ET AL.* (9)

semiarid savannas are widely accepted as the native vegetation of large areas of the globe, mesic savannas were long assumed to represent degraded forests. Only recently have mesic savannas been acknowledged for their contributions to endemic biodiversity and distinctive ecosystem function.

As their antiquity is increasingly recognized, the ecological processes that promote mesic savanna stability have come into increasing focus (*2*). Fire likely plays an important role in stabilizing mesic savannas, excluding forests by preventing tree establishment or killing trees, thereby favoring grasses. In total, grassy biomes make up >80% of the global burned area annually. Experiments, field observations, and remote sensing analyses all support fire as a mechanism allowing grassy ecosystems to expand into mesic regions. Plant traits are consistent with the history of fire in mesic savannas. The distinct, diverse, and ancient tree and shrub communities (*8*) are well adapted to enduring fires with thick bark, large belowground nonstructural carbohydrate reserves, and bud banks that promote resprouting. In addition to tolerating fire, many grasses actively spread fire (*10*). These fire adaptations have major implications for the ecosystem functioning of grassy biomes. For instance, the large belowground reserves in grassy biomes may mean a substantially larger belowground carbon storage compared with that in other biomes (*11*). Current estimates suggest that grassy biomes hold at least 17% of global biomass carbon (*12*), but this is certainly an underestimate (*11*) that needs to be adequately quantified so that the potential role of grassy biomes as carbon sinks can be fully appreciated.

Herbivores that graze on grass and eat tree leaves also influence grassland function (*13*), especially in semiarid savannas, where grass eaters decrease grass biomass accumulation and tree eaters prevent trees from establishing. Abundant herbivory-related traits have accumulated over evolutionary time in grassland plants, including herbivory defenses in trees (e.g., spines) (*8*) and grass morphologies that withstand intense grazing (e.g., growing from the base instead of from shoot tips and bud banks for resprouting) (*1*). Nevertheless, the importance for grassy biome distributions of herbivory relative to other factors, such as climate and soil conditions, remains an open question.

Overall, evidence is converging around the idea that grassy ecosystems are complex, with ecologies that depend not just on climate but also on interactions and feedbacks with fire and herbivory. These ecologies are profoundly influenced by the evolutionary history and resulting trait diversity of regional biota (*1*). Their complexity makes predicting the responses of grassy biomes to global change a particular challenge. Nonetheless, studies have shown that the combination of CO_2 fertilization, fire suppression, and livestock extensification has resulted in widespread woody encroachment (*14*) and associated degradation of grassy biomes—a trend that will likely continue into the near future.

Grassy biomes are also threatened by ongoing land use conversions and degradation while being among the least protected globally (*2*). For example, 90% of temperate grasslands have been transformed into agricultural or urban areas, with <1% of remnants currently protected from land development. Whereas rainforests in the Amazon have attracted widespread attention from the popular media, the ongoing threat to savannas, especially in Africa, South America, and Asia from afforestation, fire exclusion, and land use conversion, has gone unnoticed. The effects on savanna and grassland biodiversity will be devastating; for instance, 40% of grassland vertebrate species are projected to be lost by 2070 (*15*). Thus, the fate of evolutionarily ancient grassy biomes hangs in the balance, with terminal consequences for their functionally and evolutionarily distinct biota. ■

REFERENCES AND NOTES

1. H. P. Linder *et al.*, *Biol. Rev.* **93**, 1125 (2018).
2. W. J. Bond, C. L. Parr, *Biol. Conserv.* **143**, 2395 (2010).
3. T. J. Gallaher *et al.*, *Evolution* **73**, 927 (2019).
4. M. Schubert *et al.*, *Glob. Ecol. Biogeogr.* **28**, 1168 (2019).
5. A. T. Karp *et al.*, *Geophys. Res. Lett.* **48**, e2020GL090964 (2021).
6. T. Kukla *et al.*, *AGU Advances* **3**, e2021AV000566 (2022).
7. J. Lu *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **538**, 109454 (2020).
8. T. Charles-Dominique *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5572 (2016).
9. S. Hoetzel *et al.*, *Nat. Geosci.* **6**, 1027 (2013).
10. K. J. Simpson *et al.*, *J. Ecol.* **104**, 138 (2016).
11. Y. Zhou *et al.*, *Nature* **603**, 445 (2022).
12. S. A. Spawn *et al.*, *Sci. Data* **7**, 112 (2020).
13. A. C. Staver *et al.*, *J. Ecol.* **109**, 2804 (2021).
14. N. Stevens, C. E. R. Lehmann, B. P. Murphy, G. Durigan, *Glob. Change Biol.* **23**, 235 (2017).
15. T. Newbold, *Proc. R. Soc. B.* **285**, 20180792 (2018).

ACKNOWLEDGMENTS

This work resulted from a visit by C.A.E.S. to Yale, supported by an Edward P. Bass Visiting Fellowship from the Yale Institute for Biospheric Studies. Savanna contributions to the global biosphere were calculated with the RESOLVE biomes map and the Global Fire Emissions Database (v4). Participation by A.C.S. was supported by a grant from the National Science Foundation (NSF-MSB 1802453).

10.1126/science.add1347

REVIEW

Ancient grasslands guide ambitious goals in grassland restoration

Elise Buisson[1]†, Sally Archibald[2], Alessandra Fidelis[3], Katharine N. Suding[4,5]*†

Grasslands, which constitute almost 40% of the terrestrial biosphere, provide habitat for a great diversity of animals and plants and contribute to the livelihoods of more than 1 billion people worldwide. Whereas the destruction and degradation of grasslands can occur rapidly, recent work indicates that complete recovery of biodiversity and essential functions occurs slowly or not at all. Grassland restoration—interventions to speed or guide this recovery—has received less attention than restoration of forested ecosystems, often due to the prevailing assumption that grasslands are recently formed habitats that can reassemble quickly. Viewing grassland restoration as long-term assembly toward old-growth endpoints, with appreciation of feedbacks and threshold shifts, will be crucial for recognizing when and how restoration can guide recovery of this globally important ecosystem.

Grasslands are essential components of Earth's system, supporting a biodiverse array of plants, birds, insects, and other animals and providing important ecosystem services such as pasture forage, water regulation and freshwater supply, erosion control, pollinator health, and carbon sequestration (*1*, *2*). Yet high rates of land cover conversion for intensive agriculture and silviculture, combined with woody encroachment and species invasion driven by altered fire and grazing regimes, threaten these systems (*3*, *4*). For instance, the Cerrado has been extensively cleared for agriculture, with more than half lost in the past 50 years, exceeding the rate of forest loss in the Brazilian Amazon (*5*). The Great Plains of North America has also lost more than half its original grasslands and continues to lose 2% each year (*6*).

As we enter the United Nations Decade on Ecosystem Restoration, much of the emphasis has been on the restoration of forests (*7*). Ironically, this emphasis presents an additional threat to grasslands: Careless or poorly planned tree-planting efforts in the name of restoration can establish forests in natural grassland and savannah ecosystems. For instance, almost 1 million km^2 of Africa's grassy biomes have been targeted for tree planting by 2030 (*8*). This practice ignores the value of protecting and restoring grasslands.

The conversion and degradation of grasslands can occur rapidly, yet restoring lost ecosystem services and diversity is often a discounted or underestimated challenge. Until recently, grassland assembly was assumed to be a relatively straightforward—albeit difficult—process (*9*): Allow herbaceous species to recolonize, at times augmenting with seed of native species; reestablish appropriate grazing and fire disturbance regimes; and control ruderal, exotic, or woody species. Because many herbaceous species reach reproductive maturity in a few years, it was also assumed that this assembly process was relatively quick, achieving desired diversity and function within several years to a decade. We now know that this view of grassland restoration does not adequately acknowledge the difficulty of restoring biodiversity and functions or the time and interventions needed to restore grasslands (*10*). Here, we review recent developments that widen the view of grassland restoration to include grassland age and development, describe how this lens identifies important but overlooked restoration interventions, and highlight several key unknowns for grassland restoration into the future.

[1]Institut Méditerranéen de Biodiversité et d'Ecologie, Avignon Université, CNRS, IRD, Aix Marseille Université, 84911 Avignon, France. [2]Centre for African Ecology, School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg 2050, South Africa. [3]Instituto de Biociências, Lab of Vegetation Ecology, Universidade Estadual Paulista (UNESP), Rio Claro 13506-900, Brazil. [4]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA. [5]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, USA.
***Corresponding author. Email: suding@colorado.edu**
†These authors contributed equally to this work.

Refining the reference: The old-growth concept for grasslands

Grasslands occur in a range of biogeographical contexts (Fig. 1) including the tropical and subtropical savannas in Africa, Australia, Asia, and South America; the boreal, temperate, and southern prairies in North America; and the steppes in Eurasia. Grasslands have a continuous herbaceous layer of graminoids and herbaceous dicots, either without trees or, in the case of savannas, supporting a range of tree densities with a continuous grassy understory (*3*) (Fig. 2). The processes creating and maintaining grasslands vary across locations (*11*); these include edaphic or climatic conditions and disturbances (i.e., herbivore grazing or fire), all of which can limit the establishment of woody species (Fig. 3).

The reference condition is a cornerstone concept in ecological restoration; it encapsulates a set of desired characteristics and provides guidance for how to evaluate project success, even if a restored system is rarely able to completely reach reference conditions (*12*). In grasslands structured by edaphic or climatic conditions, with soils, low temperatures, or low precipitation constraining tree establishment, grassland is generally acknowledged to be the desired reference state for restoration. In cases where climate is suitable for forests but herbivore grazing or fire maintain them in an open state (*10*) (Fig. 3), more debate and uncertainty surrounds the reference designation. These disturbance-dependent grasslands are often assumed to be a result of deforestation (i.e., derived grasslands; grass-dominated vegetation resulting from human-caused deforestation) in an early successional stage on a forest trajectory (Fig. 4). However, climate suitability for tree growth does not preclude the likelihood that old-growth grasslands exist (or used to exist) in the region (*13*).

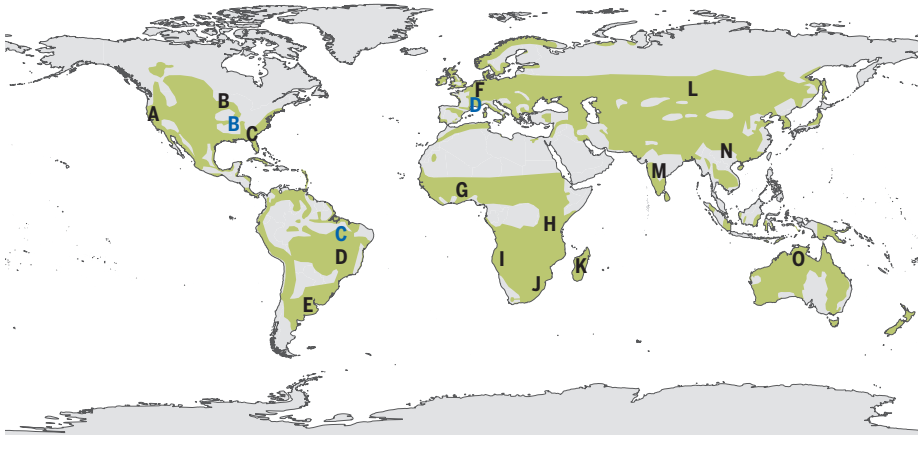


Fig. 1. The distribution of grasslands spans temperate and tropical regions of the globe. Green areas estimate the extent of grassland distribution. We note, however, that all maps of grasslands should be considered imprecise: Grasslands occur mixed within landscapes with other vegetation types and are often disturbed to an extent that masks historic distributions. Letters in black are grasslands represented in Fig. 2; letters in blue are grasslands represented in Fig. 3.

Moreover, these disturbance-dependent grasslands are often at risk from factors driving woody invasion, rearranging landscape mosaics and shifting grass-forest boundaries (*14*). If afforestation policies under the guise of restoration disregard these dynamics, irreversible damage will occur (*7*).

In forest ecosystems, old-growth forests are often used as references for restoration. These are mature forests composed of large and old trees, large snags, and a diverse tree community with structural complexity, all of which require long time periods to develop. Recent work has made it abundantly clear that the "old growth" concept is not limited to forests (*4*, *11*): Old-growth grasslands, also called ancient or pristine grasslands, assemble over centuries and contain high species diversity, long-lived perennial plants, and a substantial proportion of well-developed belowground structure from which species can resprout after natural disturbance. Old-growth grasslands are unique in their underground structures and biodiversity: They store carbon and reallocate resources aboveground after disturbances and drought. All biogeographic contexts where grasslands are present (Fig. 1) support ancient old-growth grasslands that have persisted for millennia.

As with old-growth forests, there should be little expectation that restored grasslands will ever completely recover to resemble old-growth grasslands. Even so, old-growth grasslands provide a suite of characteristics that can be the aim in restoration: long-lived perennial plants; a complex diversity of belowground structures that enable resprouting after aboveground disturbances such as fire and grazing; and substantial belowground carbon stores. Traditional management can usefully target these old-growth characteristics even in cultural landscapes where grasslands are created and maintained by human activity, and regardless of historical analogs (*15*).

With maps of grasslands contested and overlapping those of forests (*8*, *13*), it can be challenging to determine whether a grassland is one that formed after the degradation of an old-growth grassland (i.e., a secondary grassland; grass-dominated vegetation resulting from the degradation of old-growth grasslands) or a derived grassland formed after deforestation. Paleoenvironmental methods, considering

PHOTOS: (A, B, D, K) ELISE BUISSON, (C, G, H, I, J, O) SALLY ARCHIBALD, (E) ALESSANDRA FIDELIS, (F, L) RENAUD JAUNATRE, (M) AATHIRA PERINCHERY, (N) TRISTAN CHARLES-DOMINIQUE

Fig. 2. The incredible diversity of old-growth grasslands. See Fig. 1 for locations. Whether these grasslands are maintained by disturbance (such as grazing or fire) or are environmentally constrained (EC, edaphic or climatic; see Fig. 3 for details) is indicated within brackets. (**A**) California coastal grasslands on Mount Tamalpais, USA (disturbance). (**B**) Curtis Tallgrass Prairie Restoration, Wisconsin, USA (disturbance). (**C**) Longleaf pine (*Pinus palustris*) savanna, North Carolina, USA (disturbance). (**D**) Grassland in the Espinhaço mountain range, Minas Gerais, Brazil (EC, edaphic + disturbance). (**E**) Subtropical grasslands in Rio Grande do Sul, southern Brazil (disturbance). (**F**) Alpine meadow in the Alps, Vanoise National Park, France (EC, climatic). (**G**) A high-rainfall grassy savanna in Mole National Park, Ghana (disturbance). (**H**) The Serengeti ecosystem in Tanzania (EC, edaphic + disturbance). (**I**) The grasslands in the Kavango Catchment, Angola (EC, edaphic and climatic + disturbance). (**J**) Grassland in the Drakensberg, South Africa (disturbance). (**K**) Grassland and tapia savannas on Ibity mountain, Madagascar (disturbance). (**L**) Petrophytic steppe in Khakassky Zapovednik State Nature Reserve, Russia (EC, climatic). (**M**) Eravikulam Shola grasslands, India (EC, climatic + disturbance). (**N**) Oak savanna in South Yunnan, YuanJiang region, China (disturbance). (**O**) Mesic savanna in the Northern Territory, Australia (disturbance). These grasslands vary widely in composition and structure yet share key characteristics that can guide restoration: high belowground allocation, complex resprouting structures, and unique functional and taxonomic diversity.

lengthy records of pollen, phytoliths, charcoal, and *Sporormiella* fungi specific to herbivore guts, can provide evidence for past grasslands and their disturbance history (*16*). Species composition and functional diversity (e.g., of belowground structures), as well as phylogenetic studies dating the origins of endemic grassland species, can also indicate antiquity and conservation value (*17*, *18*). There are also contexts where grasslands are the desired ecosystem state for cultural or social reasons despite being created or maintained by humans.

Pathways and thresholds of grassland degradation

Grasslands are increasingly degraded by land-use change and altered disturbance regimes, which can fundamentally alter their structure and functioning (Fig. 4). Such degradation increases the need for grassland protection and restoration but can also decrease the capacity of restoring old-growth grassland characteristics.

Grazing and fire are dominant aboveground disturbances that have coevolved with grassland plants, maintaining diversity and function in grasslands (*4*). Changes to these disturbance regimes can gradually alter grasslands. Although this results in the loss of biodiversity and simplification in composition, structure, and functioning, altered grassland often maintains some belowground structures (Fig. 4). Lack of grazers (or of particular suites of grazing species) can homogenize grasslands and increase fire occurrence. On the other hand, overgrazing, particularly in grasslands with no evolutionary history of grazing, can result in loss of basal cover, soil compaction, and increased erosion (*19*). Defining the degradation point in these circumstances is difficult; for instance, naturally occurring "grazing lawns" have many of the biophysical characteristics associated with degradation (low aboveground biomass, soil compaction, sometimes even increased bare ground) even though their unique biodiversity and ecological importance is now increasingly recognized. Fire regimes can also become too frequent or infrequent or occur during the wrong season. The longer these altered disturbance regimes persist, the more risk to belowground structure (e.g., bud banks) that speed recovery. Altered disturbance regimes can also facilitate exotic grass invasion and woody encroachment (*20*), which can compound impacts to belowground structure over time.

The most detrimental disturbances are those that rapidly destroy belowground structure, such as tillage agriculture, mining, and afforestation (*10*, *21*). For instance, 50 years of pine plantation completely eliminated the viable bud bank in a once-open savannah (*22*). Several decades after cultivation or mining, the composition of secondary grassland plant communities remains very different from that of nearby old-growth grasslands, lacking species with poor dispersal abilities and species regenerating from belowground organs (*10*, *23*). Belowground degradation can therefore cause grasslands to cross a hard-to-reverse threshold where restoration may be difficult or impossible within decades of these disturbances. Given the apparent existence of this threshold, it is vital that remaining old-growth grasslands are protected, particularly from the threats that affect belowground processes and structure, as we cannot rely on restoration to guide complete recovery after such degradation.

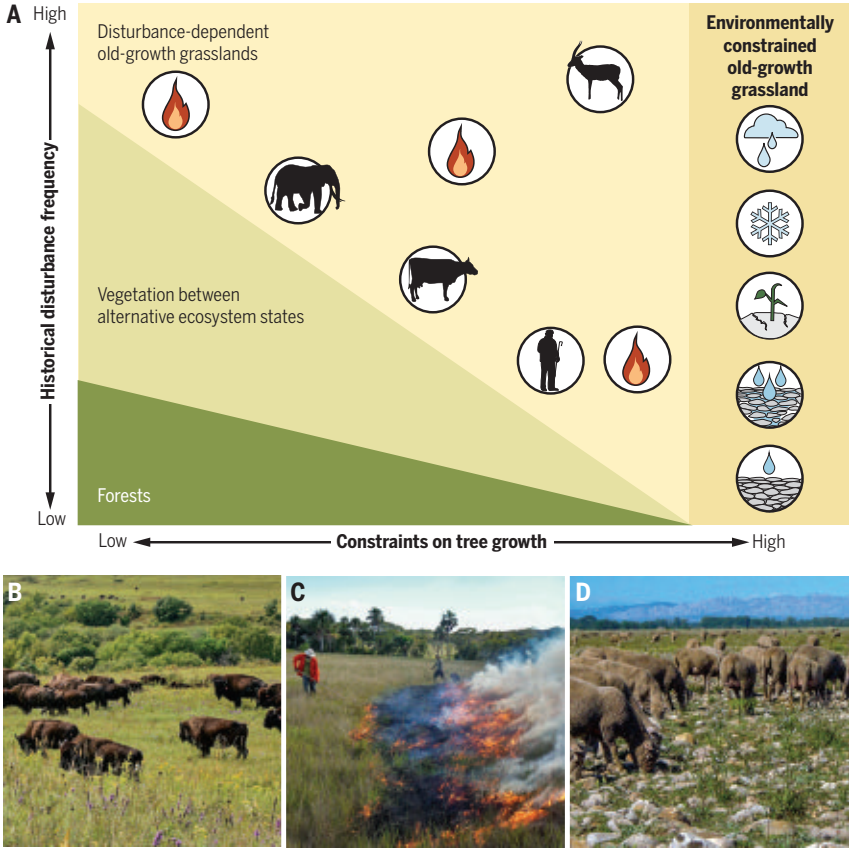


Fig. 3. Interactions among climate, soils, disturbance, and vegetation are key considerations for understanding old-growth grasslands as well as recovery trajectories in secondary grasslands. (**A**) On most soil types, the existence of disturbance-dependent grasslands (in light rose-color) is determined by interactions between soils and endogenous disturbances (fire, herbivory). Tree recruitment is limited by these disturbances. In environmentally constrained grasslands (in light brown), poor drainage (seasonally saturated or inundated soils), extremely low moisture-holding capacity (shallow, rocky soils), exceptionally low soil fertility, cold temperature, or low precipitation preclude dense tree cover, even in the absence of frequent disturbances. Disturbances and abiotic factors (circles, in no set order) that could result in exclusion of trees are placed as examples in each of the far left zones, respectively. In forests (dark green), dense tree cover constrains fire frequency and grazer abundance by limiting herbaceous plant productivity. The light green state space between disturbance-dependent old-growth grasslands and forests represents unstable vegetation (fire-excluded, tree-encroached grassland) in transition between alternative ecosystem states; old-growth grasslands and forests often co-occur in mosaics in such landscapes. (**B** to **D**) Examples of grasslands structured by different interactions. (B) Bison grazing in Konza prairie, where fire is needed to suppress woody encroachment. (C) Water saturation of the soil prevents tree establishment and fire maintains diversity in this wet grassland in Jalapão, Northern Brazil. (D) Sheep grazing in a Mediterranean grassland in Southern France, where pastoralism has coevolved with the system in a grassy state since the Holocene.

Interventions toward old-growth characteristics

In contrast to the early successional view of derived grasslands as a stage on their way to forests, restoring old-growth characteristics to altered or secondary grasslands requires attention to the development of a complex belowground structure akin to the aboveground complexity in an old-growth forest (*24*). A synthesis of 31 studies, including 92 time points on six continents, indicates that secondary grasslands may typically require at least a century, and more often millennia, to recover their former species richness (*23*). Even as their richness increases over decades to centuries, these grasslands still lack many characteristic old-growth grassland species and instead support more short-lived, early successional species than their old-growth counterparts. We know less about the timeline for belowground soil and structure development, but it likely corresponds with the timeline

MODIFIED FROM (*10*). DRAWINGS COURTESY OF S. LE STRADIC; PHOTOS: (B) BARB VAN SLYKE, (C) ALESSANDRA FIDELIS, (D) ELISE BUISSON

of these compositional dynamics (*25*). The increased appreciation of the temporal dimension of grassland assembly emphasizes the need of restoration to accelerate this trajectory and challenges the view that one initial period of active restoration will be sufficient to guide development. We highlight three advances driven by this increased appreciation below.

Focus interventions on disturbance-vegetation feedbacks

In cases where degradation has not had a catastrophic impact on belowground structure, it may be possible to reestablish broken feedbacks that then can guide recovery (*26*). Feedbacks among disturbance, vegetation, and belowground soil development have structured grasslands for millennia (*4*, *27*). Disturbance regimes select for functional traits of the vegetation, which then provide feedback to affect the intensity, frequency, and impact of disturbances (*28*). For instance, fire regimes vary in flammability depending on plant properties, and herbivore pressure varies depending on the quantity and quality of forage and habitat suitability for predator avoidance (*27*). The response of vegetation to these disturbances varies based on plant traits such as resprout ability, clonal growth, and seed recruitment (*26*, *28*). Feedbacks also extend to soils and soil organisms, as soils determine plant growth but are also products of the plants that grow on them (*29*).

As feedbacks in degraded grasslands differ in their nature and strength from those with more old-growth characteristics, reestablishing a disturbance regime in degraded grasslands may not result in expected effects of the disturbance or in the intended vegetation responses to the disturbance. Interventions simultaneously addressing disturbance and biota may be the best option to break the feedbacks that constrain recovery. For instance, there are examples of creative use of prescribed fire as a tool to recreate grazing habitat (*30*), and livestock can be managed in such a way as to initiate grazing habitat that supports large mammalian herbivores (*31*). Amendments such as biochar and mycorrhizal inoculum can shift the soil environment to be more suitable for native species, characteristics which can be maintained by slow growth and resource cycling of the vegetation (*32*, *33*). As the system recovers, these interventions also need to shift depending how the recovering biota affects disturbance dynamics and vice versa.

Breaking the cycle of invasion: Vegetation change that constrains recovery

Restoration in areas where an altered disturbance regime has resulted in woody encroachment or exotic herbaceous species invasion demonstrate the importance of viewing restoration as a set of interventions that iteratively move the system to a new system state (*10*, *34*).

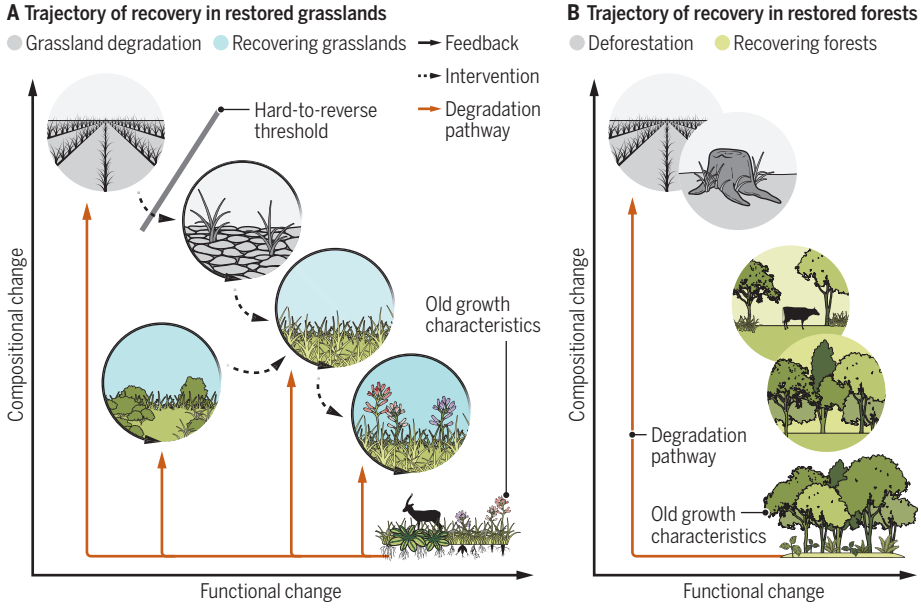


Fig. 4. Degradation pathways can result in differential loss of ecosystem function and diversity to old-growth grasslands, and the recovery of "old-growth" characteristics is dependent on the degree of functional change. Axes of functional and compositional change depict divergence from the reference characteristics [modified from (*23*)]. (**A**) The trajectory of recovery in restored grasslands (blue spheres) toward old-growth characteristics (lower right) is dependent on the degradation pathways (red arrows, ranging right to left from altered disturbance regimes to land use conversion) as well as vegetation-soil-disturbance feedbacks (black arrows) at each stage of recovery. Substantial belowground disturbance (e.g., tilling) may cause the system to cross a hard-to-reverse threshold (gray line) and woody encroachment shifts feedbacks and can lead to alternative trajectories. Iterative restoration interventions (dashed black arrows) that consider these feedbacks can result in progression back toward old-growth characteristics. (**B**) Forests show similar dynamics, where recovery to old-growth characteristics after deforestation may be hard if not impossible. An early recovery stage after deforestation may be a grassy stage (which we term a derived grassland), yet the recovery trajectory is toward forest. Restoration interventions may accelerate recovery.

Woody species can strongly influence disturbance regimes, and land managers have resorted to cutting, herbicides, and even plowing to remove trees—with striking consequences for the remaining biodiversity. Extreme fires (firestorms) have been applied in heavily encroached areas using spiral ignitions or extreme weather days to try to reverse the woody cover and reinitiate ecologically relevant feedbacks (*35*). Once the grassy understory has been reduced to the point that it cannot carry a fire or support grazers, woody encroachment becomes more difficult to reverse (*36*), requiring the replanting of herbaceous vegetation alongside the initiation of disturbance regime for recovery feedbacks.

When invasive species are grasses, they can often maintain disturbance regimes that benefit short-lived ruderal life histories, preventing transitions to the belowground complexity and allocation that characterize old-growth grasslands (*37*). High accumulation of litter and standing dead biomass changes local fire behavior, and a dependence on seed recruitment often confers advantage for invasives under this disturbance regime (*38*). Dominance in the seed bank and difficulty reestablishing long-lived natives can make this feedback particularly difficult to address. One strategy is to enhance the ability for natives to recruit by seed via seed enhancement technology (e.g., seed coating or pelleting aimed at mitigating the conditions that limit establishment) (*20*), potentially addressing priority effects (i.e., the order in which plants are reintroduced) that influence species dominance in early stages of restoration (*39*).

Overlooked old-growth grassland species

One important restoration question is how to accelerate or facilitate species turnover toward old-growth species composition and associated belowground function. Worldwide, grasslands are often restored by sowing seeds (*40*). However, as many species have developed colonization and survival strategies that are based on belowground buds and clonal growth (*23*, *41*) rather than on seeds, additional techniques may be needed to restore old-growth characteristics. Seeding fast-growing species can impede long-term restoration success by creating communities with low resilience to natural disturbance, such as fire, and excluding the longer-lived species from restoration (*42*). In fact, there may be many grasslands where seeded species maintain dominance long after restoration, spurring

reconsideration of whether actions are achieving the desired old-growth structure (*43*).

Although bud-bearing belowground organs can persist in the absence of disturbance for some time in a degraded grassland (*44*), how long is still unclear. Once these belowground structures are gone, we have little understanding of how to reintroduce this component of the vegetation (*24*). Topsoil transfer has shown some success in broadening the type of species that restoration can reintroduce (*45*), yet even this technique favors species with high seed bank allocation. Vegetative propagation—such as micropropagation, transplantation of seedlings, and individual tillers—is often needed (*24*) but is hard to conduct at scale, with open questions about protocols, spatial configuration of planting, and genetic sourcing. Techniques aimed at speeding the establishment of bud banks and belowground organs in a restoration have shown promise but are just in their infancy (*24*, *41*).

Global change as a challenge and opportunity

Global climate change frames the emerging perspective of long-term assembly toward old-growth characteristics in grassland restoration. Climate controls the distribution of grasslands in some regions, influences the feedbacks and threshold shifts that determines where grasslands persist, and, in virtually all regions, can have a strong influence on the interventions needed to restore feedbacks (*14*, *46*). Depending on the degree to which climate influences these processes, it may also affect the historical approach to the determination of grassland types and disturbance regimes (*12*). For instance, changes such as elevated atmospheric CO_2, which exacerbates invasion of woody species, would require novel disturbance regimes to aim for a grassy state.

Given the strong feedbacks between composition and disturbances in grassland recovery, shifts in climate may exert large influences on the assembly process. In some cases, it may be important to let climate effects shift restoration trajectories, as climate can guide species composition or characteristics to those most able to tolerate future conditions (*47*). Restoration efforts under a climate change scenario may thus target not only which species should be present at a given site, but also functional diversity, soil structure, and the belowground component. In this way, the system may be able to recover from an extreme event, as the presence of a viable bud bank and underground storage organs ensures the resilience of the system (*48*). However, letting climate effects shift restoration trajectories might also be undesirable if it endangers fundamental feedbacks in the trajectory of the system toward old-growth functional characteristics (*46*) by, for instance, selecting for species with greater aboveground allocation characteristics. As belowground complexity is a characteristic that develops over long time horizons, understanding how climate influences priority effects and feedbacks that affect recovery trajectories is critical.

Climate change will add difficulty to the already difficult challenge of restoring old-growth grasslands that resemble specific reference sites, as these ancient grassland references developed in a different time, disturbance regime, and climate. Yet we expect that restoring old-growth characteristics in these situations, prioritizing processes such as belowground complexity and functional diversity (*49*), should enable resilience and facilitate adaptation to future change while still maintaining character, functions, and services that embody these globally important systems.

Outlook

As we enter the United Nations Decade on Ecosystem Restoration, advances in restoration science and practice in grasslands are critical if we are to combat the loss of old-growth grasslands and the decline of biodiversity (*50*). However, in the rush to provide nature-based solutions to tackle climate change, tree planting in grasslands has become synonymous with restoration in many regions (*13*). At the same time, the high demand for arable land continues to spur conversion to agriculture. These are irreversible actions, ignoring the belowground soil-locked carbon storage in these old-growth grasslands as well as the hard road to restore their belowground complexity and their biodiversity once they are lost.

Although there are many challenges ahead, viewing grassland restoration as assembly toward old-growth characteristics with unique biota and belowground complexity will enable us to achieve ambitious restoration goals for Earth's grassy ecosystems. Given that grassland recovery involves strong feedbacks among vegetation, disturbance, and soils, as well as the lengthy time horizon for recovery, future progress depends on creative interventions that focus on iterative management, taking into account changes in grassland assembly over time. Techniques to reestablish species characteristic of old-growth grasslands, given their belowground structure and limited recruitment by seed, will require looking beyond or augmenting traditional seeding techniques. Metrics of belowground complexity and functional diversity will be critical guideposts to track trajectories in development and assess success. We urge conservation initiatives to safeguard against the conversion of old-growth grasslands for tree planting or tillage agriculture, to maintain our ancient biodiverse grasslands with appropriate disturbance regimes, and to emphasize the long-term restoration of grasslands in efforts to restore Earth's biodiversity.

REFERENCES AND NOTES

1. J. Bengtsson *et al*., *Ecosphere* **10**, e02582 (2019).
2. B. P. Murphy, A. N. Andersen, C. L. Parr, *Philos. Trans. R. Soc. London Ser. B* **371**, 20150319 (2016).
3. C. L. Parr, C. E. Lehmann, W. J. Bond, W. A. Hoffmann, A. N. Andersen, *Trends Ecol. Evol.* **29**, 205–213 (2014).
4. W. Bond, *Plant Ecol. Divers.* **14**, 205–222 (2022).
5. G. Overbeck *et al*., *Divers. Distrib.* **21**, 1455–1460 (2015).
6. T. Sohl *et al*., *Agric. Ecosyst. Environ.* **153**, 1–15 (2012).
7. N. Dudley *et al*., *Restor. Ecol.* **28**, 1313–1317 (2020).
8. W. J. Bond, N. Stevens, G. F. Midgley, C. E. R. Lehmann, *Trends Ecol. Evol.* **34**, 963–965 (2019).
9. J. P. Bakker, F. Berendse, *Trends Ecol. Evol.* **14**, 63–68 (1999).
10. E. Buisson *et al*., *Biol. Rev. Camb. Philos. Soc.* **94**, 590–609 (2019).
11. J. W. Veldman *et al*., *Front. Ecol. Environ.* **13**, 154–162 (2015).
12. N. Shackelford, J. Dudney, M. Stueber, V. Temperton, K. Suding, *Restor. Ecol.* **29**, e13541 (2021).
13. J. W. Veldman *et al*., *Science* **366**, eaay7976 (2019).
14. A. Volder, D. D. Briske, M. G. Tjoelker, *Glob. Change Biol.* **19**, 843–857 (2013).
15. O. Valkó, S. Venn, M. Zmihorski, I. Biurrun, R. Labadessa, J. Loos, *Hacquetia* **17**, 5–16 (2018).
16. A. Dabengwa, L. Gillson, W. Bond, *Environ. Res. Lett.* **16**, 055002 (2021).
17. C. L. Solofondranohatra *et al*., *Proc. R. Soc. B* **287**, 20200598 (2020).
18. R. C. R. Abreu *et al*., *Sci. Adv.* **3**, e1701284 (2017).
19. J. Kolbek, R. Alves, *Acta Univ. Carolin. Environ.* **22**, 111–130 (2008).
20. V. S. Brown *et al*., *Sci. Total Environ.* **798**, 149096 (2021).
21. R. Bardgett *et al*., *Nat. Rev. Earth Environ.* **2**, 720–735 (2021).
22. A. Ferraro *et al*., *Appl. Veg. Sci.* **24**, (2021).
23. A. N. Nerlekar, J. W. Veldman, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18550–18556 (2020).
24. E. Buisson *et al*., *Restor. Ecol.* **29**, e13292 (2021).
25. M. De *et al*., *Soil Sci. Soc. Am. J.* **84**, 568–586 (2020).
26. J. G. Pausas, B. B. Lamont, S. Paula, B. Appezzato-da-Glória, A. Fidelis, *New Phytol.* **217**, 1435–1448 (2018).
27. S. Archibald *et al*., *Environ. Res. Lett.* **13**, 033003 (2018).
28. B. Wigley *et al*., *Aust. J. Bot.* **68**, 473–531 (2020).
29. Y. A. Chung, S. L. Collins, J. A. Rudgers, *Ecology* **100**, e02756 (2019).
30. S. Archibald, G. P. Hempson, *Philos. Trans. R. Soc. London Ser. B* **371**, 20150309 (2016).
31. J. R. Goheen *et al*., *Ann. N.Y. Acad. Sci.* **1429**, 31–49 (2018).
32. G. House, J. Bever, *Restor. Ecol.* **28**, 785–795 (2020).
33. E. R. Wubs, W. H. van der Putten, M. Bosch, T. M. Bezemer, *Nat. Plants* **2**, 16107 (2016).
34. S. L. Collins *et al*., *Ecol. Lett.* **24**, 636–647 (2021).
35. C. H. Bielski, R. Scholtz, V. M. Donovan, C. R. Allen, D. Twidwell, *J. Environ. Manage.* **291**, 112550 (2021).
36. A. Bombo, F. Siebert, A. Fidelis, *Afr. J. Range Forage Sci.* **39**, 16–26 (2022).
37. M. A. Williamson *et al*., *Biol. Invasions* **22**, 663–680 (2020).
38. G. Damasceno, A. Fidelis, *J. Environ. Manage.* **271**, 111016 (2020).
39. E. Weidlich *et al*., *Restor. Ecol.* **29**, e13317 (2021).
40. D. Slodowicz, J. Humbert, R. Arlettaz, *Environ. Evid.* **8**, 28 (2019).
41. G. Ottaviani *et al*., *Trends Ecol. Evol.* **35**, 763–766 (2020).
42. A. Giles *et al*., *Restor. Ecol.* **30**, e13474 (2022).
43. E. Grman *et al*., *Restor. Ecol.* **29**, e13281 (2021).
44. A. B. Bombo, B. Appezzato-da-Glória, A. Fidelis, *Oecologia* **199**, 153–164 (2022).
45. M. Ferreira, D. Vieira, *Ecol. Eng.* **103**, 1–12 (2017).
46. B. Wilsey, *Restor. Ecol.* **29**, e13132 (2021).
47. L. W. Jochems, J. A. Lau, L. A. Brudvig, E. Grman, *Ecol. Appl.* **32**, e02487 (2022).
48. H. van Coller, F. Siebert, *Afr. J. Ecol.* **58**, 236–250 (2020).
49. J. M. Bullock *et al*., *Ecography* **2022**, ecog.05780 (2022).
50. K. Suding *et al*., *Science* **348**, 638–640 (2015).

ACKNOWLEDGMENTS

We thank the editors of *Science* for the invitation to write this review, three reviewers for comments on the manuscript, as well as the many scientists and practitioners whose work is advancing our understanding of grassland restoration. **Funding:** Future Ecosystems for Africa Program, Oppenheimer Generations Research and Conservation (S.A.); National Council for Scientific and Technological Development 312689/2021-7 (A.F.); and NSF grant 2044006 and USDA grant AFRI 1024877 (K.N.S.). **Author contributions:** Conceptualization: E.B., K.N.S.; writing–original draft: E.B., S.A., A.F., K.N.S.; writing–review and editing: E.B., S.A., A.F., K.N.S. **Competing interests:** The authors declare that they have no competing interests.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse

Submitted 3 May 2022; accepted 15 June 2022
10.1126/science.abo4605

REVIEW

Molecular, cellular, and developmental foundations of grass diversity

Paula McSteen[1]* and Elizabeth A. Kellogg[2]*

Humans have cultivated grasses for food, feed, beverages, and construction materials for millennia. Grasses also dominate the landscape in vast parts of the world, where they have adapted morphologically and physiologically, diversifying to form ~12,000 species. Sequences of hundreds of grass genomes show that they are essentially collinear; nonetheless, not all species have the same complement of genes. Here, we focus on the molecular, cellular, and developmental bases of grain yield and dispersal—traits that are essential for domestication. Distinct genes, networks, and pathways were selected in different crop species, reflecting underlying genomic diversity. With increasing genomic resources becoming available in nondomesticated species, we anticipate advances in coming years that illuminate the ecological and economic success of the grasses.

Most people reading this review will have either eaten, stepped on, or burned a grass within the past 24 hours. Humans have been cultivating grasses for at least 10,000 years and likely consumed them for millennia before that. Of the crops that feed the world, the big three—wheat, maize, and rice (Fig. 1)—provide 50% of calories consumed by humans as well as protein and micronutrients, are grown over the widest area, and have the highest economic value (*1*). In addition, so called "orphan crops," such as tef, sorghum, fonio, and various millets, most of which are native to Africa, grow well with less intense agricultural inputs and are poised to be cultivated more widely to serve a warmer, drier planet. Meat, eggs, and dairy products are the products of animals that consume forage, pasture, and prairie grasses.

Moreover, some of the most devastating agricultural weeds, such as Johnson grass in corn fields and barnyard millet in rice fields, are grasses (*2*). Grasses also underpin the beverage industry; the world wouldn't have beer without barley (Fig. 1) or rum without sugarcane, with the latter being used to produce not only sugar but also biofuel (*1*). Turf grasses beautify cultivated landscapes and provide the playing surface for golf courses, tennis courts, cricket pitches, and other sports fields. Grasses such as Miscanthus and switchgrass are being developed for lignocellulosic biomass, and perennial grasses, such as intermediate wheatgrass, may help store carbon below ground. Bamboos (and even giant reeds) are used for construction. Yet despite this diverse repertoire, only a small subset of the ~12,000 species of grasses are used by humans (*2*, *3*).

[1]Division of Biological Sciences, Bond Life Sciences Center, Interdisciplinary Plant Group, University of Missouri, 1201 Rollins Street, Columbia, MO 65211, USA. [2]Danforth Plant Science Center, St. Louis, MO 63132, USA.
*Corresponding author. Email: mcsteenp@missouri.edu (P.M.); ekellogg@danforthcenter.org (E.A.K.)

Like orchids, lilies, asparagus, and pineapples, the grasses (family Poaceae or Gramineae) have a single seedling leaf (cotyledon) and are placed in the large clade of monocotyledonous flowering plants (monocots). The grasses constitute ~20% of the ~60,000 species of monocots (*4*). Thus, all grasses are monocots, but most monocots are not grasses. Grasses that produce seed that is cultivated agronomically and eaten by humans and animals are often called cereals.

Morphological and physiological diversity

Grasses are ecologically dominant in vast areas of all the continents except for Antarctica (*2*, *5*). Even in areas with some tree cover, grasses form a broad understory. The grass family may have originated more than 80 million years ago, extending its continental reach during the late Miocene grassland expansion (8 million to 3 million years ago), although its current distribution also reflects extensive climatological change since then (*5*, *6*).

Broad physiological adaptations permit grasses to thrive in disparate environments. Most grasses are tropical, but one major group, subfamily Pooideae, has spread widely in cool and cold areas, even reaching Antarctica (*7*). Whereas some genetic components of their cold tolerance are widely shared stress responses, others represent the repurposing of loci involved in other physiological responses (*8*). Such loci include ones that regulate the induction of flowering after cold (vernalization), as in winter wheat (*9*). Among the tropical grasses, high-efficiency (C_4) photosynthesis has originated 22 to 24 times (*10*), with the physiological and anatomical bases of the pathway being subtly different each time. Our cultivated cereals are mainly annual, grown for their ability to complete their life cycle (seed to seed) in one growing season, but most species of grasses are perennial (*2*). The genetic mechanisms underlying the shift from perennial to annual are unknown but are likely diverse (*11*).

Genomic diversity

The genomes of grasses are largely collinear for all species in the family, that is, the genes are in roughly the same order (*12*). This broad similarity allows genes identified in one species to be discovered in a second species, permitting the grasses to function as "a single genetic system" (*12*). All grass genomes also share large regions of duplicated genes, which points to a polyploidization event in the common ancestor of the family [e.g., (*13*, *14*)]. Polyploidization events have continued to occur frequently throughout the evolution of the family, with some authors estimating that as many as 75 to 80% of the species are recent polyploids (*15*).

Beneath this broadly conserved genome architecture lurks extensive diversity, including variation in nucleotides (single-nucleotide polymorphisms), gene structure, and even the presence or absence of genes [e.g., (*16*, *17*)]. The nucleotide differences between two lines of

Fig. 1. Diversity of grass inflorescence morphology. (**A**) In wheat, the unbranched spike produces single spikelets (inset) with multiple florets. (**B**) In barley, the unbranched spike produces triplet spikelets (inset). In this two-row variety, only the central spikelet produces a floret. (**C**) Rice has many branches and produces single spikelets (inset) with a single floret. (**D**) Maize produces many branches with paired spikelets (inset) that each produce two florets.

PHOTOS: (A, B, C) ELIZABETH A. KELLOGG; (D) PAULA MCSTEEN

Zea mays (maize) are greater than those between humans and chimpanzees (*18*). Genes central for plant structure in maize are missing in wheat and rice, and vice versa (*19*). In other words, not all grasses have the same complement of genes, and their morphology is altered accordingly.

The grain: A grass-specific structure

The grass fruit (grain or caryopsis) is the innovation that characterizes all grasses (*2*) (Fig. 2A). The grain develops from fusion of the single seed to the inner wall of the ovary, creating a single solid structure. The wheat "seeds" sold in the grocery store are in fact grains, with the bran made up of the ovary wall plus the seed coat. Within the grain is the young embryo (the germ, in wheat), which is a well-formed little plant with multiple leaf primorida and shoot and root apical meristems (the stem cells that give rise to all organs in the plant) (Fig. 2B). Development of the grass embryo progresses a long way before the fruit is shed from the plant, distinct from that in other closely related monocot families in which the embryo is a globular, scarcely differentiated mass of cells at fruit maturation (*2*).

The grain and the inflorescence that bears it have been the focus of both natural and human selection for grain size and number and dispersal. The starch-filled endosperm and oil-filled embryo of grains made wild grasses an obvious source of food for human ancestors. The early process of converting these wild species into ones that could be cultivated year after year is well known and is described in many biology textbooks. Traits in this familiar "domestication syndrome" may include (i) cultivated plants with grains that are larger than those of their wild ancestors and do not drop off the plant, (ii) lack of dormancy, (iii) loss of awns (wheat, sorghum, oat, rice), and (iv) increased grain number. We will focus on the developmental, cellular, and molecular bases of two of these traits: failure of seed drop (called "loss of shattering") and grain number.

Shattering: Useful in the wild, a liability in cultivation

An early step in grass domestication is selection for mutations that let the plant hold onto its seeds rather than drop them in the dirt. The annual cycle of reaping and planting automatically selected for grains that were held more firmly than those in wild undomesticated plants and, over time [possibly ~1000 years (*20*)], led to domesticated plants in which the flower stalks fail to break easily, so-called nonshattering varieties. Lack of shattering was selected independently in most known domestication events in cereals (*21*).

The close relationship and genomic similarities among the cereal crops suggested that perhaps loss of shattering in wheat, sorghum, rice, and others could have occurred by repeated modifications of the same underlying genes. However, a series of quantitative genetic locus studies (*22*) and subsequent studies that looked at the expression of genes involved in forming the break point itself (abscission zone) have found extensive differences among the crops (*23*). Genes that are mutated in domesticated wheat (*brittle rachis 1* and *2*) are unrelated to those in rice (*shattering4* and *5*), which in turn are distinct from those in sorghum and millet [e.g., *less shattering1*, which is reviewed in (*24*)]. The one exception may be a locus known as *shattering1* (*sh1*) in sorghum (*25*), which is also mutated in domesticated rice and foxtail millet (*22*, *26*). *sh1* is a transcription factor in the YABBY family (named for the distinctive DNA binding domain known as a yabby domain), but its precise molecular function remains unknown.

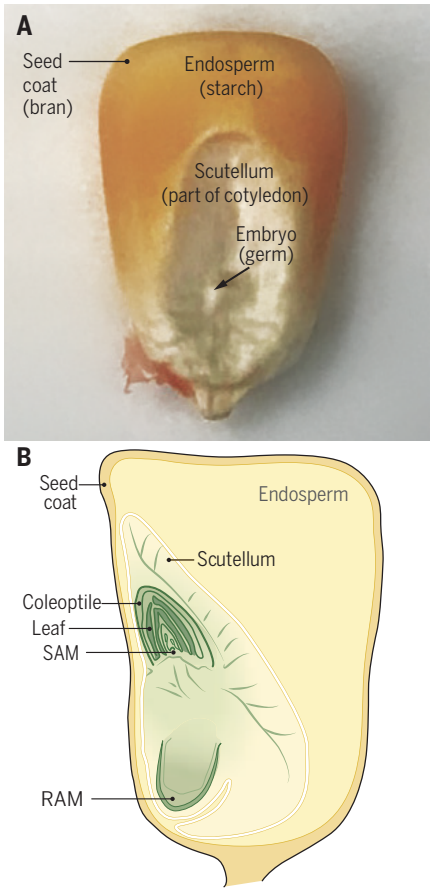


Fig. 2. Grains characterize all grasses. (**A**) Photograph of a maize grain indicating the starch-filled endosperm and the seed coat. The scutellum (part of the first leaf or cotyledon) and the embryo are visible on the adaxial side of the kernel. (**B**) Diagram of a longitudinal section through the well-developed maize embryo indicating the coleoptile and scutellum (which make up the first leaf or cotyledon), multiple leaf primordia, the shoot apical meristem (SAM), and the root apical meristem (RAM).

Spontaneous reversal of domestication, in which shattering has been reacquired independently, has created grasses that grow as weeds within the crop; such dedomestication has been documented in at least four lineages of rice, as well as in a few other grasses [reviewed in (*27*)]. The underlying domestication mutations are still present in the newly weedy rice, but the weedy populations have additional mutations that lead to shattering, each using different sets of genes (*28*).

Even among wild grasses, shattering appears to occur by different mechanisms, which may explain the distinct sets of mutations in the different domestication events. The break point forms in different positions in different lineages of grasses (*2*). Breakage occurs below the flower (often called a floret) in many species, such that the grain is shed along with floral organs and subtending bracts, but in other grasses (including the many species of millet), breakage occurs below the clusters of flowers (called spikelets) so that several flowers fall off the plant at once (Fig. 3). In still other species such as wheat and barley relatives, the inflorescence stalk breaks up. Cellular details and cell wall structures also differ among species, but the cell wall differences do not correlate with the location of the abscission zone or with evolutionary relationships (*29*).

Specific sets of genes characterize the abscission zones of rice, *Brachypodium*, and green millet, but the abscission-specific genes are almost completely nonoverlapping (*23*). Only two, a MYB transcription factor and a lysine decarboxylase, are specific to the abscission zone of all species (*23*). *sh1* is commonly upregulated in the abscission zone but is also expressed more widely, suggesting that its function in the abscission zone is part of a larger spikelet developmental network.

Despite years of investigation, the precise process of shattering in grasses remains unknown. Most of the genes that affect the process are transcription factors, often from well-known gene families that affect other aspects of plant development. One compelling hypothesis is that the process of shattering is not a single mechanism but rather a set of mechanisms that have evolved over time.

Grain yield: A diversity of mechanisms

Because a grass flower (floret) produces only one grain (at most), the number and arrangement of flowers directly affect the yield. The number of grains is thus affected by the number of flowers per spikelet, the number of spikelets per branch, and the number of branches per inflorescence, all of which vary among species (Fig. 4). Furthermore, many species have inflorescences that top vegetative

PHOTO: PAULA MCSTEEN

branches, called tillers, further contributing to grain yield. Complicating the picture, grain weight and number of grains are generally inversely correlated, so simple selection for more grains leads to more smaller grains (*30*). Because of the complexity of how flowers are produced, increased grain number can be achieved by any number of different mechanisms.

Domestication and postdomestication breeding of cereals have led to an increase in the number of grains produced (increased yield) compared with that produced by the wild ancestor. For example, hybrid maize bears 16 to 22 rows of kernels around the circumference of the cob, substantially more than the wild ancestor teosinte, which bears only two rows. The number of rows is always an even number because maize produces its spikelets in pairs (Fig. 4), as do all other members of the tribe Andropogoneae, including sorghum, sugarcane, and Miscanthus (*2*). Paired spikelets have also arisen independently in the related tribes Paspaleae (e.g., seashore paspalum) and Paniceae (e.g., fonio, crab grass). But how do these grasses produce pairs in the first place? The vast majority of grasses, like rice, produce spikelets singly, although another cereal, barley, produces spikelets in triplets. Wheat produces spikelets singly, but mutations can cause the formation of paired or triple spikelets, indicating that wheat has the underlying genetic capacity to produce additional grain. Could understanding these mechanisms be used to increase grain number in cereals or in grasses or orphan crops to be domesticated in the future?

Multiple mechanisms have led to the variation in inflorescence morphology observed in grasses during evolution and domestication (Figs. 1 and 4). Determination of the molecular, cellular, and developmental bases of these phenotypes indicates that similar phenotypes in one species can be caused by different pathways or that orthologous genes can cause different phenotypes in different species. In the following sections, we discuss three mechanisms involved in the diversity of morphology in cereal grasses.

To branch or not to branch

Multiple genetic pathways control branching in grass inflorescences (*19*). Mutations in these pathways can lead to increased branching and increased grain number, so it is not surprising that these pathways have been selected in the evolution, domestication, and breeding of cultivated cereals. However, a variety of different pathways have been used in different cereals (*19*). We discuss just two of these pathways below.

In maize, expression of *ramosa2* (*ra2*), which encodes a transcription factor in the lateral organ boundary (LOB) domain family, acts upstream of *ramosa1* (*ra1*), which encodes a zinc-finger transcription factor and controls the abrupt switch from producing branches to producing spikelet pairs [reviewed in (*19*)]. The expression pattern of *ra1* in *Miscanthus* and sorghum, both of which also produce spikelets in pairs and are in the same clade as maize, also correlates with the branch-to–spikelet pair transition, albeit later, correlating with an increased number of branches (*31*). However, *ra1* is not found in rice, barley, wheat, or other members of their subfamilies that do not produce spikelets in pairs (*19*). Conversely, mutations in the ortholog of *ra2* in barley also increase branching and are associated with phenotypic differences between two-row and six-row barley (Fig. 4) (*9*). Therefore, the genetic network regulated by *ra2* differs between major groups of grasses even though the protein itself is conserved.

Recent progress has been made in understanding the genetic basis for the unbranched spike morphology in wheat and barley by the *compositum1* (*com1*) and *com2* loci [reviewed in (*32*)]. Whereas *com1* orthologs do not regulate branching in maize and rice (*33*), the function of *com2* appears to be somewhat conserved. *com2* mutations increase branching and spikelet number in barley and cause the production of paired or, rarely, triple spikelets in "miracle wheat," which is so called because of its increased grain yield (*32*, *34*). Mutations in the orthologous gene also increase branching in maize and rice, but the additional spikelets do not produce florets and are sterile, and hence do not increase yield (*35*). However, mutations in the promoter of the rice ortholog, which cause reduced rather than complete loss of function, increase spikelet number and yield and thus may be valuable for breeding (*36*). Evolutionary analysis of *com2* orthologs identified signatures of selection at particular amino acids in rice, wheat, and barley (*37*), although their functional importance remains to be determined.

Growth suppression

Another mechanism for altering branching would be to suppress the outgrowth of structures that have already been formed. For example, increased expression of several transcription factors, including *teosinte branched1* (*tb1*) and *grassy tillers* (*gt1*), has led to the suppression of tiller buds during domestication in maize, and these transcription factors are proposed to have conserved roles in regulating tiller number in wheat and rice [reviewed in (*35*)]. *tb1* and *gt1* have been used repeatedly in cereals for different purposes other than tiller number. For example, *gt1* was co-opted in sex determination in maize (*38*), and orthologs of *tb1* or *gt1* are used in the suppression of spikelets in two-row barley (*35*). Furthermore, in wheat, loss of function of *tb1* and interactions with flowering-time genes cause production of paired spikelets (*39*, *40*). Thus, changes in expression (or the targets) of transcription factors that cause growth suppression could be very powerful in causing phenotypic changes.

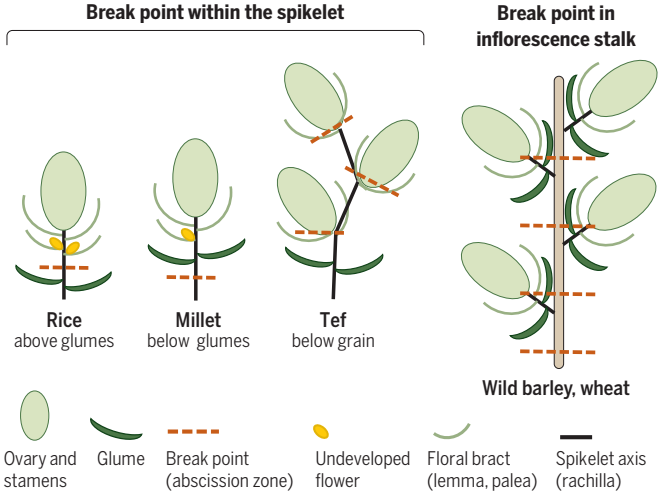


Fig. 3. Diagrams of spikelets (rice, millet, tef) and inflorescence (wheat, barley) showing different positions of break points (abscission zones). Modified leaves known as glumes (dark green arcs) mark the base of the spikelet and provide critical positional landmarks for comparisons. The break-point position above the glumes, as in rice, is common and ancestral in the grass family (*23*). The position below the glumes, as in millet, predominates in the subfamily Panicoideae. Few grasses break right below the grain, as in teff. The breakable inflorescence stalk is common not only in wild relatives of wheat and barley but also in maize.

Meristem size matters

One mechanism to increase grain number in maize and rice is to increase the size of the apical inflorescence meristem (*41*). A conserved signaling pathway involving proteins in the *CLAVATA* (*CLV*) and *WUSCHEL* (*WUS*) families regulates the plant growth hormone, cytokinin, which affects the size and number of stem cells in the meristem. Mutations that affect signaling in the *CLV-WUS* pathway can increase meristem size, row number, and yield in maize and green millet (*42*, *43*) but increase floral organ number in rice (*44*). Despite these differences, a screen for alleles with signatures of selection in both maize and rice identified the same locus, which increases yield in both species through an

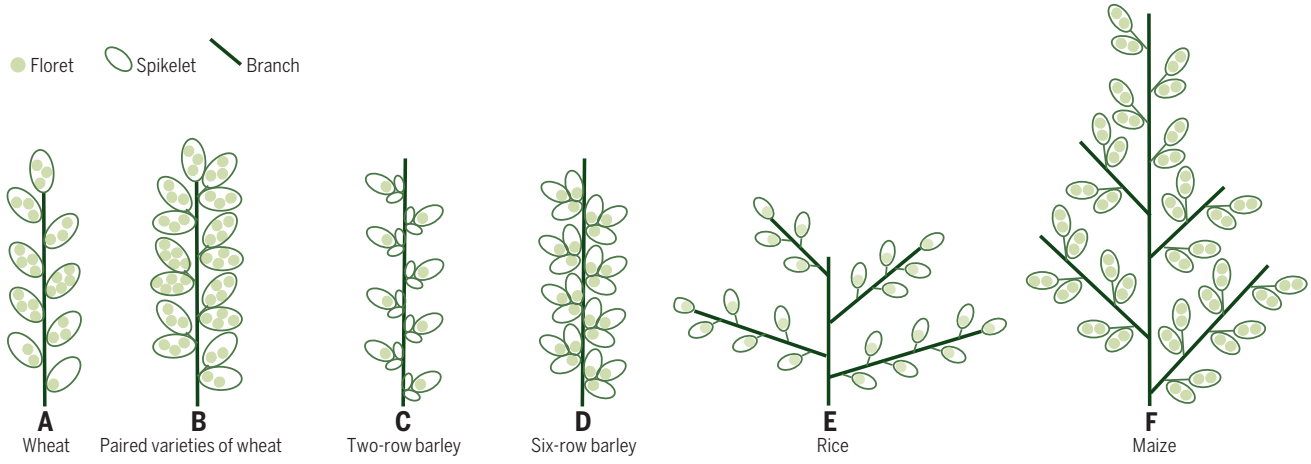


Fig. 4. Mutations that influence spikelet number in cereals provide insights into evolutionary mechanisms. (**A**) In wheat, the unbranched spike produces spikelets with a variable number of florets (average of three). (**B**) Mutations in wheat can cause the production of paired spikelets, similar to maize. (**C**) In two-row barley, only the central spikelet produces a floret. (**D**) In six-row barley, all three spikelets produce a floret and set seed. (**E**) Rice has many branches and produces single spikelets with a single floret. Mutations in rice that increase yield increase branch number and reiterate branches on the branches. (**F**) Maize produces many branches with paired spikelets, each of which produces two florets. Mutations in maize can cause the production of single spikelets, similar to rice and wheat, or can convert spikelet pairs to branches.

increase in cytokinin and cell division (*45*). Mutations that increase cytokinin levels or signaling also increase the number of branches and yield in rice (*46*), and cytokinin has been implicated in branching in barley (*47*). However, the *CLV-WUS* pathway has not yet been functionally characterized in wheat and barley, although it is an obvious target for crop improvement.

Meristems that produce multiple spikelets are larger than single-spikelet meristems. Such meristems include the spikelet-pair meristem in maize, the mutant paired-spikelet or triple-spikelet meristems in wheat, and the triple-spikelet meristem in barley; the latter extends over almost half the circumference of the inflorescence (*48*). In maize, defects in the *CLV-WUS* pathway or the plant growth hormone auxin can cause the production of single instead of paired spikelets (*49*, *50*). It seems likely that similar pathways are involved in the production of the triple-spikelet meristem in barley and in the independent origins of the paired spikelets in grasses. However, multiple ligands, receptors, and transcription factors, and even parallel pathways, converge on the *CLV-WUS* pathway in different meristem types, so the pathways that specify each meristem type in each crop will need to be identified.

Outlook

Grasses are an economic and ecological success story. We speculate that the large endosperm and well-developed embryo that are characteristic of grasses (Fig. 2) gave grains a head start in germination and seedling survival, in both ecological and agricultural settings. Grass genomic diversity provides the raw material for their morphological diversity. Genomic sequencing has provided insights into the genetic basis of domestication and post-domestication breeding of cereal genomes [reviewed in (*51*)], the development of wood in bamboo (*52*), and the multiple independent origins of cold tolerance, photoperiod insensitivity, and C_4 photosynthesis (*7*, *10*). The availability of functional genomics tools (*53*) will provide opportunities to move from genes to networks and to determine which parts of the pathway are conserved and which are species specific. These networks will enable modern-day agriculturists to determine how to domesticate orphan crops such as tef and fonio and to begin to understand how grasses have covered the world.

REFERENCES AND NOTES

1. Food and Agriculture Organization of the United Nations (FAO), FAOSTAT agriculture production tables (2022); https://www.fao.org/faostat/en/#data.
2. E. A. Kellogg, *Poaceae*, vol. 13 of *The Families and Genera of Vascular Plants*, K. Kubitzki, Ed. (Springer, 2015).
3. R. J. Soreng *et al.*, *J. Syst. Evol.* **55**, 259–290 (2017).
4. P. F. Stevens, Angiosperm phylogeny website, version 14 (2017); http://www.mobot.org/MOBOT/research/APweb/.
5. D. M. Griffith *et al.*, *New Phytol.* **228**, 15–23 (2020).
6. E. J. Edwards *et al.*, *Science* **328**, 587–591 (2010).
7. M. Schubert, A. M. Humphreys, C. L. Lindberg, J. C. Preston, S. Fjellheim, *Annu. Plant Rev. Online* **3**, apr0739 (2020).
8. J. Zhong, M. Robbett, A. Poire, J. C. Preston, *New Phytol.* **217**, 925–938 (2018).
9. L. E. Dixon, W. van Esse, D. Hirsz, V. Willemsen, S. M. McKim, *Annu. Plant Rev. Online* **5**, apr0648 (2020).
10. Grass Phylogeny Working Group II, *New Phytol.* **193**, 304–312 (2012).
11. J. Friedman, *Annu. Rev. Ecol. Evol. Syst.* **51**, 461–481 (2020).
12. J. L. Bennetzin, M. Freeling, *Trends Genet.* **9**, 259–261 (1993).
13. A. S. Seetharam *et al.*, *Front. Plant Sci.* **12**, 710383 (2021).
14. P. F. Ma *et al.*, *Plant Cell* **33**, 846–864 (2021).
15. J. H. Hunziker, G. L. Stebbins, in *Grass Systematics and Evolution*, T. R. Soderstrom, K. W. Hilu, C. S. Campbell, M. E. Barkworth, Eds. (Smithsonian Institution Press, 1987), pp. 179–187.
16. S. Mamidi *et al.*, *Nat. Biotechnol.* **38**, 1203–1210 (2020).
17. M. B. Hufford *et al.*, *Science* **373**, 655–662 (2021).
18. E. S. Buckler, N. M. Stevens, in *Darwin's Harvest*, T. Motley, N. Zerega, H. Cross, Eds. (Columbia Univ. Press, 2005), chap. 4, pp. 67–90.
19. E. A. Kellogg, *Plant Cell* **34**, koac080 (2022).
20. D. Q. Fuller, *Ann. Bot.* **100**, 903–924 (2007).
21. M. D. Purugganan, D. Q. Fuller, *Nature* **457**, 843–848 (2009).
22. A. N. Doust, M. Mauro-Herrera, A. D. Francis, L. C. Shand, *Am. J. Bot.* **101**, 1759–1769 (2014).
23. Y. Yu, H. Hu, A. N. Doust, E. A. Kellogg, *New Phytol.* **225**, 1799–1815 (2020).
24. Y. Yu, E. A. Kellogg, *Annu. Plant Rev. Online* **1**, apr0619 (2018).
25. Z. Lin *et al.*, *Nat. Genet.* **44**, 720–724 (2012).
26. S. Odonkor *et al.*, *Front. Plant Sci.* **9**, 918 (2018).
27. N. C. Ellstrand *et al.*, *Evol. Appl.* **3**, 494–504 (2010).
28. X. Qi *et al.*, *Mol. Ecol.* **24**, 3329–3344 (2015).
29. Y. Yu, P. Leyva, R. L. Tavares, E. A. Kellogg, *Am. J. Bot.* **107**, 549–561 (2020).
30. S. Sakuma, T. Schnurbusch, *New Phytol.* **225**, 1873–1882 (2020).
31. E. Vollbrecht, P. S. Springer, L. Goh, E. S. Buckler IV, R. Martienssen, *Nature* **436**, 1119–1126 (2005).
32. R. Koppolu, S. Chen, T. Schnurbusch, *Curr. Opin. Plant Biol.* **65**, 102168 (2022).
33. K. A. Levin, S. A. Boden, *J. Exp. Bot.* **71**, 6869–6871 (2020).
34. D. Du *et al.*, *New Phytol.* **231**, 814–833 (2021).
35. Z. Dong, M. Alexander, G. Chuck, *Trends Genet.* **35**, 118–128 (2019).
36. Y. Huang *et al.*, *Plant J.* **96**, 716–733 (2018).
37. J. Li *et al.*, *PeerJ* **10**, e12880 (2022).
38. H. Klein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115871119 (2022).
39. L. E. Dixon *et al.*, *Plant Cell* **30**, 563–581 (2018).
40. S. A. Boden *et al.*, *Nat. Plants* **1**, 14016 (2015).
41. S. Li, S. Meng, J. Weng, Q. Wu, *Trends Plant Sci.* **27**, 355–363 (2022).
42. C. Zhu *et al.*, *Front. Plant Sci.* **12**, 636749 (2021).
43. L. Liu *et al.*, *Nat. Plants* **7**, 287–294 (2021).
44. W. Tanaka, H. Y. Hirano, *New Phytol.* **225**, 974–984 (2020).
45. W. Chen *et al.*, *Science* **375**, eabg7985 (2022).
46. H. Wang *et al.*, *Plant Physiol.* **188**, 460–476 (2022).
47. H. M. Youssef *et al.*, *Nat. Genet.* **49**, 157–161 (2017).
48. E. A. Kellogg *et al.*, *Front. Plant Sci.* **4**, 250 (2013).
49. H. Yao *et al.*, *Mol. Plant* **12**, 374–389 (2019).
50. Z. Chen *et al.*, *Nat. Commun.* **12**, 2378 (2021).
51. M. R. Woodhouse, M. B. Hufford, *Philos. Trans. R. Soc. London Ser. B* **374**, 20180245 (2019).
52. H. Zhang *et al.*, *Sci. Rep.* **8**, 3951 (2018).
53. X. Xu *et al.*, *Dev. Cell* **56**, 557–568.e6 (2021).

ACKNOWLEDGMENTS

We apologize to the authors of original references whom we could not cite because of space limitations. **Funding:** Ideas for this review were developed under collaborative NSF grants IOS-1256373 and IOS-1256393 to P.M. and E.A.K., respectively, and IOS-1938086 and IOS-1557633 to E.A.K. **Competing interests:** P.M. and E.A.K. have no competing interests. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 18 April 2022; accepted 28 June 2022
10.1126/science.abo5035

REVIEW

Grassland soil carbon sequestration: Current understanding, challenges, and solutions

Yongfei Bai[1,2]* and M. Francesca Cotrufo[3]

Grasslands store approximately one third of the global terrestrial carbon stocks and can act as an important soil carbon sink. Recent studies show that plant diversity increases soil organic carbon (SOC) storage by elevating carbon inputs to belowground biomass and promoting microbial necromass contribution to SOC storage. Climate change affects grassland SOC storage by modifying the processes of plant carbon inputs and microbial catabolism and anabolism. Improved grazing management and biodiversity restoration can provide low-cost and/or high-carbon-gain options for natural climate solutions in global grasslands. The achievable SOC sequestration potential in global grasslands is 2.3 to 7.3 billion tons of carbon dioxide equivalents per year (CO_2e year^{-1}) for biodiversity restoration, 148 to 699 megatons of CO_2e year^{-1} for improved grazing management, and 147 megatons of CO_2e year^{-1} for sown legumes in pasturelands.

Grassland ecosystems cover an area of 52.5 million km^2, accounting for ~40.5% of the Earth's land surface excluding Greenland and Antarctica (*1*). Grasslands provide habitats for biodiversity, contribute to food production, and deliver many cultural services (*1*). They also store ~34% of the terrestrial carbon stock (*1*), with ~90% of their carbon stored belowground as root biomass and soil organic carbon (SOC), thus playing a vital role in soil carbon sequestration (*1*, *2*). However, grasslands are highly vulnerable to human disturbance (e.g., overgrazing and land-use conversion to agriculture) and climate change (*1*–*3*). Worldwide, grasslands have undergone severe decreases in biodiversity and ecosystem functions, leading to reductions in SOC storage (*2*, *4*, *5*). Here, we review the recent advances in our understanding of SOC dynamics, current challenges, and possible solutions to enhance SOC sequestration in global grassland ecosystems. We address three questions: (i) How do key biotic and abiotic factors regulate grassland SOC formation, turnover, and stability?; (ii) how do climate warming, alterations in precipitation, and fire affect SOC storage?; and (iii) how does grazing management affect SOC and how can improved practices result in SOC sequestration?

Mechanisms and drivers of SOC sequestration

In grassland ecosystems, ~60% of net primary productivity is allocated belowground (*6*). Belowground carbon inputs are more often incorporated into SOC than aboveground inputs because of their chemical composition (e.g., aliphatic compounds and root exudates) and their presence in the soil (Fig. 1) (*6*). On average, root carbon inputs have a SOC stabilization efficiency that is five times greater than aboveground carbon inputs (*6*).

Organic carbon in soil is distributed between particulate organic matter (POM) and mineral-associated organic matter (MAOM) fractions, with only a minor portion (1 to 2%) present as dissolved organic matter. POM and MAOM differ in their formation, physical and chemical properties, and mean residence times in soil (*7*, *8*). POM is formed from the fragmentation of plant and microbial residues, and therefore is composed of lightweight fragments made of large polymers (Fig. 1). MAOM, by contrast, is formed from single small molecules that are leached from plant residues or exuded from plant roots, which associate to minerals directly (ex vivo) or after microbial assimilation (in vivo) as microbial necromass (*7*, *8*). MAOM on average has a lower carbon:nitrogen ratio because of its proportionally higher microbial origin, its longer mean residence time in soils (from decades to centuries) compared with POM (<10 years to decades), and its strong chemical bonding to minerals and physical protection in fine aggregates (*7*, *8*). Therefore, MAOM contributes to longer-term carbon sequestration in soil. Root exudates such as dissolved sugars, amino acids, and organic acids are the key pathway to MAOM formation largely through microbial in vivo transformations (Fig. 1) (*8*, *9*). Plant aboveground, root, and rhizodeposition inputs exhibit different

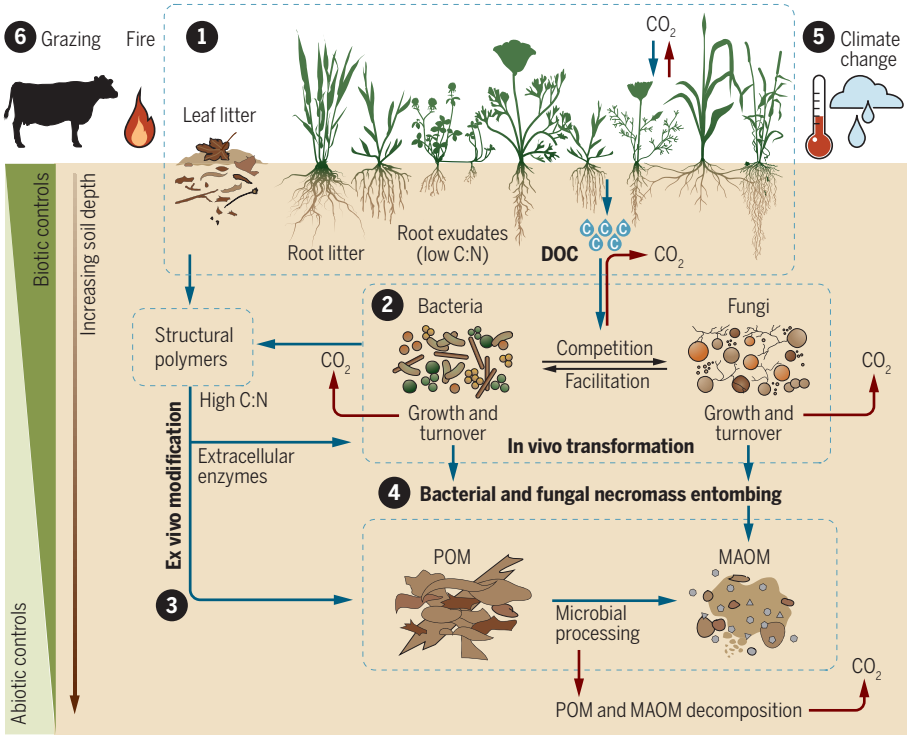
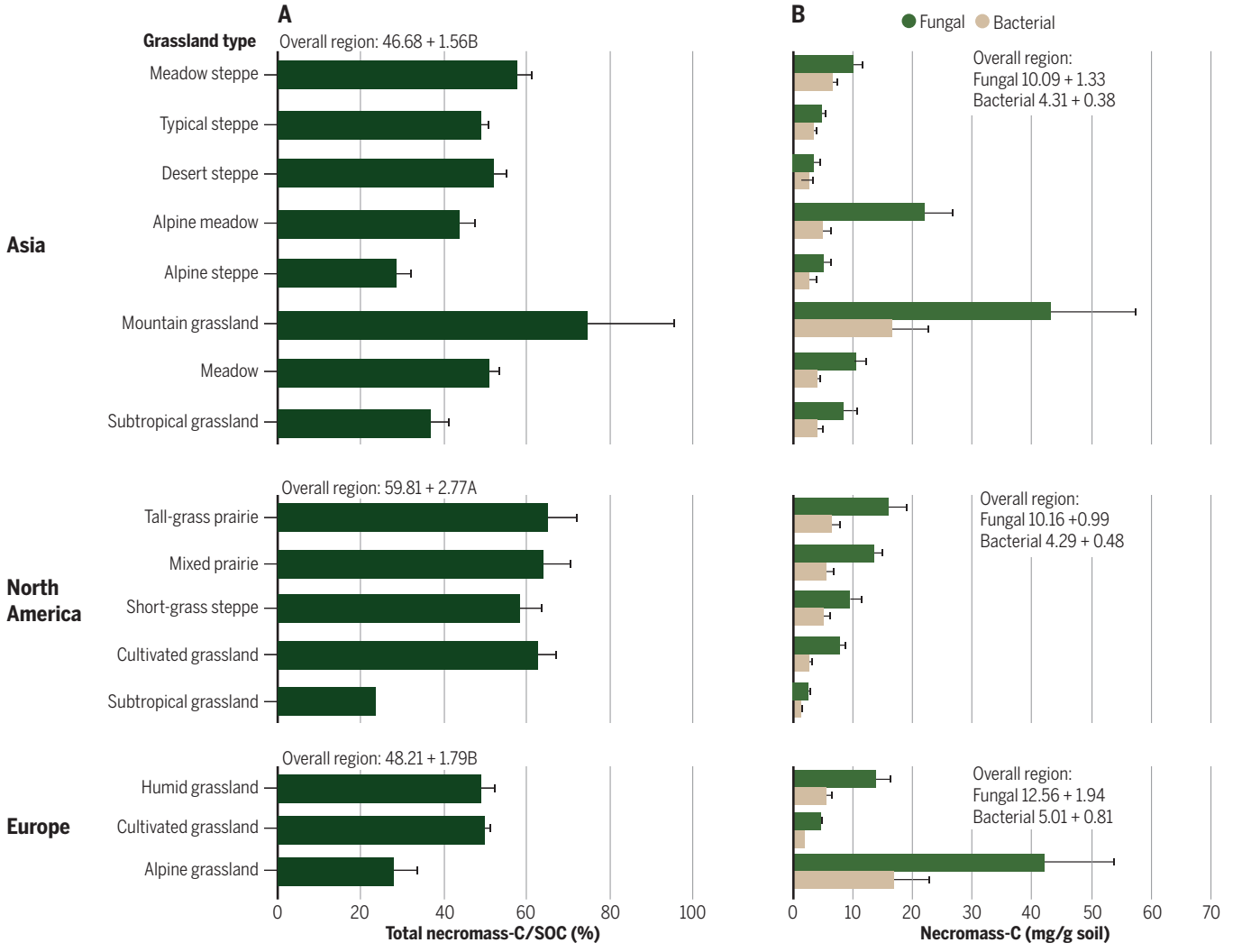


Fig. 1. Conceptual framework for key factors and mechanisms controlling SOC sequestration in grassland ecosystems. (**1**) Plant diversity controls on productivity, biomass allocation, and SOC inputs through litter and root exudates (*6*, *13*, *14*). (**2**) Key pathway of MAOM formation through microbial in vivo transformation (*8*, *17*). (**3**) Pathway of POM formation through microbial ex vivo modification (*8*, *17*). (**4**) Microbial necromass carbon (C) accumulation in MAOM (*9*, *11*). (**5**) Climate change impacts on SOC sequestration through plant and microbial pathways (*26*, *28*). (**6**) Grazing and fire impacts on SOC storage through pathways of plant and animal waste C inputs, compaction, and bioturbation (e.g., trampling and wallowing), microbial in vivo transformation, and microbial ex vivo modification (*33*, *36*, *38*, *46*). C:N, carbon: nitrogen ratio; DOC, dissolved organic carbon.

[1]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. [2]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China. [3]Department of Soil and Crop Science and Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO, USA.
*Corresponding author. Email: yfbai@ibcas.ac.cn



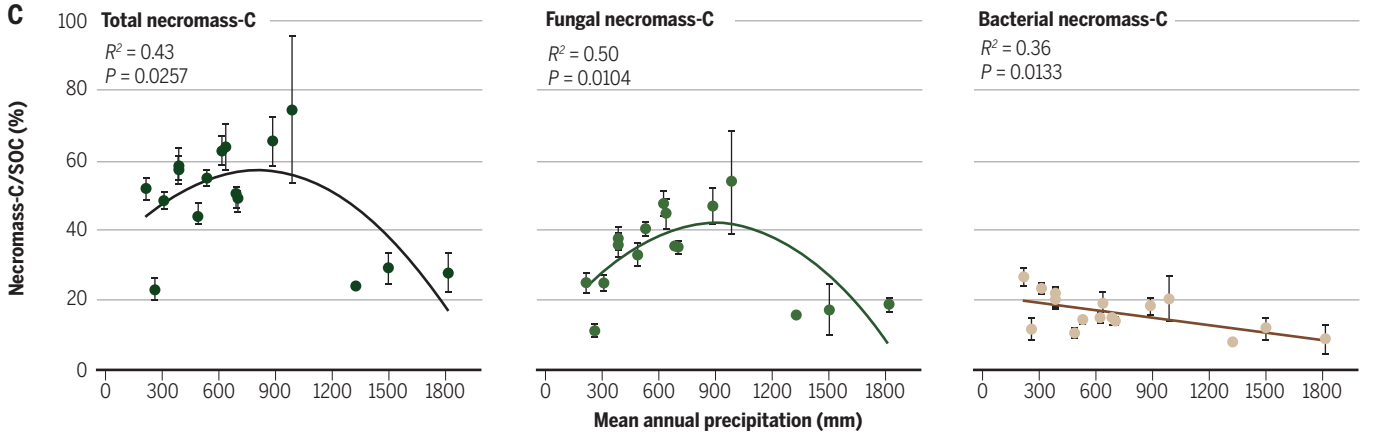


Fig. 2. Patterns and climatic drivers of microbial necromass contribution to SOC. (**A**) Microbial necromass C contribution to SOC. (**B**) Fungal and bacterial necromass C concentrations. (**C**) Relationships of total microbial, fungal, and bacterial necromass C contributions to SOC with mean annual precipitation in the topsoil of grassland systems in Asia, North America, and Europe. Data are from Liang *et al*. (*17*) and Wang *et al*. (*18*). Only the topsoil microbial necromass C and corresponding SOC data (n = 223) were used for global and regional synthesis. All data were classified into different grassland types within regions on the basis of sampling site information from the original study, Asia (eight grassland types, n = 122), North America (five grassland types, n = 47), and Europe (three grassland types, n = 54). Within each grassland type, mean and standard error for each variable were calculated across different sampling sites. General linear model analyses were performed to explore whether the total microbial necromass C contribution to SOC and fungal and bacterial necromass C concentrations differ among different regions. Values with different letters are significantly different at the $P < 0.05$ level. Simple linear regression was used to analyze to the relationship of mean annual precipitation with fungal, bacterial, and total microbial necromass C contributions to SOC across all grassland types on the global scale.

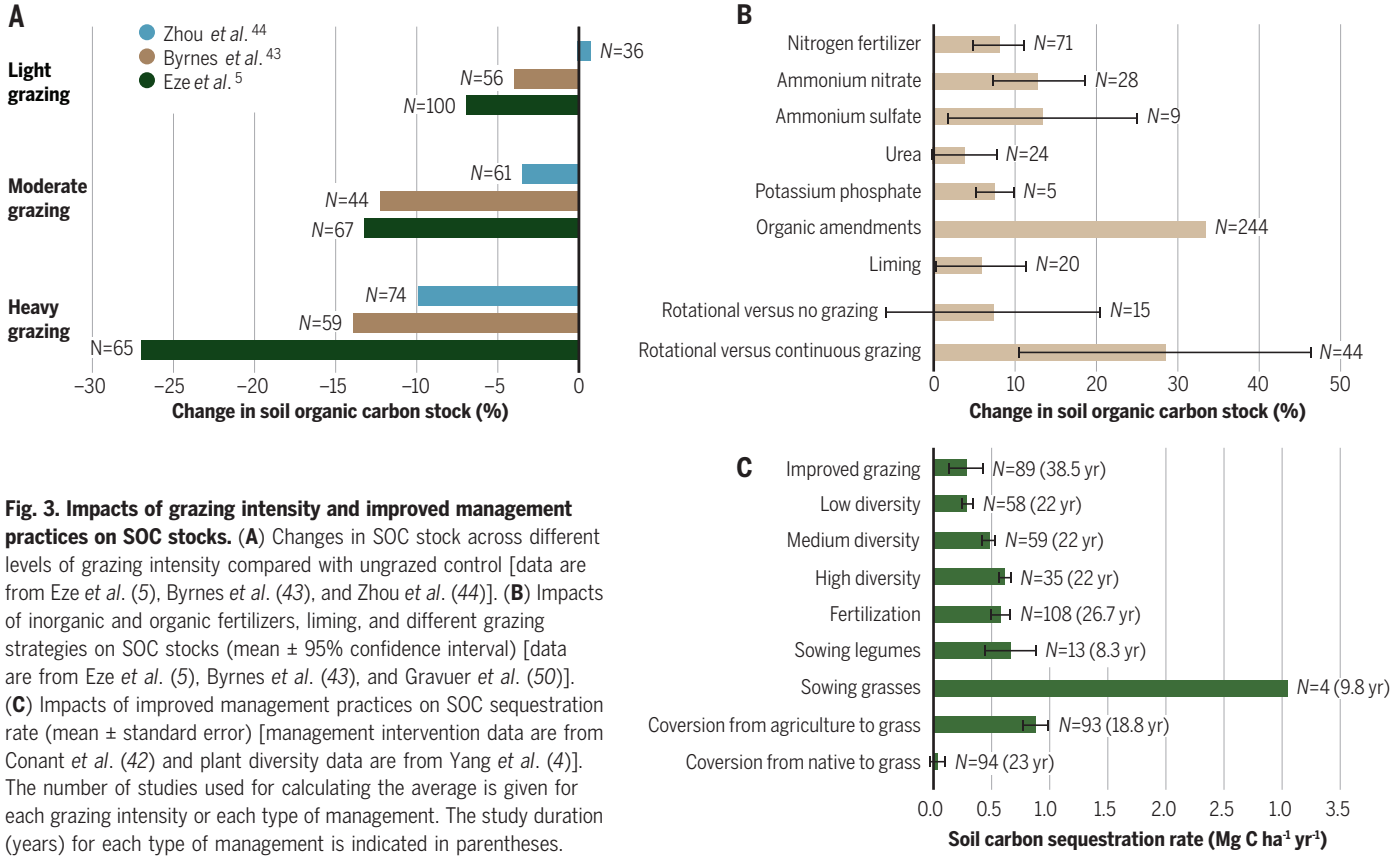


Fig. 3. Impacts of grazing intensity and improved management practices on SOC stocks. (**A**) Changes in SOC stock across different levels of grazing intensity compared with ungrazed control [data are from Eze *et al*. (*5*), Byrnes *et al*. (*43*), and Zhou *et al*. (*44*)]. (**B**) Impacts of inorganic and organic fertilizers, liming, and different grazing strategies on SOC stocks (mean ± 95% confidence interval) [data are from Eze *et al*. (*5*), Byrnes *et al*. (*43*), and Gravuer *et al*. (*50*)]. (**C**) Impacts of improved management practices on SOC sequestration rate (mean ± standard error) [management intervention data are from Conant *et al*. (*42*) and plant diversity data are from Yang *et al*. (*4*)]. The number of studies used for calculating the average is given for each grazing intensity or each type of management. The study duration (years) for each type of management is indicated in parentheses.

POM and MAOM formation efficiencies. Approximately 46% of root exudates, 9% of root tissues, and 7% of aboveground carbon residues are transformed into MAOM, whereas 19% of root litter is transformed into POM across crops, grasses, and trees growing in the field and under controlled laboratory conditions (*10*). Thus, plants with greater carbon allocation to roots contribute more to soil carbon sequestration, particularly the formation of MAOM. However, it remains largely unclear how the contributions of roots (root exudates and root litter) and aboveground inputs to SOC accumulation (POM and MAOM) change with grassland types, soil properties, and climate conditions.

In grassland topsoils, 50 to 75% of SOC is found in MAOM. The average carbon:nitrogen varies from ~10 to ~12 for MAOM and from ~16 to ~18 for POM (*3*); therefore, the accrual of SOC in MAOM requires substantially greater nitrogen than the equivalent accrual in POM (*11*). The formation of POM is primarily driven by climate (temperature and precipitation). By contrast, the accumulation of MAOM is controlled by soil properties such as silt and clay content, cation-exchange capacity, and microbial nitrogen availability, which means that it may saturate (*8*, *12*). In European grasslands, topsoil carbon storage in MAOM saturates at ~50 g C kg^{-1} soil, beyond which the additional increase in SOC storage completely depends upon accrual in POM (*11*). Currently, most European grasslands (80%) are below saturation, indicating a large capacity for SOC sequestration in their topsoils (*11*).

Plant diversity is a key driver of SOC formation and storage (*4*). High plant diversity enhances SOC storage by elevating belowground carbon (i.e., root biomass and root exudates) inputs (*13*, *14*) and promoting microbial growth, turnover, and entombment of necromass (*15*). Maintaining consistently high levels of biodiversity and root carbon inputs is essential for enhancing SOC storage and persistence in grasslands (Fig. 1).

Fungi and bacteria have a strong influence on SOC accumulation, stabilization, and turnover in grasslands (Fig. 1), as in other terrestrial ecosystems (*6*, *16*). Microbial necromass plays an important role in SOC accumulation and stabilization (*9*, *17*). In the topsoil of global grasslands, the contribution of the microbial necromass to total SOC ranges from 23 to 74%, with an average of 50% (Fig. 2A), which is greater than its contribution in agricultural and temperate forest soils (*17*, *18*). The contribution of necromass to SOC changes with soil depth (*18*) and is typically dominated by fungal necromass, with the fungi-to-bacteria necromass carbon ratio ranging from 1.2 to 4.1 across global grasslands (Fig. 2B). This is likely because fungi produce more chemically recalcitrant structural compounds and have greater carbon use efficiency than bacteria (*6*, *16*). Moreover, mycorrhizal fungi, which live in association with plant roots and derive their carbon directly from the plant, can regulate the carbon sequestration capacity in soil. Carbon sequestration capacity per unit nitrogen in soil is 1.7 times greater in ecosystems dominated by ectomycorrhizal fungi–associated plants (e.g., savannas, shrublands, and forests) than in systems dominated by arbuscular mycorrhizal fungi–associated plants (e.g., nonwoody grasslands) because ectomycorrhizal fungi can produce enzymes to degrade organic nitrogen from plant litter (*19*). However, MAOM is relatively higher in ecosystems that are dominated by arbuscular mycorrhizal fungi (*13*), such as grasslands.

Climate regulates the metabolic activity of microbes and thus controls large-scale patterns of microbial necromass and SOC storage (*18*, *20*). At the global scale, cold, moist soils promote the accumulation of microbial necromass carbon. The maximum microbial necromass carbon occurs at a mean annual precipitation of 900 to 1000 mm with a mean annual temperature <0°C (Fig. 2C), indicating high priorities for preserving the current stocks in these systems. Few studies have measured the contribution of microbial necromass carbon to SOC in grassland soils, and data are lacking from

Africa, South America, and Australia (*17*, *18*, *20*). Microbial diversity may also affect SOC storage by regulating the efficiency of microbial assimilation of carbon and the production of organo-mineral associations in soils (*21*). Recently, microbial diversity was found to promote the stabilization efficiency of grass litter–derived POM but to reduce that of MAOM (*22*).

Climate change impacts on SOC sequestration

Sixty-seven percent of the world's grasslands are distributed in semiarid, arid, and cold climates, with only 23% occurring in humid climates (*1*). Thus, carbon sequestration in most grasslands is highly sensitive to climate change, which can exert strong and diverse impacts on SOC accrual and stability through plant- and microbial-mediated mechanisms (*8*). The impacts of climate change on soil carbon sequestration often vary with grassland type, climate, and soil conditions. In semiarid steppe, warming may enhance root-derived carbon input but inhibit the decomposition of MAOM by suppressing fungal growth and soil respiration, resulting in an increase in the MAOM pool (*23*). In humid tallgrass prairies, warming may increase C_4 grass cover and C_4-derived carbon input into soil organic matter, but it also increases the decay rate of these fractions, resulting in a negligible change in soil carbon sequestration (*24*). In alpine grasslands, warming-induced permafrost degradation reduces active-layer SOC storage by decreasing the stability of microbial networks and accelerating SOC (and specifically POM) decay (*25*). A recent meta-analysis demonstrated that long-term (≥5 years) warming increases the ratios of ligninase to cellulase activity and enhances microbial utilization of recalcitrant carbon, leading to a 14% reduction in the topsoil recalcitrant carbon pool (*26*). However, warming may increase the accumulation of root-derived carbon in the subsoil MAOM pool (*27*). POM is much more climate sensitive than MAOM (*3*, *11*). The percent change in POM (–12.2%) with climate warming is on average three times greater than that in MAOM (–3.8%) in global grasslands (*28*). This suggests that grasslands with a high proportion of MAOM will contribute less to soil carbon-climate feedbacks.

Future projected precipitation anomalies and long-lasting droughts (*29*, *30*) will likely influence soil carbon sequestration of grassland ecosystems by altering plant community composition, productivity and carbon allocation, and microbial processes. In the semiarid steppe, increased precipitation promotes soil aggregation by stimulating fungal growth and increasing soil-exchangeable magnesium (*23*). Precipitation anomalies (increases and decreases) can substantially alter root-to-shoot ratios and vertical root distribution in grasslands (*31*), thus regulating soil microbial growth and SOC storage. Reduced precipitation strongly suppresses oxidase activity, whereas higher precipitation stimulates the activity of nitrogen-acquisition extracellular enzymes (*32*). However, on the global scale, only a negative tendency for POM and a positive tendency for MAOM and total SOC concentrations with increased precipitation were observed in grasslands because of the limited data availability (*28*).

Climate change–induced increases in fire frequency can substantially modify long-term SOC storage in grasslands, particularly in savanna grasslands, by intensifying nutrient limitation, which suppresses plant growth and carbon inputs. Elevated fire frequencies reduce soil carbon stocks on average by 0.21 megagrams of carbon per hectare per year (Mg C ha^{-1} $year^{-1}$) in the upper soil layer (0 to 20 cm) in global savanna grasslands (*33*). However, a recent study showed that fire suppression (i.e., >60 years of fire exclusion) has little effect on total SOC storage (0 to 60 cm) in tropical savannas because C_4 grass–derived carbon dominates the SOC, particularly in deeper soil layers, where soil carbon is less affected by changes in fire frequencies (*34*). It remains unclear to what extent different fire regimes regulate plant diversity, above- and belowground biomass allocation, microbial-mediated processes, and SOC storage in shallower and deeper soil profiles.

Impacts of grazing pressure on grassland soil carbon

Natural grasslands are grazed by wild ungulates, which can enhance SOC storage because they graze for short periods of time and move across the landscape. This results in maintained plant cover, diversity and productivity, promotion of species with deep roots, microbial processing with the formation of both POM and MAOM, and soil-mixing processing by soil fauna (*35*, *36*). Increases in ecosystem metabolism and plant labile carbon inputs (e.g., root exudates) are expected to increase both the ex vivo and in vivo formation of MAOM (*9*, *10*, *37*). Conversely, increased root inputs and allocation to depth result in higher POM in the subsoil (*6*, *38*). In addition, large herbivores create habitats for many bioturbators (e.g., fossorial mammals and soil macrofauna) to loosen up soil and expose larger aggregates of soil organic matter to organo-mineral interaction by vertical soil mixing (*36*). However, both the direction and magnitude of effects of large wild herbivores on soil carbon storage can vary strongly with soil nutrient availability, across grasslands, and under different levels of herbivore density. For example, a recent short-term study suggested that nutrient availability strongly moderates the impact of herbivore grazing on soil carbon sequestration in herbaceous grasslands (*39*). Large herbivore grazing increases the upper-layer soil carbon storage under elevated nutrient (fertilization) conditions but has no effect on soil carbon storage under ambient nutrient conditions (*39*). Sandhage-Hofmann *et al.* (*40*) report that elevated elephant densities enhance SOC stocks [4.7 tons (t) ha^{-1}] despite losses of woody biomass in moist, semiarid, wood-encroached savannas of south-central Africa. However, a synthesis of 174 experiments showed that large herbivore exclusion generally increases SOC storage across different biomes (grassland, forest, shrubland, tundra, woodland, etc.), suggesting an overall negative impact of large wild herbivores on soil carbon storage (*41*).

Livestock grazing is the most common use of grasslands worldwide. Some grasslands are managed to improve forage quantity and quality, thereby increasing livestock production and/or SOC storage (*1*, *2*, *42*). In livestock-dominated systems, these pathways are strongly controlled by grazing intensity and rest periods. Continuous livestock grazing reduces plant cover, diversity, and productivity, and thus root inputs and plant- and microbial-mediated SOC formation, while stimulating losses through microbial turnover and erosion caused by increased compaction and reduced cover (*1*, *2*, *43*). Eze *et al.* (*5*) demonstrated that livestock grazing on average decreases SOC stock by 15% across five continents, with the greatest reduction (–22.4%) in SOC stock in the tropics and the least reduction (–4.5%) in temperate grasslands. At the global scale, light grazing (e.g., seasonal and rotational grazing) shows the least negative effects or even promotes soil carbon storage, whereas moderate and heavy (continuous) grazing consistently reduces soil carbon stocks (Fig. 3A) (*5*, *43*, *44*). For a given category of grazing intensity, the discrepancy in magnitude of changes in SOC stocks between these studies may partly arise from the lack of quantitative measures of grazing intensity and the difference in data sources (*5*, *43*, *44*). Nevertheless, the magnitude and directions of grazing impacts on soil carbon sequestration are context dependent and vary with climate and soil conditions, vegetation properties, livestock type, herbivore diversity, grazing strategies (e.g., continuous versus rotational grazing), and grazing intensity and duration (*5*, *38*, *43*–*45*). The negative impact of increasing grazing intensity on SOC is lessened with greater water availability (*5*, *44*) but is more severe with warmer temperatures and longer grazing duration in temperate grasslands (*44*). With moderate and heavy grazing, SOC increases in grasslands dominated by C_4 species and decreases in grasslands dominated by C_3 species (*45*). Sheep grazing generally has a greater negative impact on SOC than cattle grazing, and the reduction in SOC with grazing is substantially greater in topsoil than that in subsoil (*44*). A mixed cattle and megaherbivore system was shown to be a sustainable management strategy in African savanna ecosystems with high

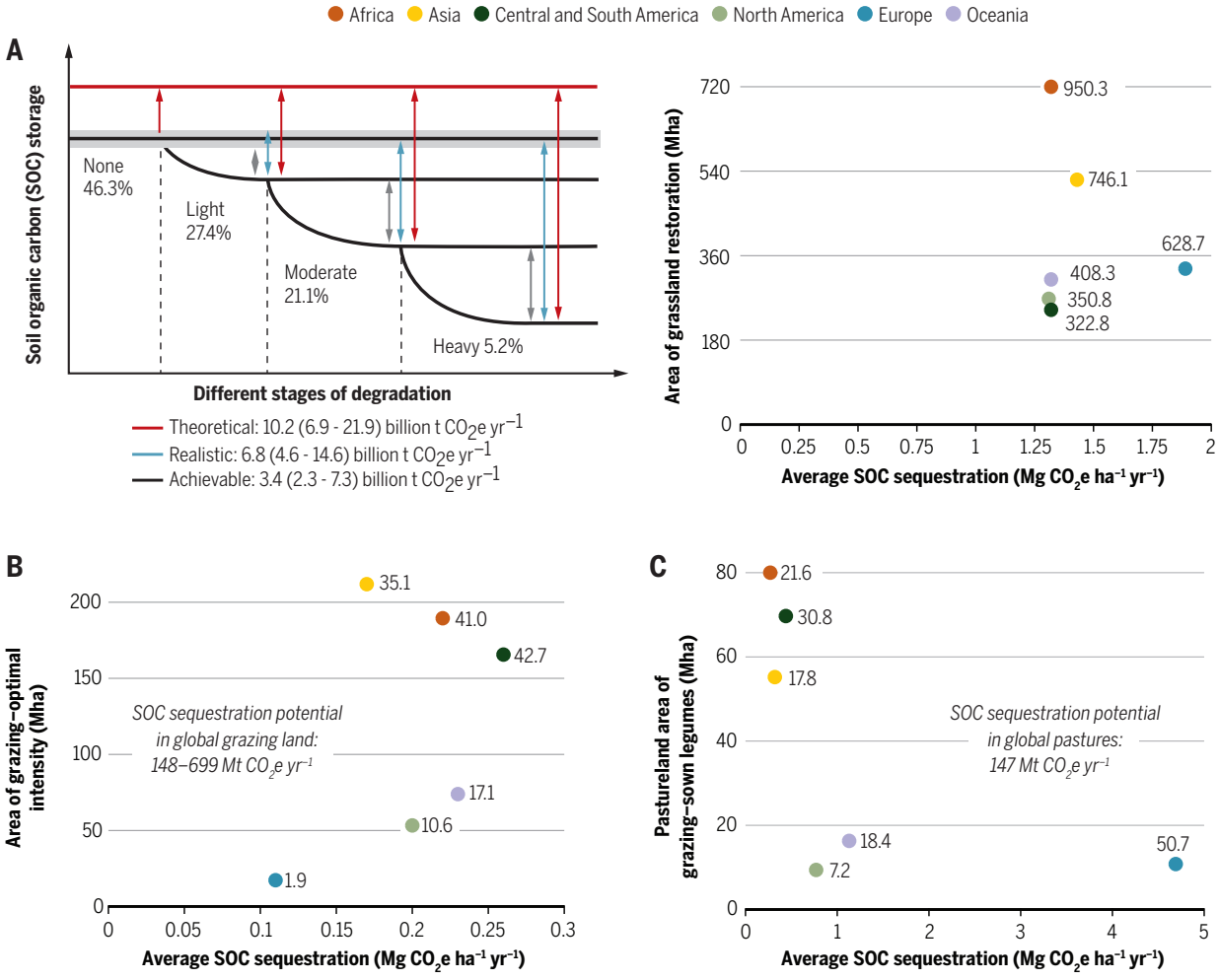


Fig. 4. Soil C sequestration potential of grassland ecosystems at the regional and global scales. SOC sequestration potentials are arranged according to average SOC sequestration rates and area of each management strategy for denoting their relative contributions. (**A**) Capacity and attainability of SOC sequestration by restoring degraded grasslands [data are from White *et al*. (*1*), De Deyn *et al*. (*52*), Deng *et al*. (*54*), and Fargione *et al*. (*53*)]. At the global scale, SOC sequestration potential is presented as theoretical, realistic, and achievable, respectively, based on Chapman (*55*). Means and the 95% confidence intervals (parentheses) are given (left panel). At the regional scale, only achievable SOC sequestration potential are estimated because of the large uncertainties for estimating the theoretical and realistic SOC sequestration potentials in each region. For each region, the mean achievable SOC sequestration potential (Mt CO_2e year^{-1}) is given (right panel). (**B** and **C**) Global SOC sequestration potential (Mt CO_2e year^{-1}) through optimizing grazing intensity in grazing lands and sowing legumes in pasturelands [data are from Griscom *et al*. (*51*)]. Only maximum climate mitigation potential with safeguards for reference year 2030 is shown.

herbivore diversity (*46*). Moreover, rotational grazing consistently shows higher SOC stocks compared with continuous grazing (or free grazing) (*43*), with gains observed specifically in the mineral associated fraction (*47*).

Managing for soil carbon storage in grasslands

Empirical and experimental studies have indicated that improving grassland management can increase SOC storage, thus mitigating carbon losses caused by climate change, long-term overgrazing, and grassland degradation (*2*, *42*, *48*). Management improvements may result in soil carbon accrual through several interrelated mechanisms (Fig. 1). Conversion from croplands to grasslands removes disturbance from tillage and increases root carbon inputs to soil (*6*, *42*). Restoring the biodiversity of degraded grasslands may increase plant production and promote microbial turnover and necromass entombment (*4*, *13*, *15*). Grazing improvement can increase higher-quality root carbon (lower carbon:nitrogen ratios) inputs (*38*) and/or nitrogen retention, thus promoting the formation and persistence of MAOM in soils (*47*). Sowing legumes increases soil carbon and nitrogen inputs by elevating root biomass, root exudates, and fine root turnover (*42*, *49*). Applications of inorganic and organic fertilizers may stimulate primary productivity and high-quality plant carbon inputs to soil, resulting in more efficient microbial carbon use (*5*, *28*, *50*).

A number of management interventions have been adopted to restore grasslands (Fig. 3, B and C). On the global scale, the improved grassland managements increase SOC stocks on average by 0.47 Mg C ha^{-1} year^{-1} (*42*). This suggests that the world's grazing lands, which occupy an area of ~34 million km^2, have substantial potential to increase SOC storage (Fig. 4). Among all improved management practices, conversion from cultivation to grasslands, increasing plant diversity, sowing legumes and grasses, and fertilization are associated with the highest soil carbon sequestration rates (Fig. 3C) (*4*, *42*). Under moderate grazing intensity, the average SOC stock increase (28.4%) is substantially greater with rotational grazing than that with continuous grazing (Fig. 3B). In the southeast United States, grassland soils managed with adaptive multi-paddock grazing that used a high-density-short-duration rotational grazing had more carbon (72.49 Mg C ha^{-1}) and nitrogen (9.26 Mg N ha^{-1}) stocks compared with continuous grazing (64.02 Mg C ha^{-1} and 8.52 Mg N ha^{-1}) in the 0 to 100 cm soil layer (*47*). However, the

direction and magnitude of management effects on soil carbon stocks are context specific, depending on factors such as climate, plant community composition, and soil properties (*5*, *43*, *50*). Therefore, grazing practices need to be implemented with an understanding of context. Moreover, further studies are required to examine the synergy and trade-offs among grassland biodiversity, primary productivity, and soil carbon sequestration under management interventions.

Soil carbon sequestration potential varies in both quantity and attainability among grasslands with different degrees of degradation and across different regions (Fig. 4). Given that ~50% of the global grassland area has been degraded (*1*, *2*), restoration of grassland cover and biodiversity is an effective strategy for promoting SOC storage and mitigating the negative impacts of global climate change (*4*, *15*, *51*–*53*). For example, the SOC accrual rate with grazing exclusion is on average 0.68 Mg C ha^{-1} $year^{-1}$ in topsoil (0 to 30 cm) and 0.62 Mg C ha^{-1} $year^{-1}$ in subsoil (30 to 100 cm) across 145 degraded grassland sites in China (*54*), indicating that it has not reached saturation over the 27-year period of grassland restoration.

Potential soil carbon sequestration capacities can be categorized as theoretical, realistic, or achievable (*55*). Theoretical soil carbon sequestration capacity refers to the estimate of restoring all soils to their natural capacity or even enhancing it through management interventions, realistic soil carbon sequestration capacity refers to the optimistic value accounting for social and economic constraints, and achievable capacity is the value of a pragmatic scenario based on the current trends (*55*). At the global scale, the mean theoretical, realistic, and achievable capacities of SOC sequestration with grassland restoration are estimated to be 10.2, 6.8, and 3.4 billion t CO_2 equivalents per year (CO_2e $year^{-1}$), respectively (Fig. 4A). At the regional scale, Africa, Asia, and Europe are projected to have the largest achievable capacity of soil carbon sequestration with grassland restoration, with Oceania and North and South America exhibiting the least SOC sequestration potential (Fig. 4A). These global patterns of SOC sequestration potential are primarily caused by the differences in average soil carbon sequestration rate and the area of degraded grassland in different regions. The greater SOC sequestration potential with grassland restoration in Africa and Asia is due to the larger areas of degraded grasslands in these continents, whereas European grasslands have a higher average soil carbon sequestration rate (Fig. 4A). In addition, optimizing grazing intensity (e.g., rotational grazing) is projected to increase soil carbon sequestration potential by 148 to 699 megatons (Mt) CO_2e $year^{-1}$ in global grazing lands (Fig. 4B), with the greatest SOC sequestration potential occurring in Central and South America, Africa, and Asia (*51*). Moreover, sowing legumes is projected to enhance SOC storage by 147 Mt CO_2e $year^{-1}$ in global pasturelands (*51*), with Europe exhibiting the greatest soil carbon sequestration potential caused by both the largest pastureland areas and the highest average soil carbon sequestration rate (Fig. 4C). At both the regional and global scales, large uncertainties exist regarding the projected soil carbon sequestration potential and rate of accrual. These uncertainties are caused by the complex interactions among climate change, human activities, and spatial and temporal variations in ecosystem and soil responses (*51*, *53*, *56*). Scientific research and management innovations are required in the future to maximize the attainable SOC storage in global grasslands.

Conclusion

Recent studies have made considerable progress toward addressing major challenges associated with identifying the capacity and key mechanisms of various grasslands to sequester and preserve carbon in soils and developing knowledge-based strategies to restore biodiversity, preserve current SOC stocks, and promote additional sequestration for climate change mitigation and sustainable management in grasslands. These advances highlight the important roles of plant and soil biodiversity in regulating the formation of microbial necromass carbon, MAOM, and POM, mediating the impacts of climate change, and promoting SOC storage through management improvements and restoration in global grasslands. They also demonstrate that the impacts of climate change, grazing, fire, grassland restoration, and mitigation solutions on soil carbon sequestration are moderated by multiple context-dependent factors. Future research is needed to address the uncertainty and context dependency of the proposed mitigation solutions and their carbon sequestration potentials and to consider their possible synergies and trade-offs for biodiversity conservation, climate mitigation, and food production.

REFERENCES AND NOTES

1. R. P. White, S. Murray, M. Rohweder, *Pilot Analysis of Global Ecosystems: Grassland Ecosystems* (World Resources Institute, 2000).
2. R. D. Bardgett *et al*., *Nat. Rev. Earth Environ.* **2**, 720–735 (2021).
3. E. Lugato, J. M. Lavallee, M. L. Haddix, P. Panagos, M. F. Cotrufo, *Nat. Geosci.* **14**, 295–300 (2021).
4. Y. Yang, D. Tilman, G. Furey, C. Lehman, *Nat. Commun.* **10**, 718 (2019).
5. S. Eze, S. M. Palmer, P. J. Chapman, *J. Environ. Manage.* **223**, 74–84 (2018).
6. R. B. Jackson *et al*., *Annu. Rev. Ecol. Evol. Syst.* **48**, 419–445 (2017).
7. J. M. Lavallee, J. L. Soong, M. F. Cotrufo, *Glob. Chang. Biol.* **26**, 261–273 (2020).
8. M. F. Cotrufo, J. M. Lavallee, *Adv. Agron.* **172**, 1–66 (2022).
9. C. Liang, J. P. Schimel, J. D. Jastrow, *Nat. Microbiol.* **2**, 17105 (2017).
10. S. H. Villarino, P. Pinto, R. B. Jackson, G. Piñeiro, *Sci. Adv.* **7**, eabd3176 (2021).
11. M. F. Cotrufo, M. G. Ranalli, M. L. Haddix, J. Six, E. Lugato, *Nat. Geosci.* **12**, 989–994 (2019).
12. E. Mitchell *et al*., *Glob. Chang. Biol.* **27**, 5383–5391 (2021).
13. S. Chen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4027–4032 (2018).
14. M. Lange *et al*., *Nat. Commun.* **6**, 6707 (2015).
15. J. Prommer *et al*., *Glob. Chang. Biol.* **26**, 669–681 (2020).
16. S. D. Frey, *Annu. Rev. Ecol. Evol. Syst.* **50**, 237–259 (2019).
17. C. Liang, W. Amelung, J. Lehmann, M. Kästner, *Glob. Chang. Biol.* **25**, 3578–3590 (2019).
18. B. Wang, S. An, C. Liang, Y. Liu, Y. Kuzyakov, *Soil Biol. Biochem.* **162**, 108422 (2021).
19. C. Averill, B. L. Turner, A. C. Finzi, *Nature* **505**, 543–545 (2014).
20. T. Ma *et al*., *Nat. Commun.* **9**, 3480 (2018).
21. J. G. Ernakovich *et al*., *Biogeochemistry* **153**, 1–15 (2021).
22. M. F. Cotrufo, M. L. Haddix, M. E. Kroeger, C. E. Stewart, *Soil Biol. Biochem.* **168**, 108648 (2022).
23. T. Bai *et al*., *Glob. Chang. Biol.* **26**, 5320–5332 (2020).
24. X. Cheng, Y. Luo, X. Xu, R. Sherry, Q. Zhang, *Biogeosciences* **8**, 1487–1498 (2011).
25. M.-H. Wu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025321118 (2021).
26. J. Chen *et al*., *Glob. Chang. Biol.* **26**, 1944–1952 (2020).
27. J. Jia *et al*., *Glob. Chang. Biol.* **25**, 4383–4393 (2019).
28. K. S. Rocci, J. M. Lavallee, C. E. Stewart, M. F. Cotrufo, *Sci. Total Environ.* **793**, 148569 (2021).
29. A. J. Felton, A. K. Knapp, M. D. Smith, *Glob. Chang. Biol.* **27**, 1127–1140 (2021).
30. V. Humphrey *et al*., *Nature* **592**, 65–69 (2021).
31. B. Zhang *et al*., *Ecology* **100**, e02828 (2019).
32. W. Xiao, X. Chen, X. Jing, B. Zhu, *Soil Biol. Biochem.* **123**, 21–32 (2018).
33. A. F. A. Pellegrini *et al*., *Nature* **553**, 194–198 (2018).
34. Y. Zhou *et al*., *Nature* **603**, 445–449 (2022).
35. O. J. Schmitz *et al*., *Science* **362**, eaar3213 (2018).
36. J. A. Kristensen, J.-C. Svenning, K. Georgiou, Y. Malhi, *Trends Ecol. Evol.* **37**, 117–128 (2022).
37. N. W. Sokol, M. A. Bradford, *Nat. Geosci.* **12**, 46–53 (2019).
38. C. H. Wilson, M. S. Strickland, J. A. Hutchings, T. S. Bianchi, S. L. Flory, *Glob. Chang. Biol.* **24**, 2997–3009 (2018).
39. J. Sitters *et al*., *Glob. Chang. Biol.* **26**, 2060–2071 (2020).
40. A. Sandhage-Hofmann, A. Linstädter, L. Kindermann, S. Angombe, W. Amelung, *Glob. Chang. Biol.* **27**, 4601–4614 (2021).
41. E. S. Forbes *et al*., *Funct. Ecol.* **33**, 1597–1610 (2019).
42. R. T. Conant, C. E. P. Cerri, B. B. Osborne, K. Paustian, *Ecol. Appl.* **27**, 662–668 (2017).
43. R. C. Byrnes, D. J. Eastburn, K. W. Tate, L. M. Roche, *J. Environ. Qual.* **47**, 758–765 (2018).
44. G. Zhou *et al*., *Glob. Chang. Biol.* **23**, 1167–1179 (2017).
45. M. E. McSherry, M. E. Ritchie, *Glob. Chang. Biol.* **19**, 1347–1357 (2013).
46. J. Sitters, D. M. Kimuyu, T. P. Young, P. Claeys, H. Olde Venterink, *Nat. Sustain.* **3**, 360–366 (2020).
47. S. Mosier *et al*., *J. Environ. Manage.* **288**, 112409 (2021).
48. Y. F. Bai, Q. M. Pan, Q. Xing, *Chin. Sci. Bull.* **61**, 201–212 (2016).
49. G.-L. Wu, Y. Liu, F.-P. Tian, Z.-H. Shi, *Land Degrad. Dev.* **28**, 1336–1344 (2017).
50. K. Gravuer, S. Gennet, H. L. Throop, *Glob. Chang. Biol.* **25**, 1152–1170 (2019).
51. B. W. Griscom *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11645–11650 (2017).
52. G. B. De Deyn *et al*., *J. Appl. Ecol.* **48**, 600–608 (2011).
53. J. E. Fargione *et al*., *Sci. Adv.* **4**, eaat1869 (2018).
54. L. Deng, Z.-P. Shangguan, G.-L. Wu, X.-F. Chang, *Earth Sci. Rev.* **173**, 84–95 (2017).
55. S. J. Chapman, in *Carbon Capture: Sequestration and Storage*, R. E. Hester, R. M. Harrison, Eds. (RSC Publishing, 2010), pp. 179–202.
56. F. Lu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4039–4044 (2018).

ACKNOWLEDGMENTS

We thank S. P. Chen, J. Q. Tian, Y. J. Zhao, S. X. Zheng, Y. Wang, X. M. Lu, B. Tang, J. Man, and J. S. Su for discussion on the topic of grassland carbon sequestration. **Funding:** This work was supported by the National Natural Science Foundation of China (grant no. 32192464) and the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA23080000). **Author contributions:** Y.F.B. conceived the concept, structure, and writing of the manuscript. M.F.C. contributed to the development of ideas and writing of the manuscript. **Competing interests:** M.F.C. provides consulting services for C quester Analytics. Services are provided to a number of companies invested in soil carbon sequestration, including by the adoption of regenerative grazing management. The remaining authors declare no competing interests. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 13 April 2022; accepted 15 June 2022
10.1126/science.abo2380

REVIEW

The planetary role of seagrass conservation

Richard K. F. Unsworth[1,2]*, Leanne C. Cullen-Unsworth[1,2], Benjamin L. H. Jones[2,3], Richard J. Lilley[2]

Seagrasses are remarkable plants that have adapted to live in a marine environment. They form extensive meadows found globally that bioengineer their local environments and preserve the coastal seascape. With the increasing realization of the planetary emergency that we face, there is growing interest in using seagrasses as a nature-based solution for greenhouse gas mitigation. However, seagrass sensitivity to stressors is acute, and in many places, the risk of loss and degradation persists. If the ecological state of seagrasses remains compromised, then their ability to contribute to nature-based solutions for the climate emergency and biodiversity crisis remains in doubt. We examine the major ecological role that seagrasses play and how rethinking their conservation is critical to understanding their part in fighting our planetary emergency.

Though commonly called grasses, seagrasses are a unique group of submarine flowering plants that belong to the monocotyledon order Alismatales, comprising four families and 72 species. Although they occupy a broad range of niches and are derived from multiple evolutionary lineages (*1*), they all share a connection to marine environments and consistently exhibit features that separate them from all other angiosperms. Seagrasses have adapted to live underwater, where light is limited, where salt and nutrients can be problematic, and where soils can become highly toxic (*2*).

Seagrass diverged from other alismatid monocots ~105 million years ago, and work by Olsen *et al*. (*3*) supports hypotheses that modern seagrass biodiversity can be linked to the materialization of multiple habitats after the Cretaceous-Paleogene extinction event. In the past decade, the seagrass science community has grown (*4*) and revealed the uniqueness of these plants and the importance of the ecosystems that they create (Fig. 1). Seagrasses bioengineer their environment by slowing water flow, trapping particles, and improving the environment within a positive feedback mechanism to facilitate the creation of habitat (*5*). Just like terrestrial plants, their reproduction can be supported by a diverse range of pollinators, such as cumacean crustaceans (*6*), and seed dispersers, such as fish (*7*). Their reproduction is not always sexual—genetic evidence has revealed that vegetative growth has led to the establishment of one single clonal organism spanning >180 km of coastline (*8*). Nitrogen-fixing bacteria living within their roots allow them to colonize nitrogen-poor environments (*9*), and associations with clams (and their bacterial symbionts) have aided their ability to inhabit otherwise toxic sulphide-rich marine soils (*10*). There is also growing evidence of the presence of fungi associated with the roots and rhizomes of seagrasses, indicating that they may play essential roles similar to those of fungal associates of terrestrial plants (*11*).

Aside from their ecological uniqueness, seagrasses are of increasing interest in a socio-political context owing to their potential to help combat the current climate and biodiversity crises that our planet faces. Seagrass meadows also support human well-being by virtue of their role in supporting fisheries, coastal protection, and water filtration (*12*), and action for their conservation supports the fulfilment of the 17 Sustainable Development Goals (SDGs) proposed by the United Nations in 2015. Seagrasses also support many species of conservation concern, such as the dugong, green turtle, and manatee (*13*), and provide interacting ecological functioning throughout the coastal seascape (*14*).

"Compared with...terrestrial grasses and even seaweeds, the body of research within seagrass is magnitudes smaller..."

To harness the power of seagrass as a nature-based solution to the climate emergency and the biodiversity crisis, seagrass systems must be in a resilient functioning state. Seagrass meadows remain globally threatened by diverse factors, including poor water quality, damage from boats and related activities, aquaculture, and coastal development (*15*). Even in areas where seagrass is protected, extreme climate drivers place seagrass at risk. For example, after a marine heatwave in 2010 to 2011, up to 699 km^2 of seagrass meadow in the Shark Bay Marine Park in Western Australia were lost or damaged, potentially releasing up to 9 Tg of CO_2 back into the atmosphere during the 3 years before regrowth occurred (*16*). Seagrass sensitivity to stressors is acute and may even extend to the effects of anthropogenic noise (*17*). In many places, the risk of seagrass loss and degradation persists (*15*), and its functional state is commonly compromised; unless this can be reversed, the potential for seagrass to contribute to the complex jigsaw of nature-based solutions remains in doubt. In this Review, we reflect on the status of seagrass ecosystems, the major ecological role that they play in the coastal environment, and how rethinking their conservation is critical to allowing them to play a role in reversing climate change.

Global decline, net-zero loss, and achieving net gain

The role that seagrass can have in reversing or mitigating climate change requires consideration of their global biogeochemical contribution. For this, we first need a better understanding of whether seagrasses are currently in a state of net loss, stasis, or net gain globally, along with the parameters that drive their greenhouse gas balance (Fig. 1). The global coverage of seagrass is currently estimated to be 160,387 to 266,562 km^2 (*18*). This range reveals that we have very limited understanding of the actual extent of seagrass populations. We also do not fully understand the extent of the ecological goods and services that seagrass provides, including to biodiversity and coastal protection. Studies have sought to place estimates on seagrass loss at 1 to 7% per year (*19*, *20*) and create global carbon storage estimates of up to 19.9 Pg (*21*, *22*). However, if we do not know how much we have or have had, we cannot hypothesize very well on what has been lost or its associated ecological relevance.

The reported trajectory of seagrass coverage (*20*, *23*) indicates that it may be recovering in some areas; however, this analysis is limited because it only focuses on locations where seagrass is mapped, monitored, and likely affected by some level of conservation action, and it may represent only a fraction of potential and unknown seagrass area. Analyses are also limited by favoring data published in academic journals and excluding available data in the gray literature. A coordinated global effort is required to create meaningful global estimates of seagrass coverage and change that are validated with open data sharing between governments, academics, nongovernmental organizations, and commercial enterprises (*18*). In the UK, a technology-focused consortium is forming to fill the gaps in our knowledge to help drive understanding of the ecological role of seagrasses (*24*), and recommendations for a methodological pathway to improve the global seagrass map have recently been proposed (*18*, *25*).

Seagrass as a nature-based solution

The growing interest in nature-based solutions is necessitating deeper understanding of the ecological role that seagrass meadows play in the context of climate change. Seagrass meadows store and sequester carbon within their sediments over long periods of time at highly efficient rates; however, this role varies over space and time along with factors such as hydrodynamics and species composition influencing

[1]Seagrass Ecosystem Research Group, Faculty of Science and Engineering, Swansea University, Singleton Park, Swansea SA2 8PP, Wales, UK. [2]Project Seagrass, The Yard, Bridgend Industrial Estate, Bridgend CF31 3EB, Wales, UK. [3]Department of Ecology, Environment and Plant Sciences, Stockholm University, SE-106 91 Stockholm, Sweden.
*Corresponding author. Email: r.k.f.unsworth@swansea.ac.uk

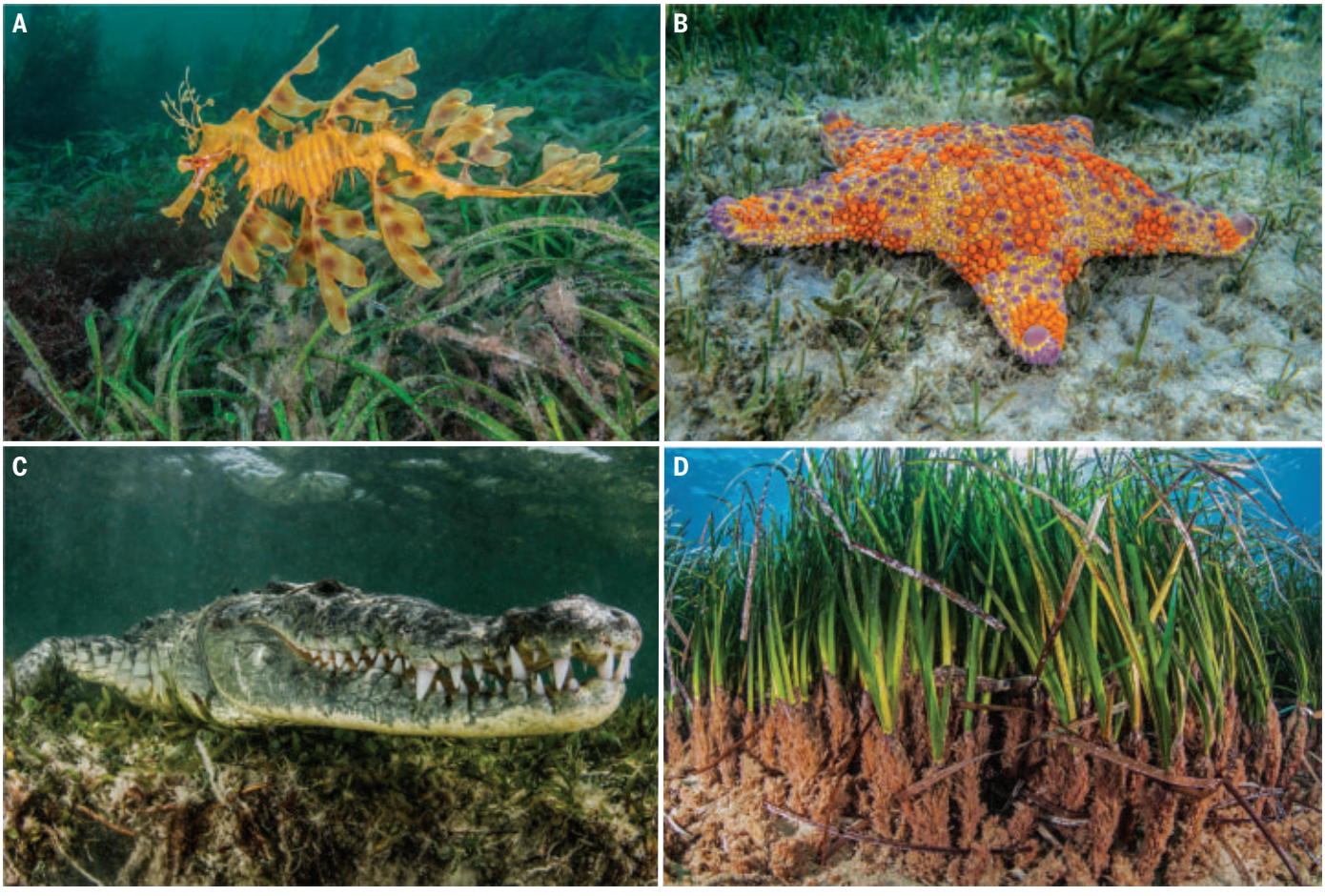


Fig. 1. Seagrass and biodiversity. (**A** to **C**) Seagrass meadows contain biodiverse and enigmatic species assemblages, including the leafy sea dragon (A), sea stars (B), and predators such as crocodiles (C). (**D**) The biodiversity and productivity of seagrass meadows also lead to them storing and sequestering substantial amounts of carbon in their sediments. Seagrass meadows provide habitat in support of biodiversity [(A) to (C)] in coastal waters globally. When healthy and in a balanced state, seagrass can be a great source of many other ecosystem services, such as water filtration, carbon storage (D), and coastal defense. Anthropogenic factors, such as coastal development and poor water quality leading to eutrophication of coastal waters, are some of the principal drivers of seagrass decline.

IMAGE CREDIT: DIMITRIS POURSANIDIS, BENJAMIN JONES, AMANDA COTTON, JAYNE JENKINS, VIA THE OCEAN AGENCY AND RICHARD UNSWORTH

this function. Additionally, despite their more obvious role in the storage of organic carbon, seagrasses, like most vegetation, also produce the greenhouse gases methane (CH_4) and nitrous oxide (N_2O). The balance of these emissions relative to the storage of carbon is of principal importance in the context of their role in influencing climate. Limited understanding exists with respect to whole–seagrass ecosystem greenhouse gas balance (Fig. 2). Available data indicate that seagrasses have broadly lower greenhouse gas emissions of CH_4 and N_2O than comparative coastal and wetland habitats and that low salinity and anthropogenic stressors are major processes driving production (*26*). Similarly, comparison with habitats such as peatlands and mangroves shows seagrasses to be relatively low in CH_4 and N_2O (*27*). However, after seagrass meadow degradation and loss, there exists a potential for high emissions of CH_4 from underlying sediment (*28*). Eutrophication of seagrasses may also drive elevated N_2O emissions. Although scientific understanding in this field is increasing rapidly, our lack of understanding of the drivers of greenhouse gas emissions by plants, least of all by seagrasses (*21*, *27*), contributes to the uncertainties that surround the marketing of blue carbon (*29*).

Although its capacity for carbon storage is of high current interest, human appreciation for the ecological role of seagrasses has changed (*30*). An historic view of seagrasses from the Northern Hemisphere shows their importance in food production and as a raw material. For example, house roofs in Denmark were thatched with dried seagrass (some of which can still be seen), and seagrass detritus was used to fertilize crops (*30*). In the late 1800s, when Indian cotton crops failed, documented discussion by British cotton traders turned to the use of seagrass as an alternative fiber. In North America, companies existed that traded in seagrass as an insulation material, which was subsequently used in the US Capitol building. The Seri people of the Gulf of California collected seagrass seed to create a gruel (*31*). In the 21st century, in many parts of the world, seagrass meadows are a source of food from the gastropod and bivalve mollusks and sea cucumbers that they shelter (*32*). The importance of seagrass habitats as a source of seafood production is both direct and indirect at local and basin-wide scales, with 20% of the world's biggest finfish fisheries having some known association with seagrass (*33*).

Seagrasses also play a fundamental role in the filtration of coastal waters, trapping particles (including microplastics), cycling nutrients, and absorbing nitrogen from the water column (*34*). This filtration role also extends to the removal of bacteria and viruses (*35*–*37*), thus contributing to improved sanitation (*38*) and human health and well-being (*12*). In the Baltic Sea, seagrass meadows have been recorded to contain 63% fewer potentially harmful *Vibrio vulnificus* and *Vibrio cholerae* bacteria compared with nonvegetated areas (*37*).

Additionally, the role of seagrass in protecting coastlines from erosion is substantial and may grow in value with sea level rise and as storms become more frequent (*11*). The locally relevant role of seagrasses in ameliorating low pH from ocean acidification may also increase the value

of these marine plants over time (*39*, *40*). Although the ecological roles that seagrasses play around the world shift with space and time, the constant across most of the world's seagrasses is that they remain at ecological risk and many are in a perilous state.

What is a pristine, healthy, or balanced seagrass ecosystem?

The extent and function of seagrass meadows are largely manifestations of current and previous human activity. We have limited capacity to appreciate the value of seagrass owing to the scale of alteration and unknown baselines for these systems (*41*, *42*). Evidence from ecological feedbacks indicates that seagrass meadows are driven by top-down and bottom-up processes (*43*, *44*). Although there is increasing appreciation for how seagrass might be influenced by excess nutrients and various pollutants in our coastal waters, we have limited appreciation for what extreme overexploitation of near-shore environments has done to seagrass meadows. We simply do not know what a so-called pristine meadow looks like, which creates a limited appreciation for the true ecological role of these poorly understood systems. A contributory factor to the poor understanding is the low relative research output on seagrasses [see (*45*)]. However, it is apparent that there has been a profound loss of predators from these systems, whereas numbers of consumers, secondary consumers, and grazers have also been affected (*46*)—in some cases, loss of predators has led to overgrazing (*47*, *48*).

In localities where associated biodiversity is high, functional redundancy may serve to protect seagrass meadows (*49*), but with decreasing diversity away from the tropics, such redundancy may be reduced. There is also a growing appreciation for seagrass as a foraging resource for seabirds; this is because they support abundant prey items, such as crustaceans, polychaetes, and fish (*50*). Given the parallel global decline of avifauna with global seagrass, we can only speculate as to what the functional role of loss of seagrass might have once been (*51*).

In recent decades, biodiversity and ecosystem functioning has evolved into a dynamic area of contemporary ecology with a rich body of research. Compared with research in terrestrial grasses and even seaweeds, the body of research within seagrass is magnitudes smaller and is fueled by a smaller community of scientists. We must understand the biodiversity associated with seagrass meadows to be able to develop management programs that secure their ecological functioning under further climate change. Global and regional studies are beginning to transform our knowledge (*44*, *52*, *53*), but tools such as sequencing environmental DNA need to be more widely applied. Reconstructions using molecular and historical evidence are needed to understand the true ecological potential of these ecosystems, to locate sites for rehabilitation and replanting, and to provide ambition to marine conservation.

Seagrass meadows and the SDGs

Improved protection and restoration of seagrasses require better recognition of the role that they play in supporting people and our planet; the state of seagrasses is symptomatic of the deteriorating state of the overall natural environment (*54*). The United Nations SDGs are a means of framing a response to this emergency by connecting the daily actions and needs of people, institutions, and communities to the sustainability of the planet and transforming these connections into measurable actions for positive environmental, social, and ecological outcomes. Articulating the ecological role of seagrass in terms of ecosystem services and natural capital promotes a scientific vision of what behavioral change might mean for seagrass, whereas the SDGs provide a framework for how change can be perceived by all people. We suggest that, of the 17 SDGs, action for seagrass conservation and restoration can make a meaningful contribution to 16 of these global goals (Fig. 3). We propose that the ecological role and value of seagrass can also be described in these terms to improve and catalyze action to halt and reverse seagrass loss.

Seagrass meadows form globally relevant habitats that support fisheries and associated economic goods; it is in this ecological role that seagrasses play a prominent role in SDGs. Thus, well-managed, sustainably exploited seagrass meadows that are in a state of ecological balance (*32*, *33*, *55*) will contribute to reducing poverty (*56*), reducing hunger (*32*), responsible consumption and production (*57*), and decent work and economic growth (*58*) (Fig. 3). Sustainably managed seagrass fisheries in many parts of the world also contribute toward gender equality and reducing other inequalities. For example, the role of women is underappreciated in intertidal and near-shore small-scale subsistence fisheries (*59*), of which seagrass meadows are a major component. Inclusion of women in these fisheries is well known to improve community adaptative capacity and resilience (*60*), leading to improved environmental outcomes (*59*).

A major ecological role of healthy seagrass systems is to make the wider environment more conducive for animal life (including humans) in both marine and coastal environments. Seagrass habitats enhance oxygenation in marine sediments; trap particles in the water column, improving water clarity; cycle and store nutrients; and reduce the bacterial and viral load in coastal waters. This creates a three-dimensional environment that harbors biodiversity, baffles wave energy to protect coastlines from erosion, and further enhances the whole coastal seascape for biodiversity (e.g., through the protection of adjacent habitats, such as coral reefs and mangroves).

The bioengineering effect that seagrasses have on their own environment also contributes to

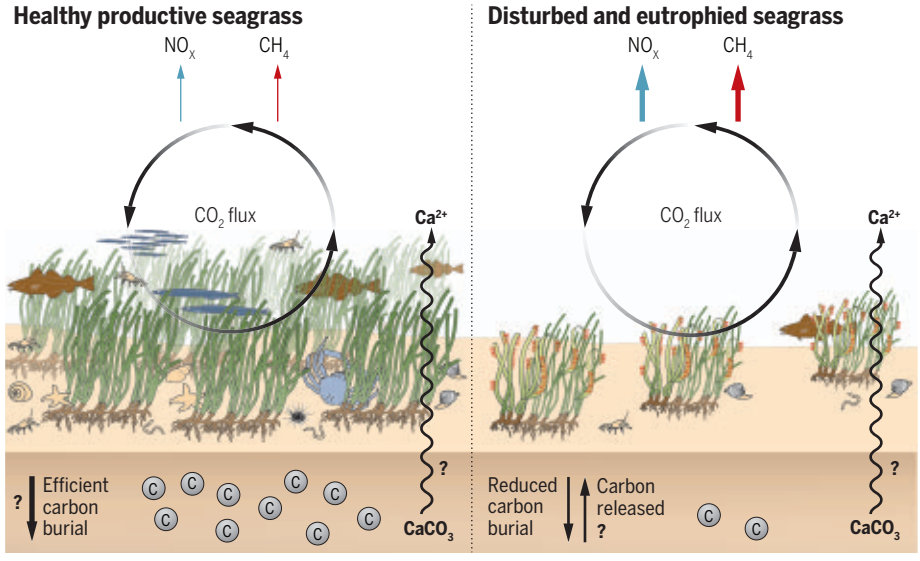


Fig. 2. The greenhouse gas balance of seagrass. There are many competing processes that result in seagrass meadows becoming net sources or sinks of greenhouse gases in our oceans. The left panel illustrates a healthy meadow where net photosynthetic productivity and dense seagrass is leading to rapid trapping and storage of carbon into the sediments. Although we lack a full understanding about greenhouse gas balance in seagrasses and the implications of disturbance, the right panel illustrates how meadow degradation and eutrophication can lead to the remobilization and loss of stored carbon and the potential increased production of CH_4 and N_2O. We also know little about the consequences of calcification by associated fauna within productive seagrass meadows on the overall carbon balance.

SDGs related to clean water and sanitation, good health, and well-being (*12*). Additionally, there is increasing appreciation of the value of seagrass for storing and sequestering carbon and the potential value of conserving seagrass meadows for climate mitigation (*61*). We understand that seagrasses enhance life below water, but less–well appreciated is that seagrass systems also enhance life on land by providing resources to shoreline habitats and populations, especially birds (*62*). The biodiversity present within seagrass meadows, the ecological processes and functions within them, and their relatively easy access also provide educational opportunities for human communities (*63*).

Without strong partnerships between communities, governments, nongovernmental organizations, and the private sector, seagrass conservation and restoration will not work effectively. The final SDG is about this bigger ambition. In the UK, the conservation charity Project Seagrass is bringing together private sector companies (e.g., CGI and Ocean Infinity), universities (e.g., Swansea and Heriot-Watt), institutes (e.g., NOC), and the government (e.g., the Hydrographic Agency) to map the UK's seagrass meadows. Similar initiatives are happening globally in places such as the Seychelles, Australia, and Indonesia.

Many aspects of the SDGs focus on the human planet, where the role that seagrasses play is changing with respect to a changing climate. With an expanding need to harness the energy of our oceans through wind, waves, and tide, there is increasing potential for new infrastructure to come into conflict with seagrass ecosystems. At the same time, this could lead to improved outcomes for seagrass, especially at a time when there is increasing global recognition of the need to develop strong criteria and indicators for pathways toward nature-positive outcomes. One such mechanism is that adopted in Australia, where marine biodiversity offsetting is accepted as a component of development consent to achieve an ambition of no net loss of biodiversity. A failed push toward tidal lagoon power in the UK provided impetus for seagrass restoration, and there is a growing focus on using seagrass restoration as a means of enhancing fish habitat as an offset to the effect of offshore wind power installations on marine biodiversity. The decline and reduced use of major historic urban coastal infrastructure, such as disused docklands, fisheries ponds, and mill ponds, are typical of many areas of the temperate Northern Hemisphere. The large empty docklands of South Wales provide an exemplary opportunity for seagrass restoration, and in southern Spain, entrepreneurial restaurateurs are bringing disused salt ponds back to life with seagrass for the growth of food products (*64*).

Charting a pathway to the net recovery of seagrass

Solutions for seagrass conservation and restoration have never been more urgent given the ongoing risks they face (*15*) and their potential role in helping mitigate climate change and the biodiversity crisis (*21*). Given the real and immediate threat of runaway climate change that places the future of humanity at risk, we need to rapidly move toward a conservation and restoration model that focuses on achieving global net recovery of seagrass (Fig. 4). Although financial mechanisms are emerging that begin to place monetary value onto seagrass carbon stores and carbon sequestration potential that will enable greater conservation and restorative action, concern exists about the potential for perverse and unintended consequences of such mechanisms (including international ownership of local resources), particularly around their role in supporting livelihoods (*56*).

It has been argued that avoiding a climate catastrophe requires at least three global transformations that are unprecedented in both magnitude and speed (*54*). One of these is a transformation of our relationship with nature to one that conserves, restores, and enhances its benefits for people and the planet (*54*). The SDGs could provide a valuable lens for securing the wider ecological role of seagrass meadows beyond carbon sequestration.

Seagrass habitats are global; estimates of loss are widespread and varied, but there is general agreement that the loss is vast. However, this does mean that there is huge potential for nature-based solutions focused on seagrass restoration. A restored seagrass meadow may take many years and be high cost in terms of

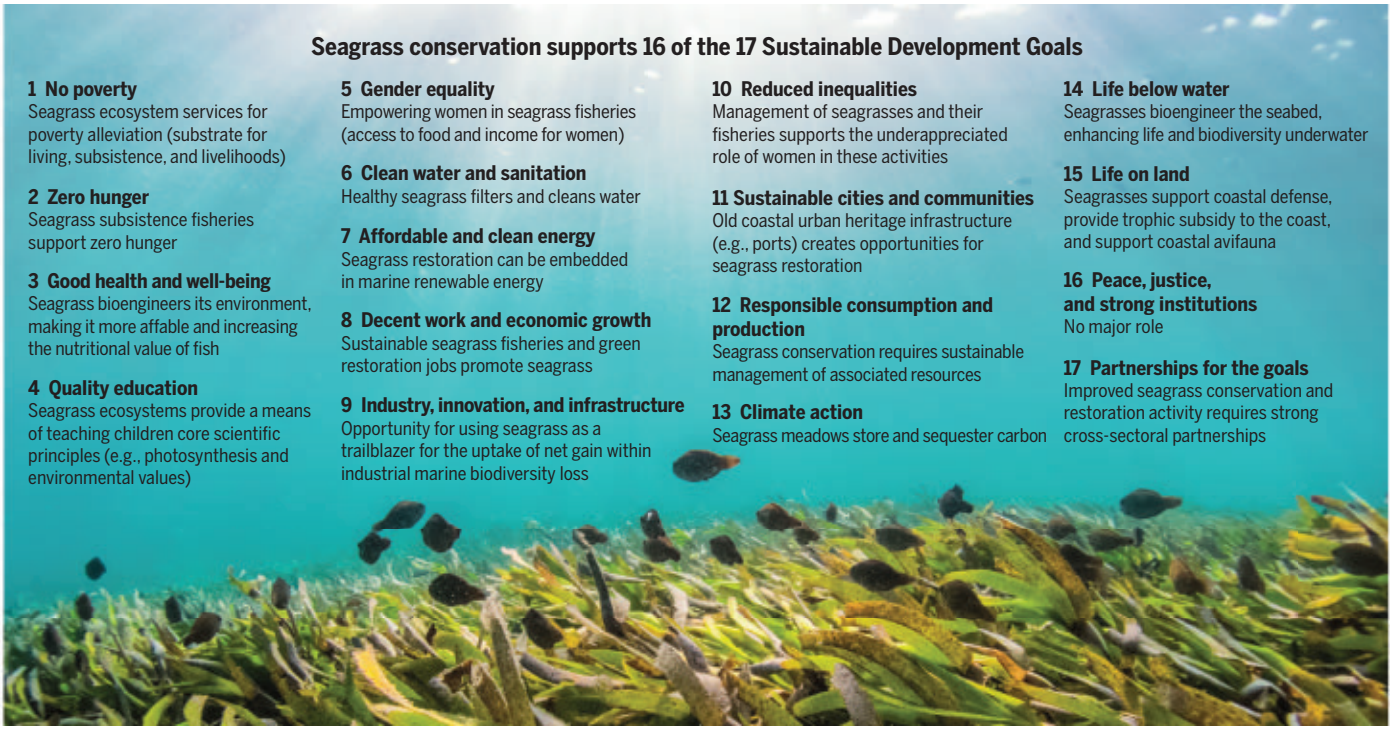


PHOTO BY BENJAMIN-L-JONES ON UNSPLASH

Fig. 3. Seagrass and sustainable development. Conservation and restoration of seagrass meadows and their ecological role can be communicated through the lens of the SDGs, of which seagrasses contribute to 16 of the 17 goals. A major part of this contribution is through the roles that they have as bioengineers and in supporting fisheries.

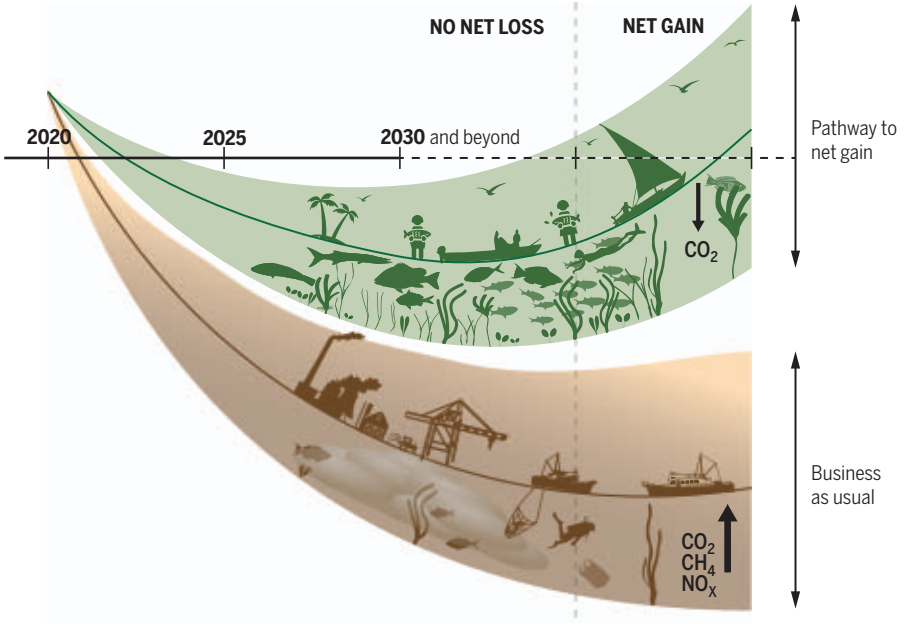


Fig. 4. A trajectory of seagrass recovery. Business as usual: Although seagrass meadows provide extensive ecosystem services and offer a major global opportunity as nature-based solutions, without intervention, they remain on a trajectory of global decline throughout the next century. Pathway to net gain: If major conservation action is taken to halt and reverse seagrass loss and degradation, then seagrasses can provide major contributions to fulfilling the aims of 16 of the 17 SDGs and for providing a major nature-based solution to climate change. Net gain of biodiversity requires avoidance of damage (e.g., legal instruments to halt bottom trawling) or minimization of effects that cannot be avoided, restoration to enhance or recreate habitats after damage (e.g., advanced mooring systems to allow recovery from boat damage), compensation, and recovery to enhance or recreate habitats known to have been historically lost or degraded (e.g., by active replanting). Image uses silhouettes created using symbols from the IAN Library, UMCES, University of Maryland.

labor and infrastructure to become ecologically functional (*65*, *66*). The opportunity provided by seagrass restoration should not detract from the urgent need to protect what we already have. As seagrass meadows become degraded, they not only begin to become net emitters of carbon, but they also release large amounts of nitrogen and sediments into the coastal ecosystem (*34*), together with any potential contaminants trapped within (e.g., heavy metals or plastics) (*67*). Achieving no net loss (and ultimately global net gain) of seagrasses requires scientific vision and political will (Fig. 3). This will not be easy, but we know that cumulative and connected conservation of seagrass over large scales can have major economic and environmental benefits (*65*). In general, plant conservation lags behind the conservation of animals (*68*), but seagrass could provide a model for how to overcome this so-called plant blindness, especially in the context of nature-based solutions (*69*).

Seagrasses have previously been described as the "ugly duckling" of marine conservation (*70*), but their star has risen with increasing interest in their potential to contribute to nature-based solutions to climate change and sustainable development. However, there are substantial ecological, social, and regulatory barriers and bottlenecks to seagrass restoration and conservation because of the scale of the interventions required. We must work inclusively at a local scale but in a globally connected network. Advances in marine robotics, molecular ecology, remote sensing, and artificial intelligence offer new opportunities to solve conservation problems in difficult environments at unprecedented global scales.

REFERENCES AND NOTES

1. L.-Y. Chen *et al*., *Mol. Biol. Evol.* **39**, msac079 (2022).
2. M. A. Hemminga, C. M. Duarte, *Seagrass Ecology* (Cambridge Univ. Press, ed. 1, 2000).
3. J. L. Olsen *et al*., *Nature* **530**, 331–335 (2016).
4. E. J. Hind-Ozan, B. L. Jones, *Mar. Pollut. Bull.* **134**, 223–227 (2018).
5. P. S. Maxwell *et al*., *Biol. Rev.* **92**, 1521–1538 (2017).
6. B. I. van Tussenbroek *et al*., *Nat. Commun.* **7**, 12980 (2016).
7. K. McMahon *et al*., *Proc. R. Soc. B.* **281**, 20140878 (2014).
8. J. M. Edgeloe *et al*., *Proc. R. Soc. B.* **289**, 20220538 (2022).
9. W. Mohr *et al*., *Nature* **600**, 105–109 (2021).
10. T. van der Heide *et al*., *Science* **336**, 1432–1434 (2012).
11. L. Panno *et al*., *New Biotechnol.* **30**, 685–694 (2013).
12. L. C. Cullen-Unsworth *et al*., *Mar. Pollut. Bull.* **83**, 387–397 (2014).
13. A. R. Hughes, S. L. Williams, C. M. Duarte, K. L. Heck Jr., M. Waycott, *Front. Ecol. Environ.* **7**, 242–246 (2009).
14. M. Huxham, D. Whitlock, M. Githaiga, A. Dencer-Brown, *Curr. For. Rep.* **4**, 101–110 (2018).
15. M. P. Turschwell *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2110802118 (2021).
16. A. Arias-Ortiz *et al*., *Nat. Clim. Chang.* **8**, 338–344 (2018).
17. M. Solé *et al*., *Commun. Biol.* **4**, 743 (2021).
18. L. J. McKenzie *et al*., *Environ. Res. Lett.* **15**, 074041 (2020).
19. M. Waycott *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12377–12381 (2009).
20. J. C. Dunic, C. J. Brown, R. M. Connolly, M. P. Turschwell, I. M. Côté, *Glob. Change Biol.* **27**, 4096–4109 (2021).
21. P. I. Macreadie *et al*., *Nat. Rev. Earth Environ.* **2**, 826–839 (2021).
22. J. W. Fourqurean *et al*., *Nat. Geosci.* **5**, 505–509 (2012).
23. C. B. de Los Santos *et al*., *Nat. Commun.* **10**, 3356 (2019).
24. CGI, "CGI announces strategic partnership with Project Seagrass to reduce CO2" (2021); https://www.cgi.com/uk/en-gb/news/climate/cgi-announces-strategic-partnership-project-seagrass-reduce-co2.
25. S. J. Pittman *et al*., "Outlining a Methodological Pathway to Improve the Global Seagrass Map" (Pew Charitable Trusts, 2022).
26. M. E. Asplund *et al*., *Front. Mar. Sci.* **8**, 811533 (2022).
27. A. N. Al-Haj, R. W. Fulweiler, *Glob. Change Biol.* **26**, 2988–3005 (2020).
28. S. Schorn *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2106628119 (2022).
29. P. I. Macreadie *et al*., *One Earth* **5**, 485–492 (2022).
30. E. Fernández, C. Barañano, S. Villasante, G. Méndez, *Ocean Coast. Manage.* **216**, 105983 (2022).
31. R. Felger, M. B. Moser, *Science* **181**, 355–356 (1973).
32. L. Nordlund, R. Unsworth, M. Gullström, L. Cullen-Unsworth, *Fish Fish.* **19**, 399–412 (2018).
33. R. K. F. Unsworth, L. M. Nordlund, L. C. Cullen-Unsworth, *Conserv. Lett.* **12**, e12566 (2019).
34. P.-O. Moksnes *et al*., *Ecosphere* **12**, e03658 (2021).
35. Y. Deng, S. Liu, J. Feng, Y. Wu, C. Mao, *Mar. Pollut. Bull.* **166**, 112229 (2021).
36. J. B. Lamb *et al*., *Science* **355**, 731–733 (2017).
37. T. B. H. Reusch *et al*., *Mar. Biol.* **168**, 149 (2021).
38. F. A. Ascioti, M. C. Mangano, C. Marcianò, G. Sarà, *Ecosyst. Serv.* **54**, 101418 (2022).
39. A. M. Ricart *et al*., *Glob. Change Biol.* **27**, 2580–2591 (2021).
40. R. K. F. Unsworth, C. J. Collier, G. M. Henderson, L. J. McKenzie, *Environ. Res. Lett.* **7**, 024026 (2012).
41. S. K. Papworth, J. Rist, L. Coad, E. J. Milner-Gulland, *Conserv. Lett.* **2**, 93–100 (2009).
42. A. E. Green, R. K. F. Unsworth, M. A. Chadwick, P. J. S. Jones, *Front. Plant Sci.* **12**, 629962 (2021).
43. J. E. Duffy *et al*., *Ecol. Lett.* **18**, 696–705 (2015).
44. C. P. Gross *et al*., *Proc. R. Soc. B.* **289**, 20211762 (2022).
45. R. K. F. Unsworth *et al*., *Ambio* **48**, 801–815 (2019).
46. S. Baden, A. Emanuelsson, L. Pihl, C. J. Svensson, P. Åberg, *Mar. Ecol. Prog. Ser.* **451**, 61–73 (2012).
47. M. J. A. Christianen *et al*., *Proc. R. Soc. B.* **281**, 20132890 (2014).
48. B. L. H. Jones *et al*., *Ecol. Soc.* **27**, 10 (2022).
49. C. R. Biggs *et al*., *Ecosphere* **11**, e03184 (2020).
50. R. K. F. Unsworth, E. G. Butterworth, *Diversity* **13**, 363 (2021).
51. M. P. Dias *et al*., *Biol. Conserv.* **237**, 525–537 (2019).
52. P. L. Reynolds *et al*., *Ecology* **99**, 29–35 (2018).
53. M. A. Whalen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28160–28166 (2020).
54. J. Rockström *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2115218118 (2021).
55. J. S. Lefcheck *et al*., *Conserv. Lett.* **12**, e12645 (2019).
56. A. P. Hejnowicz, H. Kennedy, M. A. Rudd, M. R. Huxham, *Front. Mar. Sci.* **2**, 32 (2015).
57. F. Asche *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11221–11225 (2018).
58. L. C. L. Teh, U. R. Sumaila, *Fish Fish.* **14**, 77–88 (2013).
59. M. de la Torre-Castro, *Front. Mar. Sci.* **6**, 156 (2019).
60. T. E. A. L. Quiros, M. W. Beck, A. Araw, D. A. Croll, B. Tershy, *Ocean Coast. Manage.* **157**, 56–67 (2018).
61. M. E. Röhr *et al*., *Global Biogeochem. Cycles* **32**, 1457–1475 (2018).
62. K. L. HeckJr *et al*., *Ecosystems* **11**, 1198–1210 (2008).
63. A. Syukur, L. Zulkifli, Mahrus, K. R. Dewi, in *Proceedings of the 5th Asian Education Symposium 2020 (AES 2020)* (Atlantis Press, 2021), pp. 115–118.
64. M. Goulding, "Seeding the Ocean: Inside a Michelin-Starred Chef's Revolutionary Quest to Harvest Rice From the Sea" (Time Magazine, 2021); https://time.com/5926780/chef-angel-leon-sea-rice/.
65. R. J. Orth *et al*., *Sci. Adv.* **6**, eabc6434 (2020).
66. E. Bayraktarov *et al*., *Ecol. Appl.* **26**, 1055–1074 (2016).
67. R. K. F. Unsworth *et al*., *Oceans* **2**, 162–178 (2021).
68. M. Balding, K. J. H. Williams, *Conserv. Biol.* **30**, 1192–1199 (2016).
69. S. B. Jose, C.-H. Wu, S. Kamoun, *Plants People Planet* **1**, 169–172 (2019).
70. C. M. Duarte, W. C. Dennison, R. J. W. Orth, T. J. B. Carruthers, *Estuaries Coasts* **31**, 233–238 (2008).

ACKNOWLEDGMENTS

Competing interests: The authors have no conflicts of interest.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 25 April 2022; accepted 9 June 2022
10.1126/science.abq6923

MICHELSON PHILANTHROPIES & Science PRIZE FOR IMMUNOLOGY

REWARDING HIGH-RISK RESEARCH. SUPPORTING EARLY-CAREER SCIENTISTS. HELPING TO FIND CURES FASTER. **APPLY TODAY**

Now accepting applications for the Michelson Philanthropies & *Science* Prize for Immunology.

The Michelson Philanthropies and Science Prize for Immunology focuses on transformative research in human immunology, with trans-disease applications to accelerate vaccine and immunotherapeutic discovery. This international prize supports investigators 35 and younger, who apply their expertise to research that has a lasting impact on vaccine development and immunotherapy. It is open to researchers from a wide range of disciplines including computer science, artificial intelligence/machine learning, protein engineering, nanotechnology, genomics, parasitology and tropical medicine, neurodegenerative diseases, and gene editing.

Application deadline: Oct. 1, 2022.

For more information visit:
www.michelsonmedicalresearch.org

#MichelsonPrizes

"The Michelson Philanthropies & Science Prize for Immunology will greatly impact my future work. As I am just starting my scientific career, it will illuminate my work, spark interest and support me to continue my research in this field."

Paul Bastard, MD, PhD,
Laboratory of Human Genetics of Infectious Diseases, Imagine Institute (INSERM, University of Paris), Paris, France; and The Rockefeller University, New York.

Dr. Bastard received the inaugural Grand Prize for his essay: "Why do people die from COVID-19: Autoantibodies neutralizing type I interferons increase with age."

GRAND PRIZE:
$30,000

FINALIST PRIZE:
$10,000

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk

PHOTO: ARIELLE S. FOGEL

HYBRIDIZATION

Hidden selection against interbreeding

Today, humans are the only extant members of our genus, *Homo*. This was not the case in the past, because we now know that our ancestors shared the planet with other *Homo* species. It has been suggested that selection against hybrid individuals would have acted against breeding across these species, but such a hypothesis is difficult to test today. To study this question, Vilgalys *et al*. took advantage of a decades-long dataset on two species of baboon from the Amboseli basin of Kenya. They found evidence of selection against hybrid, or admixed, ancestry similar to what has been predicted for ancestral hominids. Although evidence for selection against hybrids was clear, they also found that individual hybrids can thrive. —SNV *Science*, abm4917, this issue p. 635

Baboons in the Amboseli basin of Kenya, such as the individual here, show signs of selection against genetic admixture between species.

PLANT SCIENCE

Shift from meiosis to mitosis

In angiosperm plants, meiosis leads to a haploid but multicellular reproductive structure. Cairo *et al*. studied the molecular regulation of the shift from meiosis to mitosis that generates this reproductive structure in the small mustard plant *Arabidopsis*. They found that a germline-specific protein that drives meiotic exit is incorporated into P-bodies during the second meiotic division, where it transiently sequesters translation initiation factors, thus inhibiting translation. The authors suggest that this recognition process remodels the translatome and facilitates the transition to mitosis. —PJH

Science, abo0904, this issue p. 629

INFECTIOUS DISEASES

Repurposing against parasites

Available drugs used to combat the apicomplexan parasites *Toxoplasma gondii* and *Plasmodium falciparum* have limited efficacy and undesirable side effects. Using a drug-repurposing screen, Swale *et al*. identified altiratinib, a drug originally developed to treat glioblastoma, as a potential antiparasitic drug. Through drug resistance screening and genetic target deconvolution, a *T. gondii* kinase involved in cell cycle progression was identified as the primary target of altiratinib, which was confirmed for the corresponding kinase in *P. falciparum*. Altiratinib was shown to globally disrupt protein splicing by targeting this family of parasite kinases. These findings support the further development of pan-apicomplexan inhibitors that target this pathway. —CNF

Sci. Transl. Med. **14**, eabn3231 (2022).

QUANTUM DYNAMICS

Continuous time crystals

Time crystals are a new dynamical phase of quantum matter resulting from the breaking of time-translation symmetry and the subsequent interplay between interactions forming self-organized phases. To date, discrete time crystals have been observed in periodically driven systems. By contrast, Kongkhambut *et al*. report the observation of spontaneous breaking of a continuous time translation symmetry in an atomic Bose-Einstein condensate inside a high-finesse optical cavity (see the Perspective by LeBlanc). Using a time-independent pump, the authors observed a limit cycle phase that was characterized by emergent periodic oscillations of the intracavity photon number and was accompanied by the atomic density cycling through recurring patterns: a continuous time crystal. —ISO

Science, abo3382, this issue p. 670; see also add2015, p. 576

T CELLS

Intestinal epithelium maintains T cells

Immune surveillance of the intestinal epithelium is crucial for intestinal homeostasis and protecting against infection. Seo *et al.* used mouse knockout models to demonstrate that expression of herpes virus entry mediator (HVEM) on intestinal epithelial cells is crucial for the survival and patrolling ability of intraepithelial CD8α T cells. The protein LIGHT, a ligand for HVEM, was crucial for these HVEM-driven effects. Epithelial knockout of HVEM led to poorer protection against intestinal bacterial infection. These data show that HVEM expression in intestinal epithelial cells is involved in intraepithelial T cell–mediated immune surveillance of the small intestine.—DAE

Sci. Immunol. **7**, eabm6931 (2022).

CRYSTALLIZATION

Growing crystalline containers

Hierarchical crystals are often found in nature, and biological processes delicately control their growth and patterning. For synthetic systems, this process remains challenging. For example, skeletal crystals with concave morphology usually require additives and complicated temperature control. Oki *et al*. succeeded in growing teacup-shaped microcrystals in a controlled, uniaxial manner by casting a solution of a planar-chiral molecule on a clean quartz substrate and evaporating it, causing supersaturation and crystal growth. The authors show that despite the overall geometrical complexity of the crystal, it emerges from the highly symmetric ordering of the chiral planar molecules that stack on one another with a counterclockwise rotation of 60° along a crystallographic sixfold screw axis. —MSL

Science, abm9596, this issue p. 673

Scanning electron image of a partially formed cup-shaped crystal

OCEAN OXYGEN

Deep concentration

Between about 1.25 million and 800,000 years ago, the climate system went through a major change during a period called the Middle Pleistocene Transition. Was the inventory of dissolved oxygen in the ocean affected by this episode? Thomas *et al*. show that oxygen concentrations in glacial deep North Atlantic waters suffered a stepped reduction about 900,000 years ago, coincident with reductions in the concentration of glacial atmospheric carbon dioxide and global ice volume. —HJS

Science, abj7761, this issue p. 654

ORGANIC CHEMISTRY

Carbonyls to carbenes

Carbenes are versatile chemical intermediates because of their highly reactive divalent carbon centers. Unfortunately, typical carbene precursors such as diazo compounds are often sensitive and decompose explosively. Zhang *et al*. now report a comparatively safe protocol to generate carbenes from aldehydes, a plentiful and easily diversified substrate class (see the Perspective by West and Rousseaux). The sequence entails successive reactions of the aldehyde with pivaloyl chloride and zinc, followed by catalytic activation using iron, cobalt, or copper. Demonstrated applications included wide-ranging examples of three-membered ring formation and sigma-bond insertion. —JSY

Science, abo6443, this issue p. 649; see also abq8253, p. 580

IN OTHER JOURNALS

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

MACHINE LEARNING

Machine learning for surface catalysis

To date, only a tiny fraction of the enormous diversity of elementary surface reactions has been explored, yet it has already provided invaluable assistance in improving our understanding of heterogeneous catalysis. Microkinetic modeling of heterogeneous catalysis (i.e., exploring a detailed list of elementary surface reactions) is challenging because it requires simultaneous consideration of many different aspects of complex interfacial processes. Shi *et al.* developed a comprehensive machine learning–based approach for automated reaction pathway analysis and generated a detailed microkinetic model for carbon dioxide/carbon monoxide hydrogenation on copper–zinc surfaces, which resolved several long-standing questions about this industrial process. The proposed approach can operate with thousands of reaction pathways and different surface coverages and is potentially applicable to other heterogeneous catalytic systems. —YS

J. Am. Chem. Soc. 10.1021/jacs.2c06044 (2022).

MARINE SCIENCE

Eavesdropping on whales

Wildlife monitoring in the oceans can be challenging because of widespread undersampling. Recent developments in using undersea fiber-optic cables for sensing provide a different opportunity to solve this problem. Bouffaut *et al.* were able to successfully monitor whales with distributed acoustic

PHOTOS (LEFT TO RIGHT): OKI ET AL.; CLARKANDCOMPANY/GETTY IMAGES

BIOGEOGRAPHY

Dispersal effects on island diversity

Islands that are smaller or farther from the mainland tend to have fewer species. Walentowitz *et al.* asked how this classic pattern varies across plant species with different modes of seed dispersal (i.e., self-dispersing or moved by animals, wind, or water) and whether it is influenced by human impacts. Using data from 54 small islands off the coast of Denmark, the authors found that area had stronger effects than isolation on plant species richness and composition. Larger islands had more species and more animal- and wind-dispersed species. Human activities increased species richness and may also increase seed dispersal between islands, effectively reducing their isolation, but did not erase the effects of island area. —BEL

J. Biogeogr. 10.1111/jbi.14454 (2022).

The Danish archipelago displays a spectrum of plant biodiversity depending on island size.

sensing using fiber-optic cables in the Arctic. The authors recorded vocalizations between baleen whales, estimated their locations, and used the vocalizations for seafloor subsurface exploration. Their observations demonstrate the potential for using fiber-optic infrastructure to listen to the less-accessible parts of our oceans. —BG

Front. Mar. Sci. 10.3389/fmars.2022.901348 (2022).

NEUROSCIENCE

Memory formation in sleep

Forming indirect associations between learned items with overlapping components is crucial for abstract problem solving. This type of learning is a fundamental feature of relational memory. Sleep is important for learning relational memory tasks. However, we do not know what biophysical mechanisms enable this function. Using a biophysical model of a thalamocortical network, Tadros and Bazhenov tested the role of non–rapid eye movement sleep (during which the sleeper experiences slower muscular and brain activity) on the network's ability to perform a relational memory task. After periods of slow-wave sleep, the network could form indirect inferences that were never trained directly. Sleep replay increased connections to and from a shared conjunctive memory unit, which increased its performance during relational memory tasks. These modeling studies produced predictions that can now be tested by experiments. —PRS

J. Neurosci. **42**, 5330 (2022).

HEART REGENERATION

Old hearts learn new tricks

Aging-related diseases such as heart failure and other cardiovascular disorders are the leading causes of death in many countries, and they are becoming increasingly common worldwide as the number of older people increases. The ability of the heart to produce new cardiomyocytes decreases with age, which makes it more difficult to repair damage and increases the risk of heart failure. However, a study by Lerchenmüller *et al*. suggests that exercise may offer some help in this regard even if started late in life. The authors had previously reported that voluntary exercise can stimulate the generation of cardiomyocytes in young adult mouse hearts, and now they have also observed this phenomenon in aged animals. —YN

Circulation 10.1161/CIRCULATIONAHA.121.057276 (2022).

POLITICAL ECONOMY

Paying for patriots

US federal government spending in the 1930s under Depression-era New Deal social programs "created a new geography of patriotism" manifested a short time later during World War II. Caprettini and Voth show that government war bond purchases, volunteering for military service, and being awarded for heroism in the war, all of which entailed personal sacrifice for the country, were more common in counties that received more per capita government relief under the New Deal. This work refines our understanding of social contracts and reciprocity between governments and constituents. —BW

Q. J. Econ. 10.1093/qje/qjac028 (2022).

CELL BIOLOGY

Fusexins found in archaea

Fusexins are proteins that mediate cellular membrane fusion. These proteins were initially discovered in viruses, in which they merge the viral envelope with host cell membranes during invasion. In eukaryotes, fusexins mediate syncytial tissue development in worms and gamete fusion in plants and protists. Despite their widespread presence in eukaryotes, it is not known whether fusexins originated in the first eukaryotes and were later captured by viruses or the other way around. Moi *et al.* found that archaea contain the protein Fusexin1, which is structurally similar to fusexins from viruses, plants, and animals and can promote cell-cell fusion. This discovery raises the possibility that gamete fusion proteins originated in archaeal cells. —SMH

Nat. Commun. **13**, 3880 (2022).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

A mosaic approach to protection

The COVID-19 pandemic has been ongoing for more than 2 years now, and new variants such as Omicron are less susceptible to the vaccines developed against earlier lineages of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In addition, there is continued risk of spillovers of other animal sarbecoviruses into humans. There is thus a need for vaccines that will give broader protection. Cohen *et al*. developed mosaic nanoparticles that display the receptor-binding domains (RBDs) from SARS-CoV-2 and seven other animal sarbecoviruses. Mosaic nanoparticles protected against both SARS-CoV-2 and SARS-CoV challenges in animal models even though the SARS-CoV RBD was not present on the mosaic-8 RBD nanoparticles. By contrast, a homotypic SARS-CoV-2 RBD nanoparticle (presenting only SARS-CoV-2 RBDs) only protected against a SARS-CoV-2 challenge. —VV

Science, abq0839, this issue p. 618

HEART DISEASE

A close-up look at cardiomyopathies

Cardiomyopathies are diseases of the heart muscle that interfere with its ability to pump blood effectively and can result in heart failure. They are divided into different categories based on the clinical presentation. Reichart *et al*. performed single-nucleus RNA sequencing on heart samples from patients with cardiomyopathies with or without known genetic causes, as well as samples from controls without structural heart disease. The authors identified key cell types and their locations in the heart, cellular interactions, and biological signaling pathways, offering insights into the biology of the heart in healthy and diseased states. —YN

Science, abo1984, this issue p. 619

DEVELOPMENT

Drosophila embryo analyzed cell by cell

Animal development can progress quite rapidly, with cellular lineages proliferating and differentiation status changing minute to minute. Calderon *et al*. have now visualized development in the fruit fly *Drosophila* in greater detail than ever before. Taking advantage of the ability to produce collections of *Drosophila* embryos that differ in developmental stage by only seconds or minutes, the authors have analyzed, on a single-cell basis, how chromatin accessibility and gene expression shift during *Drosophila* embryogenesis. This single-cell atlas of *Drosophila* embryogenesis reveals cell lineages and their developmental relationships and links enhancer usage and gene expression. —BAP

Science, abn5800, this issue p. 620

MEDICINE

Mitochondrial protection from ischemia

Damage caused by lack of oxygen, or ischemia, is an important consequence of heart attack and stroke. Kynurenic acid, a product of metabolism of the amino acid tryptophan, has shown protective effects against ischemia in animal models. Wyant *et al.* found that these effects are mediated by the heterotrimeric guanine nucleotide–binding protein-coupled receptor (GPCR) GPR35 (see the Perspective by Cadenas). Mouse hearts lacking GPR35 lost the beneficial effect when treated with kynurenic acid. In mouse neonatal cardiomyocytes stimulated with kynurenic acid, GPR35 associated with mitochondria and appeared to interact indirectly with ATP synthase inhibitory factor subunit 1, reducing the loss of ATP during ischemia. The findings further support the possible therapeutic use of GPR35 agonists to reduce the deleterious effects of ischemia. —LBR

Science, abm1638, this issue p. 621; see also add4629, p. 579

SYSTEMS BIOLOGY

Cellular information processing

Prior studies have shown that the responses of individual cells to growth factors are variable, raising the question of whether cells detect more information than the mere presence or absence of growth factor. Kramer *et al*. monitored the response of cells to various concentrations of epidermal growth factor. They found that responses are indeed variable, apparently because signaling nodes store other information about the state of the cell and thus respond differently to the same concentration of growth factor. Thus, cells have a way to process information in the context of their surroundings and current cell state. —LBR

Science, abf4062, this issue p. 642

IMMUNOLOGY

The "hex" factor behind IEL development

$CD4^+CD8\alpha\alpha^+$ intraepithelial lymphocytes (CD4IELs) are a class of intestinal innate-like T cells that contribute to various immune responses, including oral tolerance. Their development depends on the gut microbiota, but the precise antigens that these cells recognize have remained elusive. Bousbaine *et al*. report that β-hexosaminidase (β-hex), a highly conserved enzyme produced by commensals from the Bacteroidetes phylum, drives CD4IEL differentiation in the gut (see the Perspective by Shenoy and Koch). T cell receptors from both tissue-resident CD4IELs and regulatory T cells in gut-draining lymph nodes recognized β-hex peptides. β-hex–specific CD4 T cells transferred into mice differentiated into CD4IELs that localized to the small intestine. There, they partially suppressed inflammation in a regulatory T cell–independent manner. This study highlights how intestinal immune responses to commensal bacteria can regulate inflammation. —STS

Science, abg5645, this issue, p. 660; see also add7145, p. 575

MUSCLE REPAIR

JMJD3 primes stem cells for inflammation

After muscle injury, muscle stem cells must coordinate with immune cells in the inflamed tissue to ensure efficient repair. Nakka *et al*. identified an essential role for the epigenetic enzyme KDM6B/JMJD3 in establishing the communication between muscle stem cells and infiltrating immune cells during muscle repair (see the Perspective by Gabellini). They found that, in response to injury, removal of the transcriptionally repressive histone $H3K27me^3$ modification by KDM6B/JMJD3 allows muscle stem cells to produce hyaluronic acid that is then incorporated into the extracellular matrix. This remodeling of the extracellular matrix allows the muscle stem cell to receive signals from the infiltrating immune cells that initiate regeneration. —SMH and BAP

Science, abm9735, this issue p. 666; see also add6804, p. 578

CANCER

CD9 tags tumorigenic packages from CAFs

Extracellular vesicles (EVs) shed from pancreatic cancer–associated fibroblasts containing the membrane protein ANXA6

induce aggressive phenotypes in tumors. By analyzing EVs from cancer–associated fibroblasts from patient samples and mouse models of pancreatic ductal carcinoma, Nigri *et al*. found that the cell-surface glycoprotein CD9 mediated the uptake of ANXA6-positive EVs. Either blocking CD9 on isolated EVs or inhibiting the kinase p38 in pancreatic ductal carcinoma cells prevented ANXA6-positive EVs from inducing cell migration and a mesenchymal protein signature. —LKF

Sci. Signal. 15, eabg8191 (2022).

CANCER

Engineered protein blocks cancer stem cells

Glioblastoma multiforme (GBM) is the most lethal brain cancer in adults. At present, there is no cure for this devastating disease, and current therapies minimally increase overall survival. Benedetti *et al*. engineered SOX2, a key transcription factor implicated GBM malignancy, with the aim of rewiring a key oncogenic gene network. The authors show that this strategy can block tumor growth and cancer stem cell activities in vitro and in mouse xenografts of human tumors. Because Sox2 has a primary role in promoting tumor development not just in GBM but also in lung, prostate, and breast cancers, this approach might be an innovative strategy against other deadly cancers. —LDC

Sci. Adv. 10.1126/sciadv.abn3986 (2022).

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Mosaic RBD nanoparticles protect against challenge by diverse sarbecoviruses in animal models

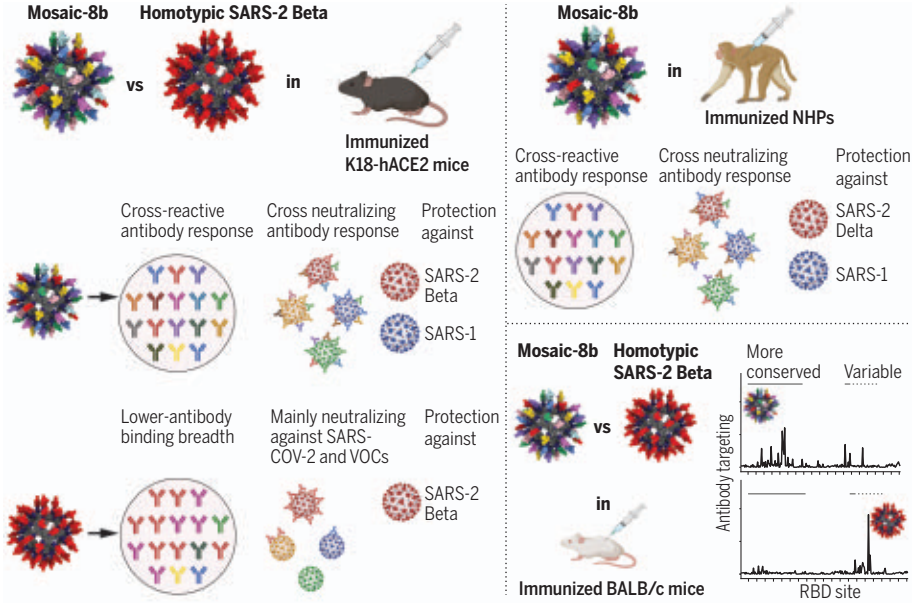
Alexander A. Cohen†, Neeltje van Doremalen†, Allison J. Greaney, Hanne Andersen, Ankur Sharma, Tyler N. Starr, Jennifer R. Keeffe, Chengcheng Fan, Jonathan E. Schulz, Priyanthi N. P. Gnanapragasam, Leesa M. Kakutani, Anthony P. West Jr., Greg Saturday, Yu E. Lee, Han Gao, Claudia A. Jette, Mark G. Lewis, Tiong K. Tan, Alain R. Townsend, Jesse D. Bloom, Vincent J. Munster, Pamela J. Bjorkman*

INTRODUCTION: Two animal coronaviruses from the severe acute respiratory syndrome (SARS)–like betacoronavirus (sarbecovirus) lineage, SARS coronavirus (SARS-CoV) and SARS-CoV-2, have caused epidemics or pandemics in humans in the past 20 years. SARS-CoV-2 triggered the COVID-19 pandemic that has been ongoing for more than 2 years despite rapid development of effective vaccines. Unfortunately, new SARS-CoV-2 variants, including multiple heavily mutated Omicron variants, have prolonged the COVID-19 pandemic. In addition, the discovery of diverse sarbecoviruses in bats raises the possibility of another coronavirus pandemic. Hence, there is an urgent need to develop vaccines and therapeutics to protect against both SARS-CoV-2 variants and zoonotic sarbecoviruses with the potential to infect humans.

RATIONALE: To combat future SARS-CoV-2 variants and spillovers of sarbecoviruses threatening global health, we designed nanoparticles that present 60 randomly arranged spike receptor-binding domains (RBDs) derived from the spike trimers of eight different sarbecoviruses (mosaic-8 RBD nanoparticles) to elicit antibodies against conserved and relatively occluded—rather than variable, immunodominant, and exposed—epitopes. The probability of two adjacent RBDs being the same is low for mosaic-8 RBD nanoparticles, a feature chosen to favor interactions with B cells whose bivalent receptors can cross-link between adjacent RBDs to use avidity effects to favor recognition of conserved, but sterically occluded, RBD epitopes. By contrast, nanoparticles that present 60 copies of SARS-CoV-2 RBDs (homotypic RBD nanoparticles) are theoretically more likely to engage B cells with receptors that recognize immunodominant and sterically accessible, but less conserved, RBD epitopes.

RESULTS: We compared immune responses elicited by mosaic-8 (SARS-CoV-2 RBD plus seven animal sarbecoviruses RBDs) and homotypic (only SARS-CoV-2 RBDs) nanoparticles in mice and macaques and observed stronger responses elicited by mosaic-8 to mismatched (not represented with an RBD on nanoparticles) strains, including SARS-CoV and animal sarbecoviruses. Mosaic-8 immunization produced antisera that showed equivalent neutralization of SARS-CoV-2 variants, including Omicron variants, and protected from both SARS-CoV-2 and SARS-CoV challenges in mice and nonhuman primates (NHPs), whereas homotypic SARS-CoV-2 immunization protected from SARS-CoV-2 challenge but not from SARS-CoV challenge in mice. Epitope mapping of polyclonal antisera by using deep mutational scanning of RBDs demonstrated targeting of conserved epitopes after immunization with mosaic-8 RBD nanoparticles, in contrast with targeting of variable epitopes after homotypic SARS-CoV-2 RBD nanoparticle immunization, which supports the hypothesized mechanism by which mosaic RBD nanoparticle immunization can overcome immunodominance effects to direct production of antibodies against conserved RBD epitopes. Given the recent plethora of SARS-CoV-2 variants that may be arising at least in part because of antibody pressure, a relevant concern is whether more conserved RBD epitopes might be subject to substitutions that would render vaccines and/or monoclonal antibodies targeting these regions ineffective. This scenario seems unlikely because RBD regions conserved between sarbecoviruses and SARS-CoV-2 variants are generally involved in contacts with other regions of spike trimer and therefore less likely to tolerate selection-induced substitutions.

CONCLUSION: Together, these results suggest that mosaic-8 RBD nanoparticles could protect against SARS-CoV-2 variants and future sarbecovirus spillovers—in particular, highlighting the potential for a mosaic nanoparticle approach to elicit more broadly protective antibody responses than those with homotypic nanoparticle approaches. ■



Mosaic RBD nanoparticle vaccination protects and elicits antibodies against conserved epitopes. Mosaic-8 elicited broader cross-reactive responses than those of homotypic nanoparticles. In a stringent infection model [K18-human angiotensin-converting enzyme 2 (K18-hACE2)], both protected against matched challenge (SARS-CoV-2 Beta), but only mosaic-8 also protected against a mismatch (SARS-CoV). Mosaic-8–immunized NHPs were protected against mismatched SARS-CoV-2 (Delta) and SARS-CoV. Mosaic-8–elicited antibodies predominantly bound conserved epitopes, whereas homotypic-elicited antibodies predominantly bound variable epitopes.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: bjorkman@caltech.edu
†These authors contributed equally to this work.
Cite this article as A. A. Cohen *et al.*, *Science* **377**, eabq0839 (2022). DOI: 10.1126/scienceabq0839

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abq0839

RESEARCH ARTICLE

CORONAVIRUS

Mosaic RBD nanoparticles protect against challenge by diverse sarbecoviruses in animal models

Alexander A. Cohen[1]†, Neeltje van Doremalen[2]†, Allison J. Greaney[3,4], Hanne Andersen[5], Ankur Sharma[5], Tyler N. Starr[3,4], Jennifer R. Keeffe[1], Chengcheng Fan[1], Jonathan E. Schulz[2], Priyanthi N. P. Gnanapragasam[1], Leesa M. Kakutani[1], Anthony P. West Jr.[1], Greg Saturday[2], Yu E. Lee[1]‡, Han Gao[1], Claudia A. Jette[1], Mark G. Lewis[5], Tiong K. Tan[6], Alain R. Townsend[6,7], Jesse D. Bloom[8,9], Vincent J. Munster[2], Pamela J. Bjorkman[1]*

To combat future severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants and spillovers of SARS-like betacoronaviruses (sarbecoviruses) threatening global health, we designed mosaic nanoparticles that present randomly arranged sarbecovirus spike receptor-binding domains (RBDs) to elicit antibodies against epitopes that are conserved and relatively occluded rather than variable, immunodominant, and exposed. We compared immune responses elicited by mosaic-8 (SARS-CoV-2 and seven animal sarbecoviruses) and homotypic (only SARS-CoV-2) RBD nanoparticles in mice and macaques and observed stronger responses elicited by mosaic-8 to mismatched (not on nanoparticles) strains, including SARS-CoV and animal sarbecoviruses. Mosaic-8 immunization showed equivalent neutralization of SARS-CoV-2 variants, including Omicrons, and protected from SARS-CoV-2 and SARS-CoV challenges, whereas homotypic SARS-CoV-2 immunization protected only from SARS-CoV-2 challenge. Epitope mapping demonstrated increased targeting of conserved epitopes after mosaic-8 immunization. Together, these results suggest that mosaic-8 RBD nanoparticles could protect against SARS-CoV-2 variants and future sarbecovirus spillovers.

Two animal coronaviruses from the sarbecovirus lineage, severe acute respiratory syndrome coronavirus (SARS-CoV) and SARS-CoV-2 (hereafter SARS-1 and SARS-2, respectively), have caused epidemics or pandemics in humans in the past 20 years. SARS-2 triggered the COVID-19 pandemic that has been ongoing for more than 2 years despite rapid development of effective vaccines (*1*). Unfortunately, new SARS-2 variants of concern (VOCs), including the heavily mutated Omicron VOCs (*2*–*7*), have prolonged the COVID-19 pandemic. In addition, the discovery of diverse sarbecoviruses in bats—some of which bind the SARS-1 and SARS-2 entry receptor, angiotensin-converting enzyme 2 (ACE2) (*8*–*14*)—raises the possibility of another coronavirus pandemic. Hence, there is an urgent need to develop vaccines and therapeutics to protect against both SARS-2 VOCs and zoonotic sarbecoviruses.

Currently approved SARS-2 vaccines include the viral spike (S) trimer (*1*), which is consistent with S being the primary target of neutralizing antibodies (*15*–*24*). A coronavirus S trimer mediates entry into a host cell after one or more of its receptor-binding domains (RBDs) adopt an "up" position that allows interactions with a host cell receptor (Fig. 1A). Many of the most potent neutralizing antibodies against SARS-2 block binding of ACE2 to the RBD (*16*–*20*, *23*–*29*), and RBD targeting has been suggested for COVID-19 vaccine development (*30*). We classified neutralizing anti-RBD antibodies into four main classes (classes 1, 2, 3, and 4) on the basis of their epitopes and whether they recognized "up" and/or "down" RBDs on S trimers (*26*). The potent class 1 and class 2 antibodies to RBD, whose epitopes overlap with the ACE2 binding footprint, recognize a portion of the RBD that exhibits high sequence variability between sarbecoviruses (*26*). By contrast, the epitopes of class 4 antibodies, and to a somewhat lesser extent class 3 antibodies, map to more conserved, but less accessible, regions of sarbecovirus RBDs (Fig. 1A). Substitutions in the RBDs of VOCs and variants of interest (VOIs) are also less common in the class 4 and class 3 regions (Fig. 1A and fig. S1), thus suggesting that a vaccine strategy designed to elicit class 3, class 4, and class 1/4 [class 4-targeting antibodies that block ACE2 binding (*31*–*33*)] could protect against potentially emerging zoonotic sarbecoviruses as well as current and future SARS-2 variants.

Here, we describe animal immunogenicity and virus challenge studies to evaluate mosaic-8 RBD nanoparticles, a potential pan-sarbecovirus vaccine in which RBDs from SARS-2 and seven animal sarbecoviruses (*14*) were covalently attached to a 60-mer protein nanoparticle (*34*). The probability of two adjacent RBDs being the same is low for mosaic-8 RBD nanoparticles, an arrangement chosen to favor interactions with B cells whose receptors can cross-link between adjacent RBDs to use avidity effects (*35*) to preferentially recognize conserved, but sterically occluded, class 3, class 4, and class 1/4 RBD epitopes (Fig. 1B). By contrast, homotypic SARS-2 RBD-mi3, a nanoparticle including 60 copies of a SARS-2 RBD (*34*, *36*), is more likely to engage B cells with receptors that recognize sterically accessible, but less conserved, class 1 and class 2 RBD epitopes (Fig. 1B). As previously observed (*34*), cross-reactivity against sarbecoviruses was more extensive in antisera from mosaic-8–immunized compared with homotypic SARS-2–immunized animals. In addition, although both mosaic-8–immunized and homotypic SARS-2–immunized animals were protected from SARS-2 challenge, only the mosaic-8–immunized animals were protected from SARS-1 challenge. Consistent with these results, polyclonal antibody epitope mapping by means of deep mutational scanning showed preferential binding to conserved RBD epitopes for mosaic-8 antisera, but binding to more variable epitopes for homotypic antisera. These results highlight the potential for a mosaic nanoparticle approach to elicit more broadly protective antibody responses than homotypic nanoparticle approaches.

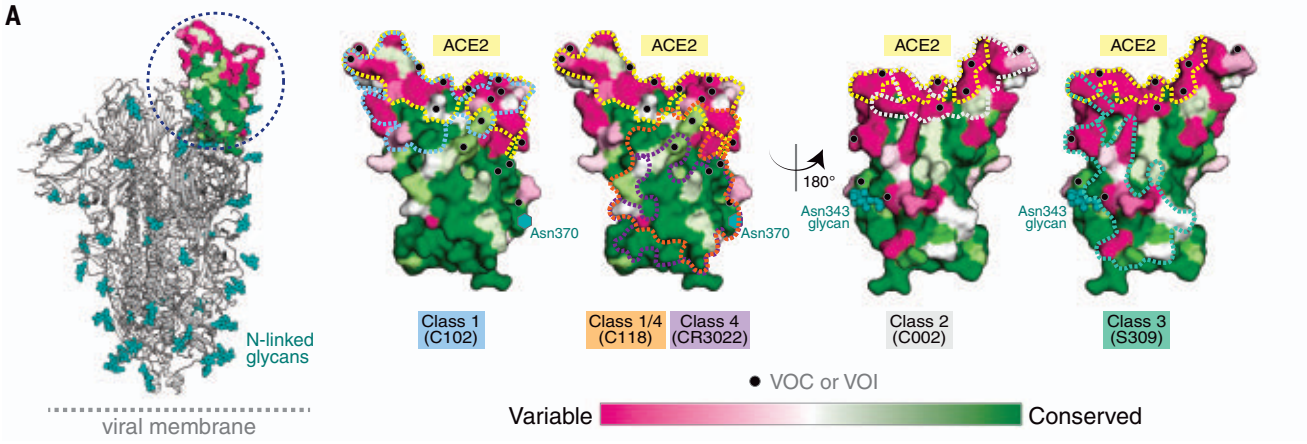
Results

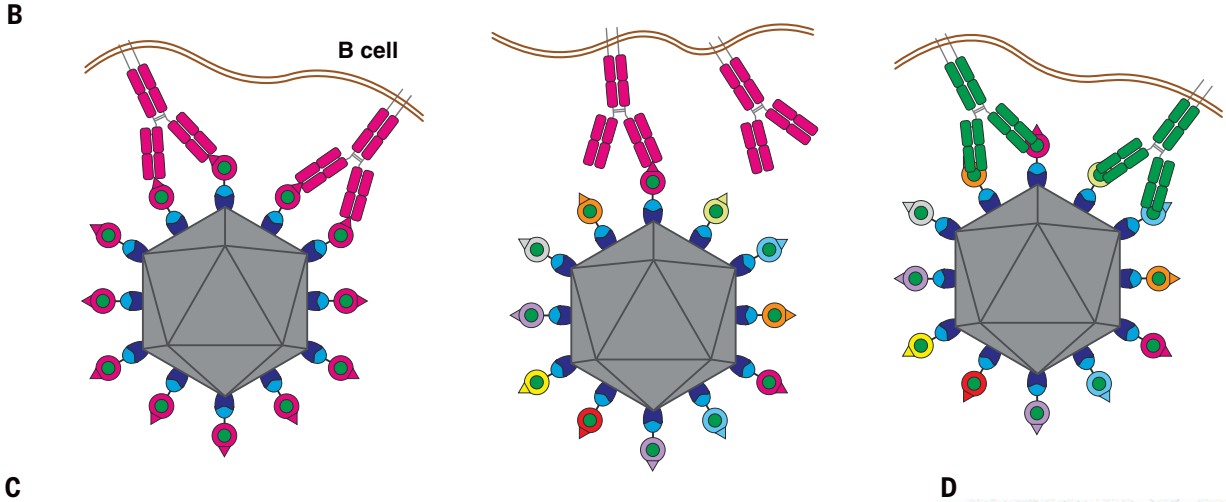
Mosaic-8b and homotypic SARS-2 Beta RBD nanoparticles were homogeneous and exhibited expected properties

We previously used SARS-2 WA1 RBD for making mosaic-8 and homotypic SARS-2 RBD nanoparticles (*34*). For the challenge studies described here, we switched to the SARS-2 Beta RBD (three RBD substitutions compared with WA1 RBD; https://viralzone.expasy.org/9556) to represent the most antigenically distinct variant at the time the experiments were initiated (fig. S1).

We used the SpyCatcher-SpyTag system (*37*, *38*) to covalently attach RBDs with C-terminal SpyTag003 sequences to a 60-mer nanoparticle (SpyCatcher003-mi3) (*39*) to make either mosaic-8b (each nanoparticle presenting the SARS-2 Beta RBD plus seven other sarbecovirus RBDs attached to the 60 sites)

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. [2]Laboratory of Virology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT 59840, USA. [3]Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [4]Department of Genome Sciences and Medical Scientist Training Program, University of Washington, Seattle, WA 98195, USA. [5]BIOQUAL, Rockville, MD, USA. [6]MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK. [7]Chinese Academy of Medical Sciences, Oxford Institute, University of Oxford, Oxford OX3 9DS, UK. [8]Basic Sciences Division and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. [9]Howard Hughes Medical Institute, Seattle, WA 98109, USA.
*Corresponding author. Email: bjorkman@caltech.edu
†These authors contributed equally to this work. ‡Present address: Department of Biology, Stanford University, Stanford, CA 94305, USA.





C

	Virus	Accession	Residue number	Clade	Host species
mosaic-8b (matched)	SARS-2 Beta	QUT64557.1.	319-539	1b	human
	RaTG13	QHR63300	319-541	1b	*Rhinolophus affinis* (bat)
	SHC014	KC881005	307-524	1a	*Rhinolophus sinicus* (bat)
	Rs4081	KY417143	310-515	2	*Rhinolophus sinicus* (bat)
	Pang17	QIA48632	317-539	1b	*Manis javanica* (pangolin)
	RmYN02	EPI_ISL_412977	298-503	2	*Rhinolophus malayanus* (bat)
	Rf1	DQ412042	310-515	2	*Rhinolophus ferrumequium* (bat)
	WIV1	KF367457	307-528	1a	*Rhinolophus sinicus* (bat)
Not included (mismatched)	SARS-1	AAP13441.1	318-510	1a	human
	LYRa3	AHX37569.1	318-511	1a	*Rhinolophus affinis* (bat)
	Yun11	JX993988	310-515	2	*Chaerephon plicata* (bat)
	BM-4831	NC014470	310-530	3	*Rhinolophus blasii* (bat)
	BtKY72	KY352407	309-530	3	*Rhinolophus sp.* (bat)
	Khosta-2	QVN46569.1	309-531	3	*Rhinolophus spp.* (bat)
	RshSTT200	EPI_ISL_852605	317-539	1b	*Rhinolophus shameli* (bat)
	SARS-2 WA1	MN985325.1	319-539	1b	human
	SARS-2 Delta	QWK65230.1	319-539	1b	human
	SARS-2 Omicron BA.1	UFO69278.1	319-539	1b	human
	SARS-2 Omicron BA.2	UFO69279.1	319-540	1b	human
	SARS-2 Omicron BA.2.12.1	UMZ92892.1	319-540	1b	human
	SARS-2 Omicron BA.4/BA.5	UPP14409.1	319-540	1b	human

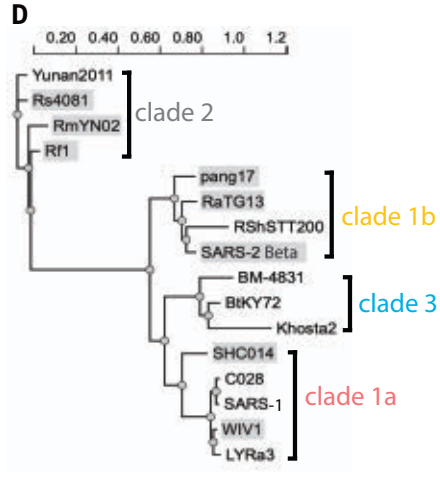


Fig. 1. Mosaic nanoparticles may preferentially induce cross-reactive antibodies through avidity effects. (**A**) (Left) Structure of SARS-2 S trimer (PDB ID: 6VYB) showing one "up" RBD (dashed circle). (Right) Sequence conservation of the 16 sarbecovirus RBDs in (D) calculated by use of the ConSurf Database (*89*) shown on four views of an RBD surface (PDB ID 7BZ5). The ACE2 binding footprint (PDB 6M0J) is outlined with a yellow dotted line. Locations of residues that are substituted in SARS-2 VOCs and VOIs as of March 2022 (https://viralzone.expasy.org/9556) are indicated as black dots. Class 1, 2, 3, 4, and 1/4 epitopes are outlined in different colored dotted lines by using information from structures of representative monoclonal antibodies bound to RBD or S trimer (C102, PDB 7K8M; C002, PDB 7K8T; S309, PDB 7JX3; CR3022, PDB 7LOP; and C118, PDB 7RKV). The *N*-linked glycan attached to RBD residue 343 is indicated with teal spheres, and the potential *N*-linked glycosylation site at position 370 in RBDs derived from sarbecoviruses other than SARS-2 is indicated with a teal circle. (**B**) Schematic showing hypothesis for how mosaic RBD nanoparticles could induce production of cross-reactive antibodies because B cells with receptors that bind to multiple antigens on a nanoparticle have a competitive advantage over B cells with receptors that likely bind with a single Fab to a one or very few antigens on a nanoparticle. (Left) Clustered membrane-bound B cell receptors (only two shown for clarity)

bind with avidity to a strain-specific epitope (dark pink triangle) on dark pink antigens attached to a homotypic particle. (Middle) B cell receptors cannot bind with avidity to a strain-specific epitope (triangle) on a dark pink antigen attached to a mosaic particle; thus, only one Fab from a single strain-specific B cell receptor is likely to bind to a mosaic particle. (Right) Clustered B cell receptors (only two shown for clarity) can bind with avidity to common epitope (green circle) presented on different antigens attached to a mosaic particle, but not to strain-specific epitopes (triangles). (**C**) Sarbecoviruses from which the RBDs in mosaic-8b RBD-mi3 were derived (matched) and sarbecoviruses from which RBDs were not included in mosaic-8b (mismatched). Clades are defined as in (*13*). The WA1 SARS-2 RBD was used in mosaic-8 g instead of the SARS-2 Beta RBD. (**D**) Phylogenetic tree of selected sarbecoviruses calculated by using PhyML 3.0 (*90*), based on amino acid sequences of RBDs aligned by using Clustal Omega (*91*). Viruses with RBDs included in mosaic-8b are indicated with gray rectangles.

(fig. S1) or homotypic (each nanoparticle presenting 60 copies of the SARS-2 Beta RBD) RBD-mi3 nanoparticles (fig. S2A). In addition to the SARS-2 Beta RBD, the other RBDs in mosaic-8b nanoparticles were chosen from clade 1a, 1b, and 2 sarbecoviruses [clades as defined in (*13*)] (Fig. 1, C and D, and figs. S1 and S2A). Immune responses against these strains in immunized animals were considered "matched" because each was represented by an RBD on mosaic-8b. Sarbecoviruses from clades 1, 2, and 3 and SARS-2 RBDs other than SARS-2 Beta that did not have RBDs represented on mosaic-8b nanoparticles were considered "mismatched" in immunological assays and challenge experiments. SARS-1 was chosen as a mismatched strain to allow challenge experiments and because it uses human ACE2 as its host receptor (*40*) and can therefore be evaluated in ACE2-dependent pseudotyped neutralization assays, although SARS-1 is closely related to WIV1, a clade 1a bat sarbecovirus represented on the nanoparticles. Two versions of mosaic-8 were used for experiments: mosaic-8b RBD-mi3 (SARS-2 Beta RBD and seven animal sarbecovirus RBDs) (Fig. 1, C and D) and mosaic-8 g (mosaic-8 with a WA1 SARS-2 RBD plus the seven zoonotic RBDs) (*34*), in which *N*-linked glycosylation site sequons at RBD position 484 were introduced in the clade 1a and 1b RBDs to occlude class 1 and 2 RBD epitopes (fig. S2A). RBD residue 484 was chosen for *N*-glycan addition because substitutions at this position in SARS-2 RBD prevented binding of some human class 1 and class 2 antibodies to RBD (*26*). Mosaic-8 and homotypic SARS-2 Beta nanoparticles were purified by means of size-exclusion chromatography (SEC) (fig. S2B) and validated with SDS–polyacrylamide gel electrophoresis (SDS-PAGE) to show near 100% conjugation efficiency (fig. S2C). Dynamic light scattering (DLS) and negative-stain electron microscopy (EM) demonstrated that conjugated nanoparticles were monodisperse and exhibited a defined diameter (fig. S2, D and E), and interactions of human ACE2 and monoclonal antibodies with known epitopes exhibited expected binding profiles (fig. S3).

RBD nanoparticles elicited binding and neutralizing antibody responses in K18-human ACE2 transgenic mice

To compare the efficacies of mosaic and homotypic RBD-mi3 nanoparticle immunizations, we evaluated immune responses and protection from viral challenge in K18-human ACE2 (K18-hACE2) transgenic mice (Figs. 2 and 3) (*41*). K18-hACE2 mice express human ACE2 driven by a cytokeratin promotor in epithelia—including airway epithelial cells, where SARS-2 infections often start—and recapitulate severe COVID-19 upon infection with SARS-2 (*41*–*43*). Viral challenges of K18-hACE2 mice result in extensive weight loss, and death usually results from SARS-2 or SARS-1 infection (*42*). We chose this lethal challenge model to evaluate the highest levels of potential protection, which might then be used to extrapolate to the expected efficacy of a vaccine in humans.

K18-hACE2 mice were primed with either mosaic-8b, mosaic-8 g, homotypic SARS-2 Beta, or unconjugated SpyCatcher-mi3 nanoparticles adjuvanted with AddaVax and boosted 4 weeks later (Fig. 2A). In these experiments, SARS-2 Beta represented a matched sarbecovirus for the mosaic-8b RBD-mi3 and homotypic SARS-2 RBD-mi3 immunogens but was mismatched for mosaic-8 g. SARS-1 was mismatched for all three nanoparticle immunogens.

We first evaluated serum antibody responses in binding and pseudovirus neutralization assays 14 days after boosting (Fig. 2, B to I). Serum enzyme-linked immunosorbent assays (ELISAs) were conducted to assess binding to the indicated RBDs and to the 6P stabilized version of the soluble WA1 S trimer (*44*). All three immunogens elicited high ELISA binding antibody titers, with titers against the WA1 RBD and S-6P modestly higher for homotypic SARS-2 Beta–immunized mice with respect to mosaic-8b– and mosaic-8 g–immunized mice (Fig. 2B). However, although the mosaic-8b and mosaic-8 g RBD-mi3 ELISA titers were not significantly different from homotypic RBD-mi3 antisera titers against the other SARS-2 variants (Fig. 2, C and D), ELISA titers were increased against sarbecovirus RBDs derived from viruses other than SARS-2 (Fig. 2, E and G to I)—for example, significantly higher binding titers when comparing mosaic-8 antisera versus the homotypic antisera against SARS-1 (mismatched) (Fig. 2E), SHC014 (matched for the two mosaic-8 immunogens) (Fig. 2G), BM48-31 (mismatched) (Fig. 2H), and Yun11 (mismatched) (Fig. 2I) (with the exception of mosaic-8 g and homotypic antisera in Fig. 2, F and H). As previously reported for RBD-mi3 immunogens (*34*), the trends for serum RBD binding were generally predictive of pseudovirus neutralization; however, the homotypic SARS-2 Beta antisera showed significantly higher neutralization titers against all three SARS-2 variants than those of the mosaic-8b and mosaic-8 g antisera (Fig. 2, B to D). This contrasts with our previous reports of equivalent neutralization for antisera from mosaic-8 RBD-mi3– and homotypic SARS-2 RBD–immunized mice (*34*). In the earlier experiments, the SARS-2 RBD in both nanoparticles was derived from the WA1 strain rather than the Beta VOC, which could elicit increased levels of potent class 2 antibodies to RBD that bind RBDs with residue Glu^{484} (E484), which is substituted in most SARS-2 variants (*45*). Equivalent levels of antibodies to SARS-2 RBD binding, but lower neutralization potencies, in mosaic-8 versus homotypic RBD antisera are consistent with a larger portion of non-neutralizing antibodies against SARS-2 induced through mosaic-8 immunization. Non-neutralizing antibodies could be involved in protection because non-neutralizing antibodies have been shown to play a role in preventing severe COVID-19 from natural infection or challenges after vaccination (*46*, *47*). Despite their reduced neutralization potencies against SARS-2, the mosaic-8b and mosaic-8 g antisera showed significantly higher neutralization titers than the homotypic antisera against clade 1a viruses such as SARS-1 (mismatched) (Fig. 2E) and WIV1 and SHC014 (matched for mosaics) (Fig. 2, F and G), which is in agreement with earlier experiments (*34*).

Mosaic-8b, but not homotypic SARS-2 Beta, RBD-mi3 immunizations protect against matched and mismatched viral challenges

The four groups of immunized K18-hACE2 mice (n = 10) were challenged with SARS-2 Beta or SARS-1 (Fig. 2A). Four mice per group were euthanized at 4 days after challenge for viral load analysis, and the remaining six mice were monitored for survival up to 28 days after challenge. Mice in each cohort were evaluated for weight loss, survival, and levels of viral genomic and subgenomic RNA, and infectious virus in lung tissue and oropharyngeal swabs (Fig. 3). Control animals immunized with unconjugated mi3 showed rapid weight loss and death 4 to 6 days after SARS-2 or SARS-1 challenge. As evaluated by relative weight loss, the mosaic-8b and homotypic SARS-2 Beta RBD-mi3 nanoparticles were equally protective against SARS-2 challenge, showing minimal to

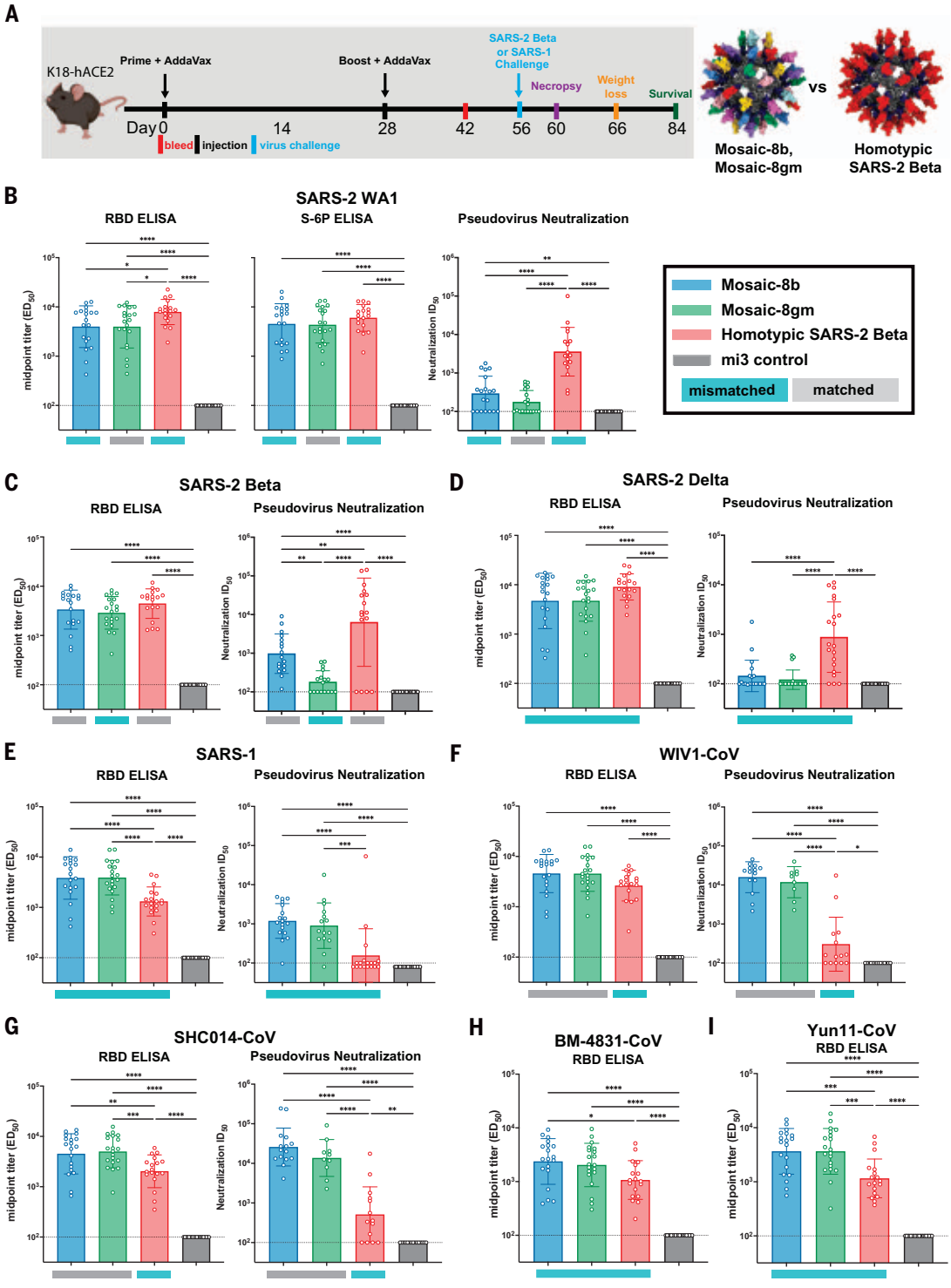


Fig. 2. Mosaic-8b and homotypic SARS-2 Beta RBD-mi3 immunizations induced binding and neutralizing antibodies in K18 mice. (**A**) (Left) Immunization schedule. K18-hACE2 mice were immunized with either 5 μg (RBD equivalents) mosaic-8b, mosaic-8 g, homotypic SARS-2 Beta, or the molar equivalent of unconjugated SpyCatcher003-mi3 nanoparticles. (Right) Structural models of mosaic-8 and homotypic RBD-mi3 nanoparticles constructed by using PDB 7SC1 (RBD), PDB 4MLI (SpyCatcher), and PDB 7B3Y (mi3). (**B** to **I**) ELISA and neutralization data from day 42 (14 days after boost) for antisera from individual mice (open circles) presented as the mean (bars) and standard deviation (horizontal lines). ELISA results are shown as midpoint titers [median effective dose (ED_{50}) values]; neutralization results are shown as half-maximal inhibitory dilutions (ID_{50} values). Dashed horizontal lines correspond to the background values representing the limit of detection. Significant differences between cohorts linked by horizontal lines are indicated with asterisks: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. Rectangles below ELISA and neutralization data indicate mismatched strains (cyan; the RBD from that strain was not present on the nanoparticle) or matched strains (gray; the RBD was present on the nanoparticle).

no weight loss, whereas some of the mosaic-8 g animals experienced transient weight loss but recovered by 10 days after challenge (Fig. 3A). By contrast, only the mosaic-8b and mosaic-8 g immunizations prevented weight loss after the SARS-1 challenge, whereas the homotypic SARS-CoV-2 Beta RBD-mi3 mice experienced similar weight loss as that of the mi3 control mice (Fig. 3A). The post-challenge survival results were consistent with weight loss: After SARS-2 Beta challenge, both the mosaic-8b and homotypic immunized animals showed complete survival (100%), whereas five of six (~83%) animals in the mosaic-8 g immunization group survived (Fig. 3B). After SARS-1 challenge, all but one of the mice in the homotypic SARS-2 Beta group reached endpoint criteria within 6 days (a delay compared with the control mi3 group, in which all animals reached endpoint criteria within 4 days), whereas all mice in the mosaic-8b and mosaic-8 g group survived the challenge during the 28 days of post-challenge monitoring (Fig. 3B).

Altogether, despite elicitation of lower neutralization titers against SARS-2 compared with homotypic Beta RBD-mi3 antisera (Fig. 2, B to D), immunization with mosaic-8b RBD-mi3

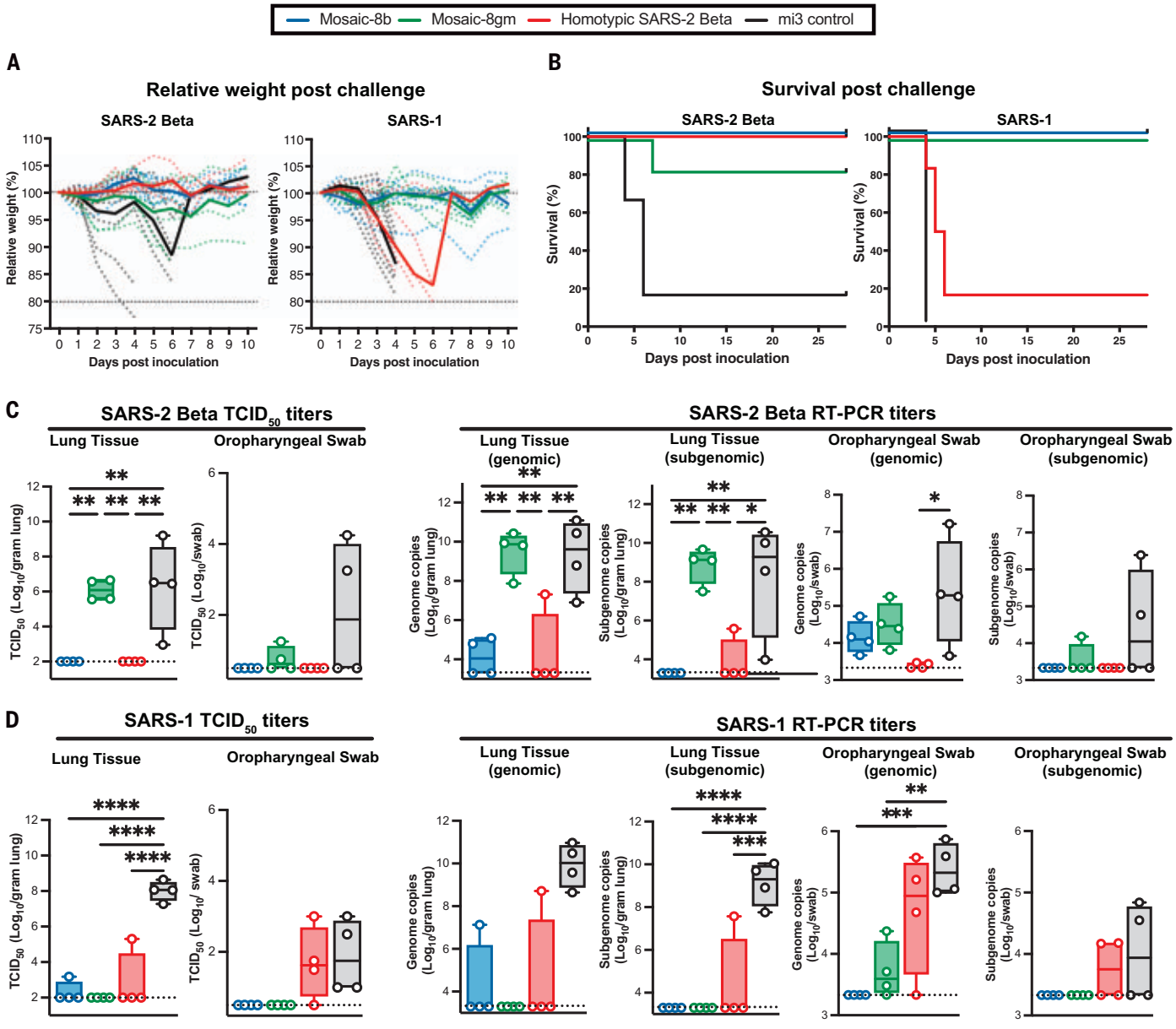


Fig. 3. Mosaic-8b immunization protected against SARS-2 and SARS-1 challenges in K18-hACE2 mice, whereas homotypic SARS-2 immunization protected only against SARS-2. Mice were immunized and boosted with the indicated mi3 nanoparticles (5 μg RBD equivalents for RBD nanoparticles; molar equivalent for unconjugated mi3 control). (**A**) Weight changes after SARS-2 Beta or SARS-1 challenge. Mean weight in each vaccinated cohort is indicated with a thick colored line. Weights of individual mice are indicated with colored dashed lines. (**B**) Survival after SARS-2 Beta or SARS-1 challenge. (**C**) (Left) SARS-2 Beta infectious titers after challenge in lung tissue and oropharyngeal swabs. (Right) Genomic and subgenomic SARS-2 Beta RNA copes determined with RT-PCR. (**D**) (Left) SARS-1 infectious titers after challenge in lung tissue and oropharyngeal swabs. (Right) Genomic and subgenomic SARS-1 RNA copies determined with RT-PCR. Significant differences between cohorts linked by horizontal lines are indicated with asterisks: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

was fully protective against both matched (SARS-2 Beta) and mismatched (SARS-1) challenges in the K18-hACE2 mouse model. By contrast, immunization with homotypic SARS-2 RBD-mi3 was protective against the matched SARS-2 Beta challenge but not against the mismatched SARS-1 challenge. Mosaic-8 g immunization protected against SARS-1 challenge but showed somewhat reduced efficacy against the SARS-2 Beta challenge, perhaps related to occluding the class 2 RBD epitope targeted by potent, but usually strain-specific, neutralizing antibodies.

Infectious virus levels and viral RNA copies correlate with protection in challenged K18-hACE2 mice

We measured levels of infectious virus and viral RNA in lung and oropharyngeal swab samples from challenged K18-hACE2 mice (n = 4), obtained at 4 days after challenge (Fig. 3, C and D). Infectious virus titers were measured by determining the median tissue culture infectious dose ($TCID_{50}$) of either oropharyngeal swab or lung tissue homogenate samples as described (*48*). In SARS-2 Beta–challenged mice, vaccination with either mosaic-8b or homotypic SARS-2 Beta completely suppressed viral replication in the lungs and oropharyngeal swabs, whereas levels of infectious virus in mosaic-8 g–immunized animals were equivalent to that in control-immunized mice in lungs. The SARS-2 Beta infectious viral

load was lower for mosaic-8 g–immunized mice with respect to control animals in the oropharyngeal swabs, suggesting partial protection in these animals (Fig. 3C, left). Almost all vaccinated animal groups displayed completely suppressed infectious SARS-1 in lungs compared with that of control-immunized mice (Fig. 3D, left). However, only mosaic-8b– and mosaic-8 g–immunized animals showed complete suppression of infectious SARS-1 in oropharyngeal swabs, whereas homotypic SARS-2 Beta–immunized animals showed infectious viral loads that were similar to those of control animals (Fig. 3D, left), possibly explaining the severity of SARS-1 in these animals (Fig. 3, A and B).

Viral RNA copies in the lung tissue and oropharyngeal swabs were measured with reverse transcription polymerase chain reaction (RT-PCR) by using both genomic and subgenomic primer sets. Genomic RT-PCR titers reflect the overall RNA copies in the sample, including both replicating virus in infected cells and viral particles and debris, whereas subgenomic RNA is a better surrogate for infectious viral titers because it is produced in infected cells and poorly packaged into virions (*49*). In lung tissue samples, genomic and subgenomic SARS-2 Beta viral RNA titers were lower in mosaic-8b–immunized and homotypic SARS-2 Beta–immunized animals compared with control-immunized animals (Fig. 3C, right), which is consistent with infectious virus titer measurements (Fig. 3C, left) and protection against SARS-2 Beta challenge in mosaic-8b–immunized and homotypic SARS-2 Beta–immunized animals (Fig. 3, A and B). SARS-2 Beta RNA copies in oropharyngeal swabs were low for all immunized cohorts compared with the control (Fig. 3C, right). Subgenomic SARS-1 viral RNA titers were completely suppressed in mosaic-8b– and mosaic-8 g–immunized animals in both lung tissue and oropharyngeal swabs with respect to the control (Fig. 3D, right), which also is consistent with infectious viral titers (Fig. 3C, left) and complete protection in these cohorts against SARS-1 challenge (Fig. 3, A and B). The lack of suppression of viral RNA and infectious viral loads in oropharyngeal swabs from homotypic SARS-2 Beta–immunized animals challenged with SARS-1 correlates with lethality from SARS-1 infection, possibly because of virus entry into the brain through nasal infection (*50*). Homotypic SARS-2–immunized animals showed significantly lower levels of subgenomic viral RNA in the lung (Fig. 3D) with respect to control animals, suggesting partial control of SARS-1 in these tissues.

The K18-hACE2 mouse experiments were designed to evaluate survival, weight loss, and reduction or absence of viral replication as the primary metrics for vaccine efficacy (*50*). However, we also obtained lung tissue 4 days after challenge for analysis by hematoxylin and eosin (H&E) staining and immunohistochemistry (IHC) (data S1 and fig. S4). For SARS-2 Beta infection, immunization with mosaic-8b and homotypic SARS-2 Beta nanoparticles showed minimal to no pathology as well as IHC staining for SARS-2 N protein with respect to control animals and mosaic-8 g–vaccinated animals (fig. S3, A to H, and Q), which is consistent with viral infectious titers and viral RNA titers in the lung (Fig. 2C). SARS-1 infection lung pathology for all animals was minimal for both vaccinated and control animals, and only control animals showed IHC staining for SARS-1 N (fig. S4, I to P and Q), which again is consistent with viral infectious titers and viral RNA titers in the lung (Fig. 2D).

Mosaic-8b RBD nanoparticles protect nonhuman primates from mismatched viral challenges

To extend these results to another animal model of SARS-2 and SARS-1 infection, we also conducted immunization and challenge studies in nonhuman primates (NHPs). Because of limitations of available NHPs, we evaluated only the most promising vaccine candidate, mosaic-8b RBD-mi3, comparing it with a non-immunized cohort for challenges with either SARS-2 Delta or with SARS-1. Both challenge viruses were mismatched because the mosaic-8b nanoparticles did not include RBDs from SARS-2 Delta or SARS-1.

Eight NHPs were immunized and boosted at day 28 with mosaic-8b RBD-mi3 adjuvanted with VAC20 [2% aluminum hydroxide wet gel (Al_2O_3)] (alum) and then boosted again at day 92 with mosaic-8b RBD-mi3 adjuvanted with EmulsiPan, an MF59-like squalene-based oil in water emulsion adjuvant (*51*). Four weeks after the second boost, half of the NHPs in the vaccinated and control groups were challenged with SARS-2 Delta, and the other half were challenged with SARS-1 (Fig. 4A).

Polyclonal antisera were evaluated with ELISA for binding to SARS-2 VOCs and evaluated for neutralization activity by use of pseudovirus and authentic virus neutralization assays (Fig. 4B). The RBD ELISA and neutralization results showed similar trends, with relatively weak binding and neutralization before the first boost, rising levels after the first boost that contracted by day 92, and then rising again and remaining above the 1:100 neutralization titers that correlate with ~90% vaccine efficiency (*52*) after the second boost. Binding antibody levels were similar for SARS-2 Beta (matched) and the mismatched WA1, Delta, and Omicron BA.1, BA.2, BA.2.12.1, and BA.4/BA.5 SARS-2 variant RBDs, which is predictive of neutralization across SARS-2 variants (*34*) as verified in available pseudovirus and authentic virus neutralization assays for SARS-2 variants (Fig. 4B). We also evaluated binding and neutralization against RBDs and pseudoviruses from other sarbecovirus lineages (clades 1a, 2, and 3), including matched (WIV1 and SHC014) and mismatched (SARS-1, LYRa3, RshSTT200, BM48-31, BtKY72, Khosta-2, and Yun11) viruses (Fig. 4, C and D). Similar trends were observed for binding and neutralization of non–SARS-2 sarbecoviruses as seen for the SARS-2 variants: All RBDs were recognized by polyclonal antisera in ELISAs and neutralized in pseudovirus assays for human ACE2 entry–dependent viral strains for which neutralization assays could be conducted, including mismatched strains: SARS-1; a mutant form of BtKY72 (K493Y/T498W) that uses ACE-2 for entry (*13*); and Khosta-2 and LYRa3 viruses, which contain chimeric spikes (RBDs from Khosta-2 or LYRa3, remaining S from SARS-1). In all cases, the contracting antibody responses at day 92 were restored by the second boost (Fig. 4, C and D).

Mosaic-8b RBD-mi3–immunized and control NHPs were challenged with either SARS-2 Delta or SARS-1 (both mismatched) 28 days after the second boost (Fig. 4A). Protection was assessed by measuring infectious virus titers (Fig. 5, A and B) and by viral RNA using RT-PCR (SARS-2 only) (Fig. 5C) in bronchial alveolar lavage (BAL) and nasal swabs 2 or 4 days after challenge. We observed no detectable SARS-2 Delta infectious virus in BAL at either time point, whereas BAL from three of four control animals showed infectious SARS-2 virus (Fig. 5A). Nasal swabs showed low levels of virus in vaccinated animals, but at about two log significantly reduced levels compared with those of control NHPs (Fig. 5A), which is consistent with reports of detectable virus replication in upper airways in animals that were protected from clinical disease (*53*). For SARS-1–challenged animals, mosaic-8b–immunized NHPs had no detectable viral titers in either BAL or nasal swabs, whereas all control animals had detectable viral titers (three of four in BAL and one of four in nasal swabs; individual animals identified by different colors) (Fig. 5B). Although the low sample numbers precluded statistical significance of BAL or nasal swab differences in immunized versus control challenged animals, the lack of detectable SARS-1 virus in mosaic-8b–immunized animals was suggestive of protection.

Viral RNA copies were measured with RT-PCR in nasal swabs, BAL, and lung tissue taken from mosaic-8b–immunized and control animals challenged with SARS-2 Delta (Fig. 3C). Both genomic and subgenomic RNA titers were significantly reduced in mosaic-8b–immunized animals in both BAL and nasal swabs (Fig. 3C), which is consistent with infectious virus titers (Fig. 3A). Genomic and subgenomic RNA copies in lung tissue were also reduced for mosaic-8b–immunized animals (Fig. 3C), which is consistent with the observed protection. Both vaccinated and unvaccinated animals challenged with SARS-2 Delta showed a lack of

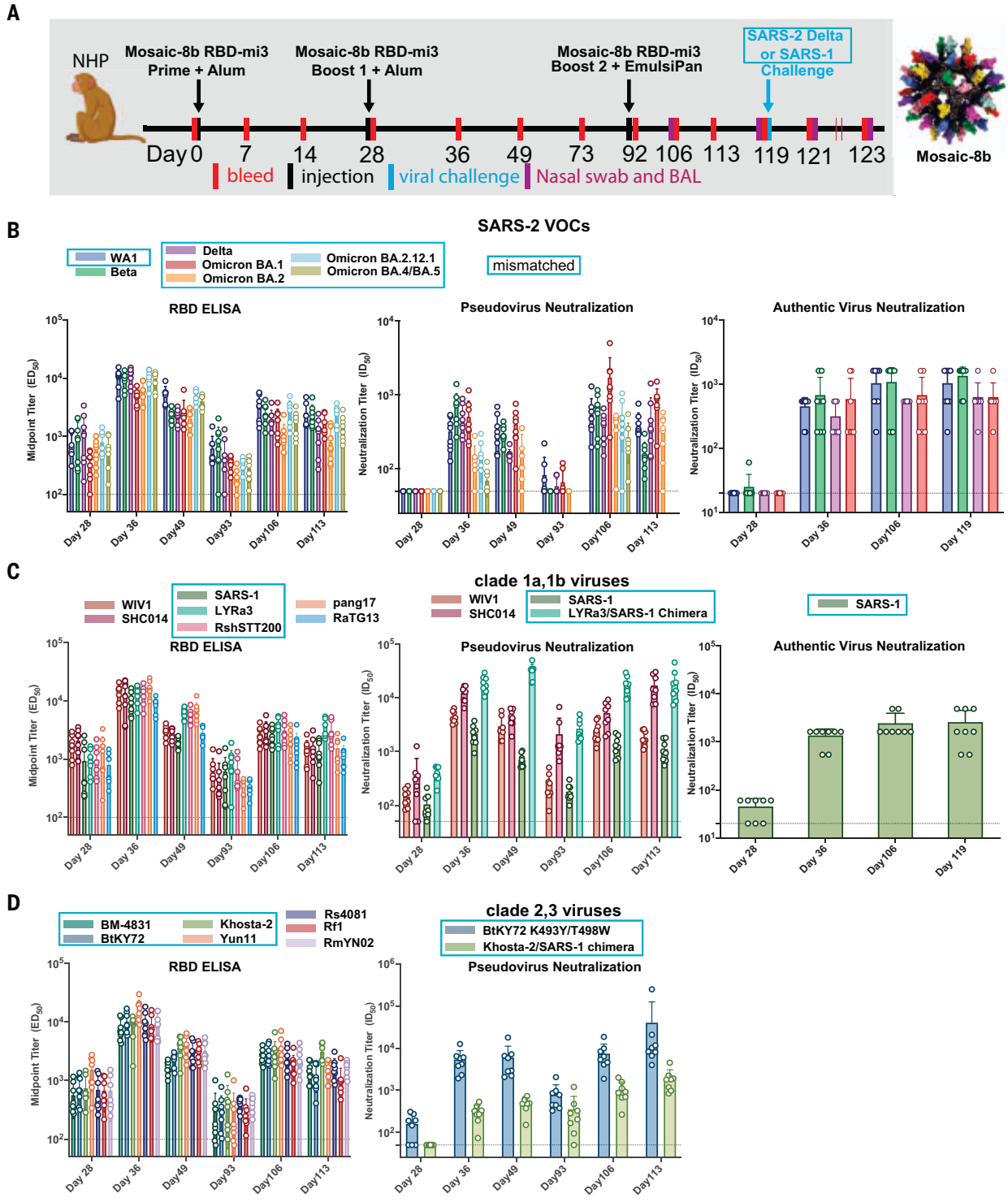


Fig. 4. Mosaic-8b RBD-mi3 immunization induced binding and neutralizing antibodies in NHPs. Mismatched RBDs or viruses with respect to mosaic-8b RBD-mi3 nanoparticles are indicated with cyan rectangular boxes. (**A**) (Left) Immunization schedule. NHPs were primed and boosted with 25 μg (RBD equivalents) mosaic-8b RBD-mi3 in alum and boosted again with 25 μg mosaic-8b RBD-mi3 in EmulsiPan. Eight immunized NHPs and eight unimmunized NHPs were then challenged with either SARS-2 Delta (four immunized and four unimmunized) or with SARS-1 (four immunized and four unimmunized). (Right) Structural model of mosaic-8b RBD-mi3 nanoparticles as shown in Fig. 2A. (**B** to **D**) RBDs or viruses for assays indicated as different colors. ELISA and neutralization data for antisera from individual NHPs (open circles) presented as the mean (bars) and standard deviation (horizontal lines). ELISA results are shown as midpoint titers (EC_{50} values); neutralization results are shown as half-maximal inhibitory dilutions (ID_{50} values). Dashed horizontal lines correspond to the background values representing the limit of detection.

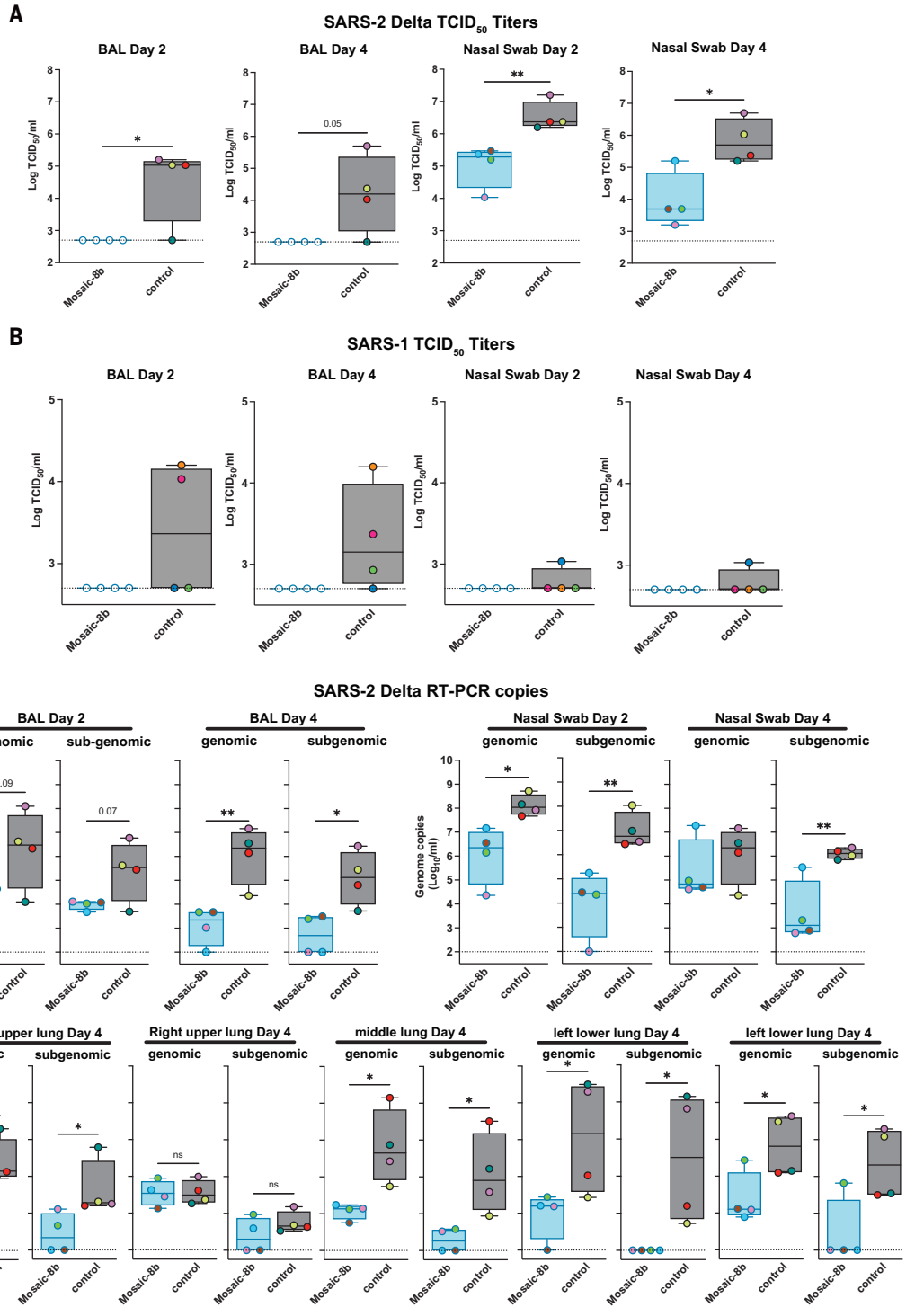


Fig. 5. Mosaic-8b immunization protected NHPs against SARS-2 Delta and SARS-1 challenges. NHPs were immunized with mosaic-8b RBD-mi3 or not immunized (control) before challenge. (**A**) SARS-2 Delta infectious titers after challenge in BAL (left) and nasal swabs (right). Individual animals are denoted with different colors. (**B**) SARS-1 infectious titers after challenge in BAL (left) and nasal swabs (right). Individual NHPs in the unvaccinated control group are denoted with different colors to show that all four animals exhibited signs of detectable SARS-1 infectious virus in BAL and/or nasal swabs. (**C**) (Top left) Genomic and subgenomic SARS-2 Beta RNA copies determined with RT-PCR in BAL day 2 and day 4 after infection. (Top right): Nasal swabs day 2 and day 4 after infection. (Bottom): Lung tissue day 4 after infection. Significant differences between cohorts linked by horizontal lines are indicated with asterisks: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

significant pathology in the lungs (fig. S5 and data S2) and no detectable differences in cytokine and chemokine levels in BAL (fig. S6).

Antisera to mosaic-8b target conserved RBD epitopes, whereas antisera to homotypic RBD-mi3 target variable epitopes

We next assessed whether mosaic and homotypic nanoparticles elicited different types of anti-RBD antibodies, as suggested by protection against matched challenge for both mosaic-8b and homotypic SARS-2 RBD-mi3 cohorts but protection against mismatched challenge only for the mosaic-8b RBD-mi3 cohort. Immunizations with either mosaic-8b or homotypic SARS-2 Beta nanoparticles adjuvanted with AddaVax were conducted in BALB/c mice (prime and boost 3 weeks later) (fig. S7A). Antibodies to RBD in serum 4 weeks after boost evaluated with ELISA and neutralization (fig. S7, B to G) exhibited similar characteristics as seen in the immunized K18-hACE2 mice (Fig. 2). To compare the characteristics of antibodies elicited by each type of RBD nanoparticle, we mapped SARS-2 Beta epitopes recognized by immunization-elicited antibodies to investigate whether mosaic-8b, but not homotypic SARS-2 Beta, preferentially elicited antibodies to RBD against conserved RBD epitopes as hypothesized (Fig. 1B). For

these experiments, we used yeast-display deep mutational scanning to map mutations in the SARS-2 Beta RBD (*54*, *55*) that escaped binding by antisera raised in BALB/c mice immunized with either mosaic-8b or homotypic SARS-2 Beta RBD-mi3 nanoparticles (Fig. 6 and fig. S8 to S10).

Mapping showed that mosaic-8b antisera from six immunized mice primarily targeted more conserved RBD epitopes, including class 4 (RBD residues 383 to 386) and class 3 (residue 357) epitopes (Fig. 6A and fig. S9), whereas homotypic SARS-2 antisera primarily targeted variable RBD epitopes, particularly class 2 (residue K484) (Fig. 6B and fig. S9). These results confirmed that mosaic-8b RBD-mi3 elicited antibodies against the conserved class 3 and class 4 epitopes, as designed in the mosaic RBD nanoparticle approach (Fig. 1, A and B). By contrast, homotypic SARS-2 RBD-mi3 primarily elicited antibodies against the more variable class 2 epitope (characterized by RBD residue 484) that varies between sarbecoviruses and in SARS-2 VOCs (Fig. 1A).

We also mapped mutations that reduce binding of four serum samples from NHPs vaccinated with three doses of mosaic-8b RBD-mi3 (day 106) (Fig. 4). The NHP antibody-escape

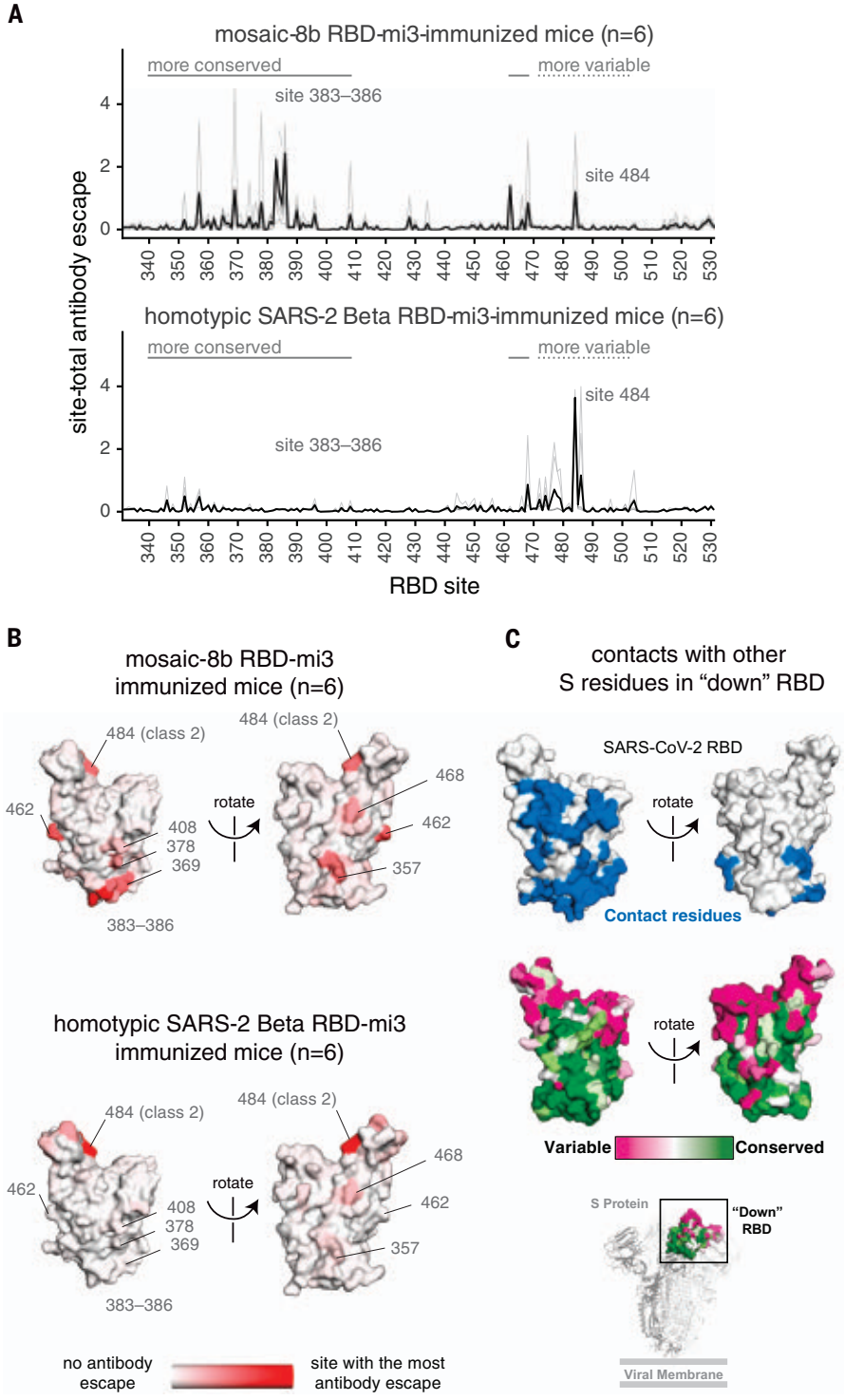


Fig. 6. Antibodies elicited by mosaic-8b immunization map to conserved RBD epitopes, as compared with antibodies elicited by homotypic SARS-2 Beta immunization. (**A**) Deep mutational scanning was used to identify mutations that reduced binding of sera from BALB/c mice immunized with mosaic-8b RBD-mi3 (top) or homotypic SARS-2 Beta RBD-mi3 (bottom) to the SARS-2 Beta RBD. The *y* axis shows the site-total antibody escape (sum of the antibody escape of all mutations at a site), with larger numbers indicating more antibody escape. Each light gray line indicates one antiserum, and the heavy black lines indicate the average across the $n = 6$ sera per group. RBD sites 340 to 408 and 462 to 468, which include the more conserved class 3/4 epitopes, are indicated with solid gray lines, and sites 472 to 503, which include sites from the more variable class 1/2 epitopes, are indicated with dashed lines. The "conserved" and "variable" epitopes presented here were generalized for simple visualization and are not identical to more specific epitope-class definitions (*26*, *60*). The highly variable RBD class 2 site 484 that is immunodominant among humans infected with SARS-2 (*45*, *60*) and the subdominant class 4 sites 383 to 386 are labeled. (**B**) The average site-total antibody escape for mice immunized with mosaic-8b RBD-mi3 (top) or homotypic SARS-2 RBD-mi3 (bottom) mapped to the surface of the SARS-2 Beta RBD (PDB 7LYQ), with white indicating no escape and red indicating sites with the most escape. Key sites are labeled, all of which are class 3/4 sites, except for the class 2 484 site. Interactive logo plots and structure-based visualizations of the antibody-escape maps are at https://jbloomlab.github.io/SARS-CoV-2-RBD_Beta_mosaic_np_vaccine. Individual antibody-escape maps are in fig. S9; raw data are in data S3 and at https://github.com/jbloomlab/SARS-CoV-2-RBD_Beta_mosaic_np_vaccine/blob/main/results/supp_data/all_raw_data.csv. (**C**) (Top) Residues in a "down" RBD that contact other regions of spike shown in blue on an RBD surface (PDB 7BZ5). Interacting residues were identified by using the PDBePISA software server (https://www.ebi.ac.uk/pdbe/prot_int/pistart.html) and the RBD from chain A of the spike trimer structure in PDB 7M6E. (Middle) Variable to conserved sarbecovirus sequence gradient (dark pink, variable; green, conserved) shown on RBD surface as in Fig. 1A. (Bottom) Structure of SARS-2 S trimer (PDB 6VYB) showing "down" RBD (boxed) colored with the variable to conserved sarbecovirus sequence gradient.

profiles were relatively broad, suggesting that no single mutation had a disproportionately large effect on binding (fig. S10). The NHP sera showed some skewing toward class 4 RBD epitopes and slight targeting of K484 (class 2) and T500 (class 3) (fig. S10). Differences in antibody-escape profiles for mosaic-8b–immunized mice and NHPs could be related to species differences (*56*) and/or different immunization regimens (Figs. 2A and 4A). The broad escape profiles from NHP antisera may suggest either antibody binding to a broad set of RBD epitopes and/or a population of affinity-matured antibodies that are less affected by single-point mutations. Nevertheless, the antibody-escape mapping results for mice and NHPs immunized with mosaic-8b RBD-mi3 are consistent with the hypothesis that the mosaic-8b nanoparticles elicit antibodies that target conserved RBD epitopes (Fig. 1B).

Mapping of sera from mosaic-8b–immunized mice and NHPs demonstrated relatively low, but nonzero, targeting of variable epitopes—for example, typified by RBD residue 484. Immunization with mosaic-8 g RBD-mi3 nanoparticles, in which class 1 and 2 epitopes in clade 1a and 1b RBDs were likely at least partially occluded by *N*-glycosylation at residue 484, was less protective against SARS-2 challenge than immunization with mosaic-8b, in which these epitopes were intact (Fig. 3). This result implies that retaining at least a subset of antibodies that target the immunodominant class 1 and 2 epitopes may be important for protection against SARS-2 challenge. Strategies to occlude these epitopes by introducing *N*-glycans (*57*, *58*) may thus impede optimal protection against SARS-2.

Discussion

Antibodies to RBD raised by infection and vaccination can potently neutralize SARS-2 through blocking S trimer binding to the ACE2 receptor required for viral entry (*16*–*20*, *23*–*29*). Although neutralizing antibodies recognize multiple RBD epitopes (*26*, *59*), immunoglobulins G (IgGs) in human polyclonal plasmas tend to target the class 1 and class 2 epitopes that are undergoing rapid evolution in SARS-2 and that vary between zoonotic and human sarbecoviruses (*28*, *45*, *60*). Some monoclonal antibodies against these epitopes maintain breadth (*61*) but more commonly show partial or complete loss of potency against SARS-2 VOCs and only rarely cross-react with animal sarbecoviruses (*45*, *62*, *63*). By contrast, although less common and usually less potent than antibodies against class 1 and class 2 anti-RBD epitopes, antibodies against the more conserved class 3, 4, and 1/4 epitopes exhibit increased cross-reactivity across sarbecoviruses and SARS-2 VOCs (*27*, *31*–*33*, *64*, *65*). Therefore, a vaccine that elicits such antibodies could serve to protect against SARS-2, its variants, and emerging zoonotic sarbecoviruses without the need for updating in the event of new VOCs and/or another sarbecovirus epidemic or pandemic.

Homotypic SARS-2 RBD or S trimer nanoparticles elicit potent neutralizing antibody responses that exhibit some degree of cross-reactivity across SARS-2 variants and sarbecoviruses (*34*, *36*, *66*–*75*). We reproduced results for homotypic nanoparticles and extended them in challenge studies comparing protection against SARS-2 and SARS-1 conferred by a mosaic-8 RBD nanoparticle versus a homotypic SARS-2 RBD nanoparticle. We demonstrated protection from SARS-2 challenge in animals immunized with homotypic SARS-2 RBD-mi3 and with mosaic-8b RBD-mi3 (challenges for the mosaic-8b included both a matched SARS-2 variant and a mismatched variant), despite mosaic-8b containing one-eighth as many SARS-2 RBDs as that of its homotypic SARS-2 counterpart. These results suggest that a mosaic RBD nanoparticle could be used as a COVID-19 vaccine option to protect from current and future SARS-2 variants. In addition, mosaic-8b, but not homotypic SARS-2 RBD-mi3 nanoparticles, protected K18-hACE2 mice against lethality with a mismatched SARS-1 challenge, suggesting that a mosaic nanoparticle vaccine could also protect from disease caused by future mismatched zoonotic sarbecoviruses.

Regarding the potential for mismatched protection conferred by immunization with the homotypic SARS-2 RBD-mi3 nanoparticles, although the homotypic nanoparticles did not confer complete protection against SARS-1 in K18-hACE2 mice, viral loads in lung tissue obtained from immunized animals were reduced compared with those from control animals. Thus, only one of four animals had detectable viral loads in the lungs compared with four of four in the control group, demonstrating that some level of protection was achieved. A similar outcome was reported upon vaccination of aged mice with RBD-scNP, a homotypic SARS-2 RBD–conjugated ferritin nanoparticle that induced neutralizing antibodies against SARS-2 and preemergent sarbecoviruses: Upon subsequent challenge with mouse-adapted SARS-1, viral loads were significantly reduced, but not absent, in lung tissue (four of five vaccinated animals exhibited reduced but detectable SARS-1 virus, as compared with five of five with higher viral loads in the control group) (*75*).

The presence of immunodominant epitopes on viral antigens have contributed to preventing the development of a universal influenza vaccine and vaccines against other antigenically variable viruses, such as HIV-1 and hepatitis C (*76*). Presentation of related, but antigenically different, viral antigens on a mosaic nanoparticle, a method to subvert immunodominance (*77*), is appropriate for making a pan-sarbecovirus vaccine because the SARS-2 S trimer contains immunodominant epitopes within its RBD that limit the breadth of antibodies elicited by SARS-2 infection or vaccination. Another potential advantage of a mosaic approach is that presentation of conserved T cell epitope peptides across the RBDs present on a mosaic nanoparticle could facilitate increased protection from disease that results from infection with a mismatched sarbecovirus. Our demonstrations that a mosaic RBD protected against challenges from both matched and mismatched sarbecoviruses, as compared with homotypic SARS-2 RBD nanoparticles that protected fully only against a matched challenge, are consistent with RBD mapping experiments showing that mosaic-8b but not homotypic SARS-2 RBD-mi3 nanoparticles primarily elicited antibodies against conserved RBD regions rather than the immunodominant class 1 and class 2 RBD epitopes. By including eight different RBD antigens arranged randomly on a 60-mer nanoparticle, as compared with a smaller number of different RBDs that are not arranged randomly (*78*), the chances of stimulating production of cross-reactive antibodies against conserved regions is maximized because adjacent antigens are unlikely to be the same (*77*). We recently described isolation of cross RBD-binding monoclonal antibodies from mosaic-8–immunized mice that included potent neutralizing antibodies against class 1/4 and class 3 RBD epitopes, thus also demonstrating targeting of conserved epitopes at the level of individual antibodies (*79*).

The plug-and-display approach facilitated by SpyCatcher-SpyTag methodology (*37*, *80*) allows the straightforward production of mosaic nanoparticles with different RBDs attached randomly. Such nanoparticles could be used to protect against COVID-19 and future sarbecovirus spillovers and easily adapted to make other pan-coronavirus vaccines—for example, against Middle East respiratory sydrome (MERS)–like betacoronaviruses and/or against alpha- or deltacoronaviruses. Given the recent plethora of SARS-2 VOCs and VOIs that may be arising at least in part because of antibody pressure, a relevant concern is whether more conserved RBD epitopes might be subject to substitutions that would render ineffective vaccines and/or monoclonal antibodies that target these regions. Although direct proof remains to be established, this scenario seems unlikely because RBD regions conserved between sarbecoviruses and SARS-2 variants are generally involved in contacts with other regions of spike trimer (Fig. 6C) and therefore less likely to tolerate selection-induced substitutions.

Materials and methods

Protein expression

Mammalian expression vectors encoding the RBDs of SARS-2 Beta (GenBank QUT64557.1),

SARS-2 WA1 (GenBank MN985325.1), SARS-2 Delta (GenBank QWK65230.1), SARS-2 Omicron BA.1 (GenBank UFO69278.1), SARS-2 Omicron BA.2 (GenBank UFO69279.1), SARS-2 Omicron BA.2.12.1 (GenBank UMZ92892.1), SARS-2 Omicron BA.4/BA.5 1 (GenBank UPP14409.1), RaTG13-CoV (GenBank QHR63300), SHC014-CoV (GenBank KC881005), Rs4081-CoV (GenBank KY417143), pangolin17-CoV (GenBank QIA48632), RmYN02-CoV (GSAID EPI_ISL_412977), Rf1-CoV (GenBank DQ412042), W1V1-CoV (GenBank KF367457), Yun11-CoV (GenBank JX993988), BM-4831-CoV (GenBank NC014470), BtkY72-CoV (GenBank KY352407), Khosta-2-CoV (QVN46569.1), RsSTT200-CoV (EPI_ISL_852605), LYRa3-CoV (AHX37569.1) and SARS-1 (GenBank AAP13441.1) with an N-terminal human IL-2 or Mu phosphatase signal peptide were constructed as previously described (*34*, *81*). 5 of the 8 RBD genes (SARS-2 WA1, RaTG13, pang17, WIV1, and SHC014) used to make mosaic-8 g were altered by site-directed mutagenesis to include a potential *N*-linked glycosylation site (PNGS) (N at position 484 and T at position 486). Each RBD was expressed to include a C-terminal hexahistidine tag (G-HHHHHH) and SpyTag003 (RGVPHIVMVDAYKRYK) (*39*) (for coupling to SpyCatcher003-mi3) or only a 15-residue Avi-tag (GLNDIFEAQKIEWHE) followed by a 6xHis tag (for ELISAs). RBDs were purified from transiently-transfected Expi293F cell (Gibco) supernatants by Ni-NTA and SEC as described (*81*), and RBDs with an introduced PNGS used for making mosaic-8 g RBD-mi3 were compared to their counterpart RBDs by SDS-PAGE to verify addition of extra *N*-glycans. SEC RBD fractions identified by SDS-PAGE were pooled and stored at 4°C or frozen in liquid nitrogen and stored at −80°C for longer term storage. A soluble SARS-2 trimer with 6P stabilizing mutations (*44*) was expressed and purified as described (*26*). Monoclonal human IgGs and human ACE-2 fused to human IgG Fc (hACE2-Fc) was expressed and purified as described (*26*, *32*).

Preparation of RBD-mi3 nanoparticles

SpyCatcher003-mi3 nanoparticles (*80*) were expressed in BL21 (DE3)-RIPL *Escherichia coli* (Agilent) transformed with the pET28a His6-SpyCatcher003-mi3 gene (Addgene) as described (*34*, *82*). Briefly, transformed bacterial cell pellets were lysed in the presence of 2.0 mM PMSF (Sigma). Lysates were spun at 21,000 × g for 30 min, filtered with a 0.2 μm filter, and mi3 particles were isolated by Ni-NTA chromatography using a HisTrap™ HP column (GE Healthcare). Eluted particles were concentrated using an Amicon Ultra 15 ml 30K concentrator (MilliporeSigma) and SEC purified using a HiLoad® 16/600 Superdex® 200 (GE Healthcare) column equilibrated with 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.02% NaN_3 (TBS). SpyCatcher003-mi3 particles were stored at 4°C for up to 1 month and used for conjugations after 0.2 μm filtering or spinning at 21,000 × g for 10 min.

Purified SpyCatcher003-mi3 nanoparticles were incubated with a 2-fold molar excess (RBD to mi3 subunit) of SpyTagged RBD (either a single RBD for homotypic SARS-2 RBD particles or an equimolar mixture of eight RBDs for mosaic particles) overnight at room temperature in TBS. The nanoparticles included the following RBDs: SARS-2 Beta (homotypic RBD-mi3); SARS-2 Beta, RaTG13, SHC014, Rs4081, RmYN02, pang17, Rf1, and WIV1 (mosaic-8b RBD-mi3); and N-glycan modified versions of the clade 1a and 1b RBDs (SARS-2 WA1, RaTG13, SHC014, pang17, and WIV1) together with unmodified Rs4081, RmYN02, and Rf1 RBDs (mosaic-8 g RBD-mi3). For mosaic-8 mi3 nanoparticles, equivalent conjugation of each of the eight SpyTagged RBDs was verified as described by SEC and SDS-PAGE analysis of conjugations to make homotypic nanoparticles (*34*).

Conjugated RBD-mi3 particles were separated from free RBDs by SEC on a Superose 6 10/300 column (GE Healthcare) equilibrated with PBS (20 mM sodium phosphate pH 7.5, 150 mM NaCl) and fractions corresponding to conjugated RBD-mi3 and free RBD were identified by SDS-PAGE. Concentrations of conjugated mi3 particles were determined using a Bio-Rad Protein Assay and are reported based on RBD content.

RBD-mi3 nanoparticles were evaluated for binding to a human ACE2-Fc construct (*32*) and to human monoclonal antibodies that recognize known RBD epitopes by ELISA. Duplicate samples of 20 μl of a 2.5 μg/ml solution of a purified RBD-mi3 nanoparticle in 0.1 M $NaHCO_3$ pH 9.8 was coated onto Nunc® MaxiSorp 384-well plates (Sigma) and incubated overnight at 4°C. After blocking with 3% bovine serum albumin (BSA) in TBS containing 0.1% Tween 20 (TBS-T) for 1 hour at room temperature, plates were washed with TBS-T, and purified hACE2-Fc or human IgG (50 μg/ml with 8 4-fold serial dilutions in TBS-T/3% BSA) was added to plates for 3 hours at room temperature. Plates were then washed again for 1 hour at room temperature, and a 1:100,000 dilution of secondary HRP-conjugated goat anti-human IgG (Abcam) was added. SuperSignal ELISA Femto Maximum Sensitivity Substrate (ThermoFisher) was added to plates following manufacturer instructions, and plates were read at 425 nm. For the homotypic SARS-2 Beta ELISA shown in fig. S3F, 50 μl of a 2.5 μg/ml solution of a purified RBD-mi3 nanoparticle in 0.1 M $NaHCO_3$ pH 9.8 was coated onto Corning® 96 well plates (Sigma) and incubated overnight at 4°C. After blocking with 3% bovine serum albumin (BSA) in TBS containing 0.1% Tween 20 (TBS-T) for 1 hour at room temperature, plates were washed with TBS-T, and purified hACE2-Fc or human IgG (50 μg/ml with 8 serial dilutions 4-fold in TBS-T/3% BSA) was added to plates for 3 hours at room temperature. Plates were then washed again for 1 hour at room temperature, and a 1:10,000 dilution of secondary HRP-conjugated goat anti-human IgG (Abcam) was added. 1-step Ultra TMB-ELISA (ThermoFisher) was added to plates following manufacturer instructions, and plates were read at 450 nm.

Prior to shipping for immunization and challenge studies, aliquots of conjugated RBD-mi3 nanoparticles were frozen in liquid nitrogen and then lyophilized (*36*) in PBS pH 7.4 using a Labconco CentriVap Benchtop Concentrator at −4°C. For immunizations, distilled water was added to rehydrate to a concentration of 1 mg/ml for a working stock, and the solution was gently pipetted and then spun at 20,000 × g for 10 min to remove aggregates.

DLS and EM characterizations of RBD-mi3 nanoparticles

DLS was used to determine the hydrodynamic radii of conjugated nanoparticles. Lyophilized nanoparticles were rehydrated as described above. Sample sizes of 100 μl were loaded into a disposable cuvette, and DLS measurements were performed on a DynaPro NanoStar (Wyatt Technology) using settings suggested by the manufacturer. A fit of the second order autocorrelation function to a globular protein model was used to derive the hydrodynamic radius and plotted on Graphpad Prism 9.3.1.

Mosaic-8b RBD-mi3 and homotypic SARS-2 RBD-mi3 were compared by negative-stain EM. Ultrathin, holey carbon-coated, 400 mesh Cu grids (Ted Pella, Inc.) were glow discharged (60 s at 15 mA), and a 3 μl aliquot of SEC-purified RBD-mi3 nanoparticles were diluted to ~40-100 μg/ml and applied to the grids for 60 s, Grids were then negatively stained with 2% (w/v) uranyl acetate for 30 s. Images were collected with a 120 keV FEI Tecnai T12 transmission electron microscope at 42,000× magnification.

K18-hACE2 mice

The Institutional Animal Care and Use Committee at Rocky Mountain Laboratories provided animal study approvals, which were conducted in an Association for Assessment and Accreditation of Laboratory Animal Care-accredited facility, following the basic principles and guidelines in the *Guide for the Care and Use of Laboratory Animals* eighth edition, the Animal Welfare Act, US Department of Agriculture, and the US Public Health Service Policy on Humane Care and Use of Laboratory Animals.

Animals were kept in climate-controlled rooms with a fixed light/dark cycle (12 hours/12 hours). Mice were cohoused in rodent cages,

fed a commercial rodent chow with ad libitum water, and monitored at least once daily. The Institutional Biosafety Committee (IBC)–approved work with infectious SARS-1 and SARS-2 viruses was conducted under biosafety level 3 (BSL3) conditions. All sample inactivation was performed according to IBC-approved standard operating procedures for removal of specimens from high containment.

Cells and virus for K18-hACE2 mouse studies

Virus propagation was performed in VeroE6 cells in DMEM containing 2% FBS, 1 mM L-glutamine, penicillin (50 U/ml), and streptomycin (50 μg/ml (DMEM2). The consensus sequence of the virus stock (SARS-CoV-2 Beta, isolate hCoV-19/USA/MD-HP01542/2021) used for these experiments was identical to the initial sequence deposited on GISAID (EPI_ISL_890360), and no contaminants or additional mutations were detected. VeroE6 cells were maintained in DMEM supplemented with 10% fetal bovine serum, 1 mM L-glutamine, penicillin (50 U/ml), and streptomycin (50 μg/ml). VeroE6 cells were provided by R. Baric (University of North Carolina at Chapel Hill). Mycoplasma testing is performed at monthly intervals, and no mycoplasma was detected.

Vaccination and infection of K18-hACE2 mice

K18-hACE2 mice (4 to 6 weeks old) were vaccinated with 2 × 50 μl of 5 μg (RBD equivalents)/(11.4 μg of total RBD-mi3) of RBD-mi3 or 5 μg unconjugated mi3 adjuvanted with Addavax 1:1 (1:1 v/v) intramuscularly at day 0 and day 28 (and challenged 28 days post the second immunization). Fourteen days before virus challenge, animals were bled via the submandibular vein. 10 animals per group were challenged with 30 μl of 10^5 $TCID_{50}$ SARS-2/human/USA/MD-HP01542/2021) or SARS-1 (Tor2) diluted in sterile Dulbecco's modified Eagle's medium (DMEM). Weight was recorded daily. Six mice per group were observed for survival up to 28 days post challenge or until they reached end-point criteria. End-point criteria were as follows: labored breathing or ambulatory difficulties or weight loss exceeding 20%. Four animals per group were euthanized on day 4 post challenge to collect oropharyngeal swabs and lung tissue for virology and histology analysis.

Virus titration after K18-hACE2 mouse challenge

Lung tissue sections were weighed and homogenized in 750 μl of DMEM. Virus titrations were performed by end point titration in VeroE6 cells expressing transmembrane protease serine 2 (TMPRSS-2) and human ACE2 (BEI resources, NR-54970), which were inoculated with 10-fold serial dilutions of virus swab medium or tissue homogenates in 96-well plates. When titrating tissue homogenate, cells were washed with PBS and 100 μl of DMEM2. Cells were incubated at 37°C and 5% CO_2, and cytopathic effect (CPE) was assessed 6 days later.

RNA extraction and quantitative RT-PCR

RNA was extracted from oropharyngeal swabs swabs using a QIAamp Viral RNA kit (Qiagen) according to the manufacturer's instructions. Tissue was homogenized and extracted using the RNeasy kit (Qiagen) according to the manufacturer's instructions. Viral gRNA- and sgRNA-specific assays (*48*) were used for the detection of viral RNA. The RT-PCR reaction (with 5 μl template viral RNA) was performed using the QuantStudio (Thermo Fisher Scientific) according to instructions of the manufacturer. Dilutions of SARS-2 with known genome copies were run in parallel to be used to generate the standard curves.

Histopathology and immunostaining of K18-hACE2 mouse samples

Lungs were collected upon necropsy on day 4 post challenge and perfused with 10% neutral-buffered formalin. Fixation was done for at least 7 days. Tissues were placed in cassettes and processed with a Sakura VIP-6 Tissue Tek on a 12-hour automated schedule using a graded series of ethanol, xylene, and PureAffin. Embedded tissues were sectioned at 5μm and dried overnight at 42°C prior to staining. Sections were stained with Harris hematoxylin (Cancer Diagnostics, no. SH3777), decolorized with 0.125% HCl/70% ethanol, blued in Pureview PH Blue (Cancer Diagnostics, no. 167020), counterstained with eosin 615 (Cancer Diagnostics, no. 16601), dehydrated, and mounted in Micromount (Leica, no. 3801731). An anti-SARS-2 nucleocapsid protein rabbit antiserum (generated by GenScript) was used at a 1:1000 dilution to detect specific anti–SARS-2 immunoreactivity using the Discovery ULTRA automated staining instrument (Roche Tissue Diagnostics) with a Discovery ChromoMap DAB (Ventana Medical Systems) kit. All slides were examined by a board-certified veterinary anatomic pathologist who was blinded to study group allocations. Scoring was done as follows. H&E; no lesions = 0; less than 1% = 0.5; minimal (1 to 10%) = 1; mild (11 to 25%) = 2; moderate (26 to 50%) = 3; marked (51 to 75%) = 4; severe (76 to 100%) = 5. IHC attachment; none = 0; less than 1% = 0.5; rare/few (1 to 10%) = 1; scattered (11 to 25%) = 2; moderate (26 to 50%) = 3; numerous (51 to 75%) = 4; diffuse (76 to 100%) = 5. Histopathology report is summarized in data S1.

BIOQUAL Ethics Statement and Animal Exposure

Rhesus macaques were housed and cared for at BIOQUAL, Inc., Rockville, MD. The study was performed under a BIOQUAL-approved IACUC protocol (no. 21-092P), in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the NIH, and in accordance with BIOQUAL standard operating procedures. BIOQUAL is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) and through OLAW, assurance number A-3086. All animal procedures were done under anesthesia to minimize pain and distress, in accordance with the recommendations of the Weatherall report "The use of non-human primates in research." Teklad 5038 primate diet was provided once daily according to macaque size and weight. The diet was supplemented daily with fresh fruit and vegetables. Fresh water was given ad libitum.

Vaccination of NHPs

The study included 16 rhesus macaques (Macaca mulatta), 8 of which were immunized with mosaic-8b RBD-mi3 (n = 8), and 8 of which served as unimmunized controls for SARS-2 and SARS-1 challenges. Four immunized and four unimmunized control NHPs were challenged with SARS-2, and four immunized and four unimmunized control NHPs were challenged with SARS-1. Due to a shortage of available NHPs, we could not compare mosaic-8b RBD-mi3 and homotypic SARS-2 Beta RBD-mi3 immunizations in this study. Macaques were 3 to 5 years old and ranged from 3.2 to 5.1 kg in body weight. Male and female macaques per group were balanced. Studies were performed unblinded. Macaques were evaluated by BIOQUAL veterinary staff before, during, and after immunizations.

NHPs were immunized intramuscularly with 25 μg (calculated based on RBDs; 56.8 μg of total RBD-mi3) of mosaic-8b RBD-mi3 adjuvanted with VAC20 (2% aluminum hydroxide wet gel, Al_2O_3) (alum) (Prime and Boost 1) (kind gift of Francis Laurent and Ruben Caputo, SPI Pharma) and subsequently with EmulsiPan adjuvant (an MF59-like, squalene-based oil in water emulsion adjuvant) (*51*) (kind gift of Harshet Jain, Panacea Biotec) for Boost 2. Each macaque received 0.5 ml into the right forelimb.

SARS-2 and SARS-1 intranasal and intratracheal NHP challenges

All macaques were challenged at week 11 (3 weeks after last vaccination) through combined intratracheal (1.0 ml) and intranasal (0.5 ml per nostril) inoculation with an infectious dose of 10^5 $TCID_{50}$ of SARS-2 B.1.617.2 (Delta, BEI NR-55612) or SARS-1 (Urbani). Virus was stored at −80 °C before use, thawed by hand and placed immediately on wet ice. Stock was diluted to 5×10^4 $TCID_{50}$ ml^{-1} in PBS and vortexed gently for 5 s before inoculation. Nasal swabs, BAL, plasma, and serum samples were collected 7 days before and 2 and 4 days after challenge. Protection from SARS-2 and SARS-1 infection was

determined by quantitative infectious viral load assay ($TCID_{50}$), and for SARS-2, also by RT-PCR of subgenomic N RNA (N sgRNA) as described above except that amplification was done use the Applied Biosystems 7500 Sequence detector.

$TCID_{50}$ and SARS-2 and SARS-1 virus $PRNT_{50}$ assays in NHP samples

$PRNT_{50}$ (50% plaque reduction neutralization test) assays for NHP samples were performed in a biosafety level 3 facility at BIOQUAL, Inc. (Rockville, MD). The $TCID_{50}$ assay was conducted by addition of 10-fold graded dilutions of samples to Vero/TMPRSS2 cell monolayers. Serial dilutions were performed in cell culture wells in quadruplicates. Positive (virus stock of known infectious titer in the assay) and negative (medium only) control wells were included in each assay set-up. The plates were incubated at 37°C, 5.0% CO_2 for 4 days. The cell monolayers were visually inspected for CPE, complete destruction of the monolayer. $TCID_{50}$ values was calculated using the Reed-Muench formula (*83*). For samples that had less than 3 CPE positive wells, the $TCID_{50}$ could not be calculated using the Reed-Muench formula, and these samples were assigned a titer of below the limit of detection (<2.7 log10 $TCID_{50}$/ml). For acceptable assay performance, the $TCID_{50}$ value of the positive control tested within 2-fold of the expected value.

To measure neutralization activity, sera from each NHP were diluted to 1:10 followed by a 3-fold serial dilution. Diluted samples were then incubated with ~30 plaque-forming units of wild-type SARS-2 USA-WA1/2020 (BEI NR-52281), B.1.351 (Beta, 501Y.V2.HV, NR-54974), or B.1.617.2 (Delta, BEI NR-55612) variants, in an equal volume of culture medium for 1 hour at 37°C. The serum-virus mixtures were added to a monolayer of confluent Vero E6 cells and incubated for one hour at 37°C in 5% CO_2. Each well was then overlaid with culture medium containing 0.5% methylcellulose and incubated for 3 days at 37°C in 5% CO_2. The plates were then fixed with methanol at −20°C for 30 min and stained with 0.2% crystal violet for 30 min at room temperature. $PRNT_{50}$ were estimated by determining the dilution at which plaques were reduced by 50% with respect to viral control.

Histopathology of SARS-2 infected NHPs

At necropsy, lung was collected and placed in 10% neutral buffered formalin for histopathologic analysis. Tissue sections were trimmed and processed for H&E-stained slides and examined by a board-certified pathologist at Experimental Pathology Laboratories (EPL) in Sterling, Virginia. Histopathologic findings are presented in the Individual Data Listing of Histopathology tables. Findings were graded from 1 to 5, depending upon severity, and summarized by treatment group in the Incidence Summary of Microscopic Findings by Sacrifice. An explanation of descriptive severity grades is provided at the end of the tables; equivalent numbered grades are 1 = minimal, 2 = mild, 3 = moderate, 4 = marked, 5 = severe.

Luminex chemokine and cytokine assays

Samples were analyzed using a Cytokine/Chemokine/Growth Factor 37-Plex NHP ProcartaPlex Panel (EPX370-40045-901). Target list [bead region] were as follows:

T helper: GM-CSF [44], IFN gamma [43], IL-1 beta [18], IL-2 [19], IL-4 [20], IL-5 [21], IL-6 [25], IL-8 (CXCL8) [27], IL-10 [28], IL-12p70 [34], IL-13 [35], IL-17A (CTLA-8) [36], IL-18 [66], IL-23 [63], TNF alpha [45]

Cytokines: CD40L [74], G-CSF (CSF-3) [42], IFN alpha [48], IL-1RA [38], IL-7 [26], IL-15 [65]

Chemokines: BLC (CXCL13) [15], Eotaxin (CCL11) [33], IP-10 (CXCL10) [22], I-TAC (CXCL11) [46], MCP-1 (CCL2) [51], MIG (CXCL9) [54], MIP-1 alpha (CCL3) [12], MIP-1 beta (CCL4) [47], SDF-1 alpha [13]

Growth Factors: BDNF [57], FGF-2 [75], NGF beta [55], PDGF-BB [77], SCF [39], VEGF-A [78], VEGF-D [53]

BAL samples were thawed at room temperature, quickly vortexed for 10s, and then centrifuged to pellet and remove debris. Samples were analyzed in duplicate using the above panel according to the manufacturer's protocol. Briefly, 200 µl of bead mix solution was added to all wells and washed using a Bio-Plex Pro Wash Station (BIORAD). 25 µl of Universal Assay Buffer was added to all wells, and then 25µl of samples/standards/blanks (blanks utilized Universal Assay Buffer) was added to the relevant wells, after a 2hr incubation on an orbital shaker (400 to 500 rpm) at room temperature. Plates were then washed on the magnetic plate washer, and 25 µl of prepared detection antibody was added to all wells following a 30m incubation, with agitation at room temperature. Plates were washed again, and 50 µl of provided Streptavidin-PE solution was added to all wells following a 30m incubation with agitation at room temperature. Once the incubation was complete, the plates were again washed using the magnetic plate washer. Inactivation was then performed overnight using 100 µl per well of 10% Formalin solution (VWR). Plates were run on a Bio-Plex 200 System (BIORAD), which was preprogrammed according to kit specifications for optimal signal detection. Values out of range (OOR<) are designated with the lowest standard cut off point, based on manufacturer's standard concentrations.

Mouse and NHP serum ELISAs

20 µl of a 2.5 µg/ml solution of an affinity purified His-tagged RBD in 0.1 M $NaHCO_3$ pH 9.8 was coated onto Nunc MaxiSorp 384-well plates (Sigma) and incubated overnight at 4°C. After blocking with 3% bovine serum albumin (BSA) in TBS containing 0.1% Tween 20 (TBS-T) for 1 hour at room temperature, plates were washed with TBS-T, and mouse or NHP serum diluted 1:100 and then serially diluted by 4-fold with TBS-T/3% BSA was added to the plates for 3 hours at room temperature. Plates were then washed again for 1 hour at room temperature, and a 1:50,000 dilution of secondary HRP-conjugated goat anti-mouse IgG (Abcam) was added. SuperSignal ELISA Femto Maximum Sensitivity Substrate (ThermoFisher) was added to plates following manufacturer's instructions, and plates were read at 425 nm. Curves were plotted and analyzed to obtain midpoint titers (EC_{50} values) using Graphpad Prism 9.3.1 (Graphpad Software, San Diego, CA) assuming a one-site binding model with a Hill coefficient. Titer differences were evaluated for statistical significance between groups using analysis of variance (ANOVA) test followed by Tukey's multiple comparison post hoc test calculated using Graphpad Prism 9.3.1.

Mouse and NHP serum pseudovirus neutralization assays

Lentiviral-based SARS-2 variants (WA1, Beta, Delta, Omicron BA.1, BA2, BA.2.12.2, and BA.4/BA.5), SARS-1, WIV1, SHC014, and BtKY72 K493Y/T498W (*13*) (kind gift of A. Walls and D. Veesler, University of Washington) pseudoviruses were prepared as described (*20*, *84*) using genes encoding S protein sequences lacking C-terminal residues in the cytoplasmic tail: 21 amino acid deletions for SARS-2 variants, WIV1, SHC014, and BtKY72 and a 19 amino acid deletion for SARS-CoV. LYRa3 and Khosta-2 pseudoviruses were made as described (*85*) using genes encoding chimeric spikes containing the RBD from LYRa3 (residues 330-506) or Khosta-2 (residues 324 to 500) and the remaining portions of the spike protein from SARS-1 (substituting for SARS-1 RBD residues 323 to 501). For neutralization assays, threefold serially diluted sera from immunized mice or NHPs were incubated with a pseudovirus for 1 hour at 37°C, then the serum-virus mixture was added to $293T_{ACE2}$ target cells and incubated for 48 hours at 37°C. Media was removed, cells were lysed with Britelite Plus reagent (Perkin Elmer), and luciferase activity was measured as relative luminesce units (RLUs). Relative RLUs were normalized to RLUs from cells infected with pseudotyped virus in the absence of antiserum. Half-maximal inhibitory dilutions (ID_{50} values) were derived using 4-parameter nonlinear regression in AntibodyDatabase (*86*). Statistical significance of titer differences between groups were evaluated using ANOVA test followed by Tukey's multiple comparison post hoc test of ID_{50} values converted to log10 scale using Graphpad Prism 9.3.1.

Mouse serum samples for RBD epitope mapping

Animal procedures and experiments were performed at Labcorp Drug Development (formerly Covance, Inc.) according to protocols approved by the IACUC to obtain serum samples for RBD epitope mapping experiments. Immunizations of mosaic-8b or homotypic SARS-2 Beta (5 μg each based on RBD content, 11.4 μg of total RBD-mi3) in 100 μl of 50% v/v AddaVax adjuvant (InvivoGen) were done using intramuscular (IM) injections of 7-8-week-old female BALB/c mice (Envigo) (8 animals per cohort). Animals were boosted 3 weeks after the prime with the same quantity of antigen in adjuvant. Animals were bled under anesthesia approximately every 2 weeks via orbital sinus and then euthanized 7 weeks after the prime (Day 49) after blood collection from the jugular vein. Blood samples were stored at room temperature in serum separator tubes (BD Microtainer) to allow clotting. Serum was then harvested into microtubes (Mikro-Schraubrohre) and stored at −80°C until use.

RBD sequencing library construction and SARS-2 enrichment

To construct sequencing libraries for RBD epitope mapping of mouse sera, 25 μl of ds-cDNA was brought to a final volume of 53 μl in elution buffer (Agilent Technologies) and sheared on a Covaris LE220 (Covaris) to generate an average size of 180 to 220 base pairs (bp). The following settings were used: peak incident power, 450 W; duty factor, 15%; cycles per burst, 1000; and time, 300 s. The Kapa HyperPrep kit was used to prepare libraries from 50 μl of each sheared cDNA sample following modifications of the Kapa HyperPrep kit (version 8.20) and SeqCap EZ HyperCap Workflow (version 2.3) user guides (Roche Sequencing Solutions Inc.). Adapter ligation was performed for 1 hour at 20°C using the Kapa Unique-Dual Indexed Adapters diluted to 1.5 μM concentration (Roche Sequencing Solutions Inc.). After ligation, samples were purified with AmPure XP beads (Beckman Coulter) and subjected to double-sided size selection as specified in the SeqCap EZ HyperCap Workflow User's Guide. Precapture polymerase chain reaction (PCR) amplification was performed using 12 cycles, followed by purification using AmPure XP beads. Purified libraries were assessed for quality on the Bioanalyzer 2100 using the High-Sensitivity DNA chip assay (Agilent Technologies). Quantification of pre-capture libraries was performed using the Qubit dsDNA HS Assay kit and the Qubit 3.0 fluorometer following the manufacturer's instructions (Thermo Fisher Scientific).

The myBaits Expert Virus bait library was used to enrich samples for SARS-2 according to the myBaits Hybridization Capture for Targeted NGS (version 4.01) protocol. Briefly, libraries were sorted according to estimated genome copies and pooled to create a combined mass of 2 μg for each capture reaction. Depending on estimated genome copies, two to six libraries were pooled for each capture reaction. Capture hybridizations were performed for 16 to 19 hours at 65°C and subjected to 8 to 14 PCR cycles after enrichment. SARS-2–enriched libraries were purified and quantified using the Kapa Library Quant Universal quantitative PCR mix in accordance with the manufacturer's instructions. Libraries were diluted to a final working concentration of 1 to 2 nM, titrated to 20 pM, and sequenced as 2 × 150 bp reads on the MiSeq sequencing instrument using the MiSeq Micro kit version 2 (Illumina).

Sorting of yeast libraries to identify mutations that reduced binding by polyclonal antisera

Plasma mapping experiments were performed in biological duplicate using the independent mutant RBD libraries as previously described (*45*). Prior to the yeast-display deep mutational scanning experiments, 100 μl of each serum sample was heat-inactivated at 56°C for 30 min and twice-depleted of nonspecific yeast-binding antibodies by incubating with 50 OD units of AWY101 yeast containing an empty vector (*55*). Mutant yeast libraries that were pre-sorted for RBD expression and ACE2 binding (*55*) were induced to express RBD in galactose-containing synthetic defined medium with casamino acids (6.7 g/liter Yeast Nitrogen Base, 5.0 g/L Casamino acids, 1.065 g/liter MES acid, and 2% w/v galactose + 0.1% w/v dextrose). 16 to 18 hours post-induction, cells were washed and incubated with plasma at a range of dilutions for 1 hour at room temperature with gentle agitation. For each plasma, we chose a subsaturating dilution such that the amount of fluorescent signal due to plasma antibody binding to RBD was approximately equal across samples. The exact dilution used for each plasma is given in fig. S8B. The libraries were washed and secondarily labeled for 1 hour with 1:100 fluorescein isothiocyanate-conjugated anti-MYC antibody (Immunology Consultants Lab, CYMC-45F) to label for RBD expression and 1:200 Alexa Fluor-647-conjugated goat anti-human-IgG Fc-gamma (Jackson ImmunoResearch 109-135-098) to label for bound NHP antibodies or Alexa Fluor-647-conjugated goat anti-mouse-IgG Fc-gamma (Jackson ImmunoResearch 115-605-008) to label for bound mouse antibodies. A flow cytometric selection gate was drawn to capture RBD mutants with reduced antibody binding for their degree of RBD expression (fig. S8C). For each sample, 7.5×10^6 to 1.1×10^7 cells were processed on the BD FACSAria II cell sorter (fig. S8C). Antibody-escaped cells were grown overnight in synthetic defined medium with casamino acids (6.7 g/liter Yeast Nitrogen Base, 5.0 g/L Casamino acids, 1.065 g/liter MES acid, and 2% w/v dextrose + 100 U/ml penicillin + 100 μg/ml streptomycin) to expand cells prior to plasmid extraction.

DNA extraction and Illumina sequencing

Plasmid samples were prepared from 30 optical density (OD) units [1.6×10^8 colony forming units (CFUs)] of pre-selection yeast populations and approximately 5 OD units (~3.2×10^7 CFUs) of overnight cultures of plasma-escaped cells (Zymoprep Yeast Plasmid Miniprep II) as previously described (*54*, *87*). The 16-nucleotide barcode sequences identifying each RBD variant were amplified by polymerase chain reaction (PCR) and prepared for Illumina sequencing as described (*54*, *87*). Barcodes were sequenced on an Illumina HiSeq 2500 with 50 bp single-end reads. Raw sequencing data are available on the NCBI SRA under BioProject PRJNA770094, BioSample SAMN26315988.

Analysis of deep sequencing data to compute each mutation's escape fraction

Escape fractions were computed essentially as described in (*54*) and exactly as described in (*55*). We used the dms_variants package (https://jbloomlab.github.io/dms_variants, version 0.8.10) to process Illumina sequences into counts of each barcoded RBD variant in each pre-selection and antibody-escape population. We computed the escape fraction for each barcoded variant using the deep sequencing counts for each variant in the original and plasma-escape populations and the total fraction of the library that escaped antibody binding via the formula in (*55*). These escape fractions represent the estimated fraction of cells expressing that specific variant that falls in the escape bin, such that a value of 0 means the variant is always bound by plasma and a value of 1 means that it always escapes plasma binding.

We then applied a computational filter to remove variants with >1 amino-acid mutation, low sequencing counts, or highly deleterious mutations that might escape antibody binding due to poor RBD expression or folding as described (*55*). The reported antibody-escape scores are the average across duplicate libraries; these scores are also in data S3. Correlations in final single-mutant escape scores are shown in fig. S8D. Full documentation of the computational analysis is at https://github.com/jbloomlab/SARS-CoV-2-RBD_Beta_mosaic_np_vaccine.

Data visualization

The static logo plot visualizations of the escape maps in the paper figures were created using the dmslogo package (https://jbloomlab.github.io/dmslogo, version 0.6.2) and in all cases the height of each letter indicates the escape fraction for that amino-acid mutation calculated as described above. For the mouse sera, the static logo plots feature any site where for

>=1 serum, the site-total antibody escape was >10x the median across all sites and at least 10% the maximum of any site. Due to the relative breadth of the NHP sera, a more sensitive threshold for displaying sites on logo plots was used: we include any site where the site-total antibody escape is >5x the median across all sites and at least 5% maximum of any sites. This resulted in sites 383, 386, and 500. Thus, sites 346, 352, 357, 369, 378, 384, 385, 390, 396, 408, 462, 468, 477, 478, 484, 485, 486, and 501 were also added to the logo plots to facilitate comparison to the mouse sera. For each sample, the y-axis is scaled to be the greatest of (a) the maximum site-wise escape metric observed for that sample, or (b) 20x the median site-wise escape fraction observed across all sites for that plasma. The code that generates these logo plot visualizations is available at https://github.com/jbloomlab/SARS-CoV-2-RBD_Beta_mosaic_np_vaccine/blob/main/results/summary/escape_profiles.md. In many of the visualizations, the RBD sites are categorized by epitope region (23) and colored accordingly. Specifically, we define the class 1 epitope as residues 403+405+406+417+420+421+453+455–460+473–478+486+487+489+503+504, the class 2 epitope as residues 472+479+483–485+490–495, the class 3 epitope to be residues 341+345+346+354–357+396+437-452466–468+496+498–501, and the class 4 epitope as residues 365–390+408+462.

For the structural visualizations in figures, the Beta RBD surface [Protein Data Bank (PDB) ID 7LYQ] was colored by the site-wise escape metric at each site, with white indicating no escape and red scaled to be the same maximum used to scale the y-axis in the logo plot escape maps, determined as described above. We created interactive structure-based visualizations of the escape maps using dms-view (*88*) that are available at https://jbloomlab.github.io/SARS-CoV-2-RBD_Beta_mosaic_np_vaccine.

REFERENCES AND NOTES

1. C. Zheng *et al*., Real-world effectiveness of COVID-19 vaccines: A literature review and meta-analysis. *Int. J. Infect. Dis.* **114**, 252–260 (2022). doi: 10.1016/j.ijid.2021.11.009; pmid: 34800687
2. D. Planas *et al*., Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nat. Med.* **27**, 917–924 (2021). doi: 10.1038/s41591-021-01318-5; pmid: 33772244
3. N. L. Washington *et al*., Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* **184**, 2587–2594. e7 (2021). doi: 10.1016/j.cell.2021.03.052; pmid: 33861950
4. T. K. Burki, Omicron variant and booster COVID-19 vaccines. *Lancet Respir. Med.* **10**, e17 (2022). doi: 10.1016/S2213-2600(21)00559-2; pmid: 34929158
5. L. Liu *et al*., Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* **602**, 676–681 (2022). doi: 10.1038/s41586-021-04388-0; pmid: 35016198
6. D. Yamasoba *et al*., Virological characteristics of SARS-CoV-2 BA.2 variant. bioRxiv 480335 [Preprint] (2022); doi: 10.1101/2022.02.14.480335
7. F. Konings *et al*., SARS-CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse. *Nat. Microbiol.* **6**, 821–823 (2021). doi: 10.1038/s41564-021-00932-w; pmid: 34108654
8. S. Alkhovsky *et al*., SARS-like coronaviruses in horseshoe bats (*Rhinolophus* spp.) in Russia, 2020. *Viruses* **14**, 113 (2022). doi: 10.3390/v14010113; pmid: 35062318
9. D. Delaune *et al*., A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *Nat. Commun.* **12**, 6563 (2021). doi: 10.1038/s41467-021-26809-4; pmid: 34753934
10. H. Zhou *et al*., Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* **184**, 4380–4391.e14 (2021). doi: 10.1016/j.cell.2021.06.008; pmid: 34147139
11. S. Wacharapluesadee *et al*., Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 972 (2021). doi: 10.1038/s41467-021-21240-1; pmid: 33563978
12. S. N. Seifert, M. C. Letko, A sarbecovirus found in Russian bats uses human ACE2. bioRxiv 471310 [Preprint] (2021); doi: 10.1101/2021.12.05.471310
13. T. N. Starr *et al*., ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature* **603**, 913–918 (2022). doi: 10.1038/s41586-022-04464-z; pmid: 35114688
14. M. Letko, A. Marzi, V. Munster, Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* **5**, 562–569 (2020). doi: 10.1038/s41564-020-0688-y; pmid: 32094589
15. T. S. Fung, D. X. Liu, Human Coronavirus: Host-Pathogen Interaction. *Annu. Rev. Microbiol.* **73**, 529–557 (2019). doi: 10.1146/annurev-micro-020518-115759; pmid: 31226023
16. P. J. M. Brouwer *et al*., Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643–650 (2020). doi: 10.1126/science.abc5902; pmid: 32540902
17. Y. Cao *et al*., Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73–84.e16 (2020). doi: 10.1016/j.cell.2020.05.025; pmid: 32425270
18. C. Kreer *et al*., Longitudinal isolation of potent near-germline SARS-CoV-2-neutralizing antibodies from COVID-19 patients. *Cell* **182**, 843–854.e12 (2020). doi: 10.1016/j.cell.2020.06.044; pmid: 32673567
19. L. Liu *et al*., Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020). doi: 10.1038/s41586-020-2571-7; pmid: 32698192
20. D. F. Robbiani *et al*., Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020). doi: 10.1038/s41586-020-2456-9; pmid: 32555388
21. R. Shi *et al*., A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **584**, 120–124 (2020). doi: 10.1038/s41586-020-2381-y; pmid: 32454512
22. S. J. Zost *et al*.., Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat. Med.* **26**, 1422–1427 (2020). doi: 10.1038/s41591-020-0998-x; pmid: 32651581
23. T. F. Rogers *et al*., Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* **369**, 956–963 (2020). doi: 10.1126/science.abc7520; pmid: 32540903
24. E. Seydoux *et al*., Analysis of a SARS-CoV-2-infected individual reveals development of potent neutralizing antibodies with limited somatic mutation. *Immunity* **53**, 98–105.e5 (2020). doi: 10.1016/j.immuni.2020.06.001; pmid: 32561270
25. S. J. Zost *et al*.., Potently neutralizing and protective human antibodies against SARS-CoV-2. *Nature* **584**, 443–449 (2020). doi: 10.1038/s41586-020-2548-6; pmid: 32668443
26. C. O. Barnes *et al*., SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682–687 (2020). doi: 10.1038/s41586-020-2852-1; pmid: 33045718
27. D. Pinto *et al*., Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020). doi: 10.1038/s41586-020-2349-y; pmid: 32422645
28. L. Piccoli *et al*., Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042.e21 (2020). doi: 10.1016/j.cell.2020.09.037; pmid: 32991844
29. Z. Wang *et al*., mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* **592**, 616–622 (2021). doi: 10.1038/s41586-021-03324-6; pmid: 33567448
30. H. Kleanthous *et al*., Scientific rationale for developing potent RBD-based vaccines targeting COVID-19. *NPJ Vaccines* **6**, 128 (2021). doi: 10.1038/s41541-021-00393-6; pmid: 34711846
31. H. Liu *et al*., Cross-neutralization of a SARS-CoV-2 antibody to a functionally conserved site is mediated by avidity. *Immunity* **53**, 1272–1280.e5 (2020). doi: 10.1016/j.immuni.2020.10.023; pmid: 33242394
32. C. A. Jette *et al*., Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies. *Cell Rep.* **36**, 109760 (2021). doi: 10.1016/j.celrep.2021.109760; pmid: 34534459
33. D. L. Burnett *et al*., Immunizations with diverse sarbecovirus receptor-binding domains elicit SARS-CoV-2 neutralizing antibodies against a conserved site of vulnerability. *Immunity* **54**, 2908–2921.e6 (2021). doi: 10.1016/j.immuni.2021.10.019; pmid: 34788600
34. A. A. Cohen *et al*., Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* **371**, 735–741 (2021). doi: 10.1126/science.abf6840; pmid: 33436524
35. L. Bongini *et al*., A dynamical study of antibody-antigen encounter reactions. *Phys. Biol.* **4**, 172–180 (2007). doi: 10.1088/1478-3975/4/3/004; pmid: 17928656
36. T. K. Tan *et al*., A COVID-19 vaccine candidate using SpyCatcher multimerization of the SARS-CoV-2 spike protein receptor-binding domain induces potent neutralising antibody responses. *Nat. Commun.* **12**, 542 (2021). doi: 10.1038/s41467-020-20654-7; pmid: 33483491
37. K. D. Brune *et al*., Plug-and-Display: Decoration of Virus-Like Particles via isopeptide bonds for modular immunization. *Sci. Rep.* **6**, 19234 (2016). doi: 10.1038/srep19234; pmid: 26781591
38. B. Zakeri *et al*., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E690–E697 (2012). doi: 10.1073/pnas.1115485109; pmid: 22366317
39. A. H. Keeble *et al*., Approaching infinite affinity through engineering of peptide-protein interaction. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 26523–26533 (2019). doi: 10.1073/pnas.1909653116; pmid: 31822621
40. A. M. Davidson, J. Wysocki, D. Batlle, Interaction of SARS-CoV-2 and other coronavirus with ACE (angiotensin-converting enzyme)-2 as their main receptor: Therapeutic implications. *Hypertension* **76**, 1339–1349 (2020). doi: 10.1161/HYPERTENSIONAHA.120.15256; pmid: 32851855
41. W. Dong *et al*., The K18-human ACE2 transgenic mouse model recapitulates non-severe and severe COVID-19 in response to an infectious dose of the SARS-CoV-2 virus. *J. Virol.* **96**, e0096421 (2022). doi: 10.1128/JVI.00964-21; pmid: 34668775
42. C. K. Yinda *et al*., K18-hACE2 mice develop respiratory disease resembling severe COVID-19. *PLOS Pathog.* **17**, e1009195 (2021). doi: 10.1371/journal.ppat.1009195; pmid: 33465158
43. E. S. Winkler *et al*., SARS-CoV-2 infection of human ACE2-transgenic mice causes severe lung inflammation and impaired function. *Nat. Immunol.* **21**, 1327–1335 (2020). doi: 10.1038/s41590-020-0778-2; pmid: 32839612
44. C. L. Hsieh *et al*., Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020). doi: 10.1126/science.abd0826; pmid: 32703906
45. A. J. Greaney *et al*., Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476.e6 (2021). doi: 10.1016/j.chom.2021.02.003; pmid: 33592168
46. T. Zohar *et al*., Compromised humoral functional evolution tracks with SARS-CoV-2 mortality. *Cell* **183**, 1508–1519.e12 (2020). doi: 10.1016/j.cell.2020.10.052; pmid: 33207184
47. J. Yu *et al*., DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* **369**, 806–811 (2020). doi: 10.1126/science.abc6284; pmid: 32434945
48. N. van Doremalen *et al*., Intranasal ChAdOx1 nCoV-19/AZD1222 vaccination reduces viral shedding after SARS-CoV-2 D614G challenge in preclinical models. *Sci. Transl. Med.* **13**, eabh0755 (2021). doi: 10.1126/scitranslmed.abh0755; pmid: 34315826
49. G. Dagotto *et al*., Comparison of subgenomic and total RNA in SARS-CoV-2 challenged rhesus macaques. *J. Virol.* **95**, e02370-20 (2021). doi: 10.1128/JVI.02370-20; pmid: 33472939
50. P. Kumari *et al*., Neuroinvasion and encephalitis following intranasal inoculation of SARS-CoV-2 in K18-hACE2 mice. *Viruses* **13**, 132 (2021). doi: 10.3390/v13010132; pmid: 33477869
51. B. Pulendran, P. S Arunachalam, D. T. O'Hagan, Emerging concepts in the science of vaccine adjuvants. *Nat. Rev. Drug Discov.* **20**, 454–475 (2021). doi: 10.1038/s41573-021-00163-y; pmid: 33824489
52. P. B. Gilbert *et al*., Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science* **375**, 43–50 (2022). pmid: 34812653

53. F. Hansen *et al*., SARS-CoV-2 reinfection prevents acute respiratory disease in Syrian hamsters but not replication in the upper respiratory tract. *Cell Rep.* **38**, 110515 (2022). doi: 10.1016/j.celrep.2022.110515; pmid: 35263638
54. A. J. Greaney *et al*., Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57.e9 (2021). doi: 10.1016/j.chom.2020.11.007; pmid: 33259788
55. A. J. Greaney *et al*., A SARS-CoV-2 variant elicits an antibody response with a shifted immunodominance hierarchy. *PLOS Pathog.* **18**, e1010248 (2022). doi: 10.1371/journal.ppat.1010248; pmid: 35134084
56. A. C. Walls *et al*., Distinct sensitivities to SARS-CoV-2 variants in vaccinated humans and mice. bioRxiv 479468 [Preprint] (2022); doi: 10.1101/2022.02.07.479468
57. R. Shinnakasu *et al*., Glycan engineering of the SARS-CoV-2 receptor-binding domain elicits cross-neutralizing antibodies for SARS-related viruses. *J. Exp. Med.* **218**, e20211003 (2021). doi: 10.1084/jem.20211003; pmid: 34623376
58. J. Heeney *et al*., Gene delivery of a single, structurally engineered coronavirus vaccine antigen elicits SARS-CoV-2 Omicron and pan-sarbecovirus neutralisation. *Research Square* 995273 [Preprint] (2021); doi: 10.21203/rs.3.rs-995273/v1
59. W. Dejnirattisai *et al*., The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell* **184**, 2183–2200.e22 (2021). doi: 10.1016/j.cell.2021.02.032; pmid: 33756110
60. A. J. Greaney *et al*., Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat. Commun.* **12**, 4196 (2021). doi: 10.1038/s41467-021-24435-8; pmid: 34234131
61. T. N. Starr *et al*., SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**, 97–102 (2021). doi: 10.1038/s41586-021-03807-6; pmid: 34261126
62. T. N. Starr *et al*., Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021). doi: 10.1126/science.abf9302; pmid: 33495308
63. E. Cameroni *et al*., Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* **602**, 664–670 (2022). doi: 10.1038/s41586-021-04386-2; pmid: 35016195
64. D. J. Sheward *et al*., Structural basis of Omicron neutralization by affinity-matured public antibodies. bioRxiv 474825 [Preprint] (2022); doi: 10.1101/2022.01.03.474825
65. K. Westendorf *et al*., LY-CoV1404 (bebtelovimab) potently neutralizes SARS-CoV-2 variants. *Cell Rep.* **39**, 110812 (2022). doi: 10.1016/j.celrep.2022.110812; pmid: 35568025
66. M. G. Joyce *et al*., SARS-CoV-2 ferritin nanoparticle vaccines elicit broad SARS coronavirus immunogenicity. *Cell Rep.* **37**, 110143 (2021). doi: 10.1016/j.celrep.2021.110143; pmid: 34919799
67. K. O. Saunders *et al*., Neutralizing antibody vaccine for pandemic and pre-emergent coronaviruses. *Nature* **594**, 553–559 (2021). doi: 10.1038/s41586-021-03594-0; pmid: 33971664
68. A. E. Powell *et al*., A single immunization with spike-functionalized ferritin vaccines elicits neutralizing antibody responses against SARS-CoV-2 in mice. *ACS Cent. Sci.* **7**, 183–199 (2021). doi: 10.1021/acscentsci.0c01405; pmid: 33527087
69. P. T. Heath *et al*., Safety and efficacy of NVX-CoV2373 COVID-19 vaccine. *N. Engl. J. Med.* **385**, 1172–1183 (2021). doi: 10.1056/NEJMoa2107659; pmid: 34192426
70. W. Wang, B. Huang, Y. Zhu, W. Tan, M. Zhu, Ferritin nanoparticle-based SARS-CoV-2 RBD vaccine induces a persistent antibody response and long-term memory in mice. *Cell. Mol. Immunol.* **18**, 749–751 (2021). doi: 10.1038/s41423-021-00643-6; pmid: 33580169
71. X. Ma *et al*., Nanoparticle vaccines based on the receptor binding domain (RBD) and heptad repeat (HR) of SARS-CoV-2 elicit robust protective immune responses. *Immunity* **53**, 1315–1330.e9 (2020). doi: 10.1016/j.immuni.2020.11.015; pmid: 33275896
72. Q. Geng *et al*., Novel virus-like nanoparticle vaccine effectively protects animal model from SARS-CoV-2 infection. *PLOS Pathog.* **17**, e1009897 (2021). doi: 10.1371/journal.ppat.1009897; pmid: 34492082
73. Y. F. Kang *et al*., Rapid development of SARS-CoV-2 spike protein receptor-binding domain self-assembled nanoparticle vaccine candidates. *ACS Nano* **15**, 2738–2752 (2021). doi: 10.1021/acsnano.0c08379; pmid: 33464829
74. A. C. Walls *et al*., Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382.e17 (2020). doi: 10.1016/j.cell.2020.10.043; pmid: 33160446
75. D. Li *et al*., Breadth of SARS-CoV-2 neutralization and protection induced by a nanoparticle vaccine. bioRxiv 477915 [Preprint] (2022); doi: 10.1101/2022.01.26.477915
76. A. I. Mosa, Antigenic variability. *Front. Immunol.* **11**, 2057 (2020). doi: 10.3389/fimmu.2020.02057; pmid: 33013870
77. M. Kanekiyo *et al*., Mosaic nanoparticle display of diverse influenza virus hemagglutinins elicits broad B cell responses. *Nat. Immunol.* **20**, 362–372 (2019). doi: 10.1038/s41590-018-0305-x; pmid: 30742080
78. A. C. Walls *et al*., Elicitation of broadly protective sarbecovirus immunity by receptor-binding domain nanoparticle vaccines. *Cell* **184**, 5432–5447.e16 (2021). doi: 10.1016/j.cell.2021.09.015; pmid: 34619077
79. C. Fan *et al*., Neutralizing monoclonal antibodies elicited by mosaic RBD nanoparticles bind conserved sarbecovirus epitopes. bioRxiv 497989 [Preprint] (2022); doi: 10.1101/2022.06.28.497989
80. T. U. J. Bruun, A. C. Andersson, S. J. Draper, M. Howarth, Engineering a rugged nanoscaffold to enhance plug-and-display vaccination. *ACS Nano* **12**, 8855–8866 (2018). doi: 10.1021/acsnano.8b02805; pmid: 30028591
81. C. O. Barnes *et al*., Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell* **182**, 828–842.e16 (2020). doi: 10.1016/j.cell.2020.06.025; pmid: 32645326
82. A. A. Cohen *et al*., Construction, characterization, and immunization of nanoparticles that display a diverse array of influenza HA trimers. *PLOS ONE* **16**, e0247963 (2021). doi: 10.1371/journal.pone.0247963; pmid: 33661993
83. L. J. Reed, H. Muench, A simple method of estimating fifty per cent endpoints12. *Am. J. Epidemiol.* **27**, 493–497 (1938). doi: 10.1093/oxfordjournals.aje.a118408
84. K. H. D. Crawford *et al*., Protocol and reagents for pseudotyping lentiviral particles with SARS-CoV-2 spike protein for neutralization assays. *Viruses* **12**, 513 (2020). doi: 10.3390/v12050513; pmid: 32384820
85. S. N. Seifert *et al*., An ACE2-dependent sarbecovirus in Russian bats is resistant to SARS-CoV-2 vaccines. bioRxiv 471310 [Preprint] (2022); doi: 10.1101/2021.12.05.471310
86. A. P. WestJr *et al*., Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10598–10603 (2013). doi: 10.1073/pnas.1309215110; pmid: 23754383
87. T. N. Starr *et al*., Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20 (2020). doi: 10.1016/j.cell.2020.08.012; pmid: 32841599
88. S. K. Hilton *et al*., *dms-view*: Interactive visualization tool for deep mutational scanning data. *J. Open Source Softw.* **5**, 2353 (2020). doi: 10.21105/joss.02353; pmid: 34189395
89. M. Landau *et al*., ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299-302 (2005). doi: 10.1093/nar/gki370; pmid: 15980475
90. S. Guindon *et al*., New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010). doi: 10.1093/sysbio/syq010; pmid: 20525638
91. F. Sievers *et al*., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011). doi: 10.1038/msb.2011.75; pmid: 21988835

ACKNOWLEDGMENTS

We thank J. Vielmetter and the Caltech Protein Expression Center for assistance with protein production; M. Howarth (Oxford) and S. Biswas (SpyBiotech) for helpful discussions about SpyCatcher-SpyTag methodology; F. Laurent and R. Caputo (SPI Pharma) and H. Jain (Panacea Biotec) for VAC20 and EmulsiPan adjuvants, respectively; A. Walls and D. Veesler (University of Washington) for a modified BtKY72 S gene for pseudovirus neutralization assays; M. Holbrook, E. de Wit, B. Williamson, A. Pekosz, C. Martens, K. Barbian, S. Ricklefs, S. Anzick, R. Feldmann, and the Public Health Agency of Canada for viruses used at RML; K. Almquist, A. Darrow, K. Bauer, A. Weidow, R. Cole, L. Heaney, M. Culbert, B. Bailes, C. Henderson, S. Gallogly, L. Crawford, and T. Lippincott for animal care at RML; and C. Kwe Yinda (RML) for technical support. The following reagents were obtained through BEI Resources, NIAID, NIH: *Cercopithecus aethiops* Kidney Epithelial Cells Expressing Transmembrane Protease, Serine 2 and Human Angiotensin-Converting Enzyme 2 (Vero E6-TMPRSS2-T2A-ACE2), NR-54970. We thank Labcorp Drug Development–Antibody Reagents and Vaccines (Denver, PA) (formerly Covance) for carrying out BALB/c mice immunizations for the RBD epitope mapping study. **Funding:** This work was supported by Wellcome Leap (P.J.B.); Bill and Melinda Gates Foundation INV-034638 (P.J.B.); INV-004949 and INV-016575 (J.D.B.); Caltech Merkin Institute (P.J.B.); George Mason University Fast Grant (P.J.B.); Intramural Research Program of the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH) (1ZIAAI001179-01) (V.J.M.); NIAID/NIH contract numbers HHSN272201400006C and 75N93021C00015 (J.D.B.); NIH grant R01AI141707 (J.D.B.); ORIP grant S10OD028685 (The Scientific Computing Infrastructure at Fred Hutch); Caribbean Primate Research Center (grants P40 OD012217 and 2U42OD021458 from ORIP/OD/NIH) (M.G.L.). T.N.S. is a Howard Hughes Medical Institute fellow of the Damon Runyon Cancer Research Foundation (DRG-2381-19). J.D.B. is an investigator of the Howard Hughes Medical Institute. T.K.T. is funded by the Townsend-Jeantet Charitable Trust (charity number 1011770) and the EPA Cephalosporin Early Career Researcher Fellowship. **Author contributions:** Conceptualization: A.A.C., T.K.T., A.R.T., and P.J.B. Methodology: A.A.C., N.v.D., A.J.G., H.A., A.S., T.N.S., J.R.K., J.E.S., M.G.L., J.D.B., V.J.M., and P.J.B. Investigation: A.A.C., N.v.D., A.J.G., H.A., A.S., J.R.K., C.F., J.E.S., P.N.P.G., L.M.K., A.P.W., G.S., Y.E.L., H.G., and C.A.J. Visualization: A.A.C., C.F., A.J.G., J.E.S., and J.D.B. Funding acquisition: P.J.B., J.D.B., and V.J.M. Project administration: J.R.K., M.G.L., J.D.B., V.J.M., and P.J.B.; Supervision: A.A.C., N.v.D., H.A., A.S., J.R.K., A.P.W., M.G.L., J.D.B., V.J.M., and P.J.B.; Writing – original draft: A.A.C. and P.J.B. Writing – review and editing: A.A.C., N.v.D., A.J.G., H.A., J.R.K., C.F., T.K.T., A.R.T., T.N.S., J.D.B., V.J.M., P.J.B. **Competing interests:** P.J.B. and A.A.C. are inventors on a US patent application (17/523,813) filed by the California Institute of Technology that covers the mosaic nanoparticles described in this work, and P.J.B., A.A.C., and C.F. are inventors on provisional US patent applications (63/341,314; 63/285,441) filed by the California Institute of Technology that cover antibodies elicited by the mosaic nanoparticles. J.D.B. consults for Moderna and Flagship Labs 77 on topics related to viral evolution. A.J.G., T.N.S., and J.D.B have the potential to receive a share of intellectual property revenue as inventors on a Fred Hutchinson Cancer Center–optioned technology related to deep mutational scanning of viral proteins and RBD-based vaccine formulations. **Data and materials availability:** All data are available in the main text or the supplementary materials. Raw Illumina sequencing data for the antibody-escape mapping experiments are available on the NCBI Short Read Archive (SRA) at BioProject PRJNA770094, BioSample SAMN26315988. Antibody-escape scores are available at https://github.com/jbloomlab/SARS-CoV-2-RBD_Beta_mosaic_np_vaccine/blob/main/results/supp_data/all_raw_data.csv. Materials are available upon request to the corresponding author with a signed material transfer agreement. **License information:** This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq0839
Materials and Methods
Figs. S1 to S10
Data S1 to S3
MDAR Reproducibility Checklist

Submitted 17 March 2022; accepted 29 June 2022
10.1126/science.abq0839

RESEARCH ARTICLE SUMMARY

HEART DISEASE

Pathogenic variants damage cell composition and single-cell transcription in cardiomyopathies

Daniel Reichart†, Eric L. Lindberg*†, Henrike Maatz†, Antonio M. A. Miranda, Anissa Viveiros, Nikolay Shvetsov, Anna Gärtner, Emily R. Nadelmann, Michael Lee, Kazumasa Kanemaru, Jorge Ruiz-Orera, Viktoria Strohmenger, Daniel M. DeLaughter, Giannino Patone, Hao Zhang, Andrew Woehler, Christoph Lippert, Yuri Kim, Eleonora Adami, Joshua M. Gorham, Sam N. Barnett, Kemar Brown, Rachel J. Buchan, Rasheda A. Chowdhury, Chrystalla Constantinou, James Cranley, Leanne E. Felkin, Henrik Fox, Ahla Ghauri, Jan Gummert, Masatoshi Kanda, Ruoyan Li, Lukas Mach, Barbara McDonough, Sara Samari, Farnoush Shahriaran, Clarence Yapp, Caroline Stanasiuk, Pantazis I. Theotokis, Fabian J. Theis, Antoon van den Bogaerdt, Hiroko Wakimoto, James S. Ware, Catherine L. Worth, Paul J. R. Barton, Young-Ae Lee, Sarah A. Teichmann, Hendrik Milting†, Michela Noseda†, Gavin Y. Oudit†, Matthias Heinig†, Jonathan G. Seidman*†, Norbert Hubner*†, Christine E. Seidman*†

INTRODUCTION: Human heart failure is a highly morbid condition that affects 23 million individuals worldwide. It emerges in the setting of an array of different cardiovascular disorders, which has propelled the notion that diverse stimuli converge on a common final pathway. Consistent with this, initiating etiologies do not direct heart failure treatments, which are often inadequate and necessitate mechanical interventions and cardiac transplantation.

The recent application of single-nucleus RNA sequencing (snRNAseq) transcriptional analyses to characterize the cellular composition and molecular states in the healthy adult human heart provides an emerging benchmark by which disease-related changes can be assessed. Moreover, the discovery of human pathogenic variants that cause dilated cardiomyopathy (DCM) and arrhythmogenic cardiomyopathy (ACM), disorders associated with high rates of heart failure, provides direct opportunities to evaluate whether genotype influences heart failure pathways.

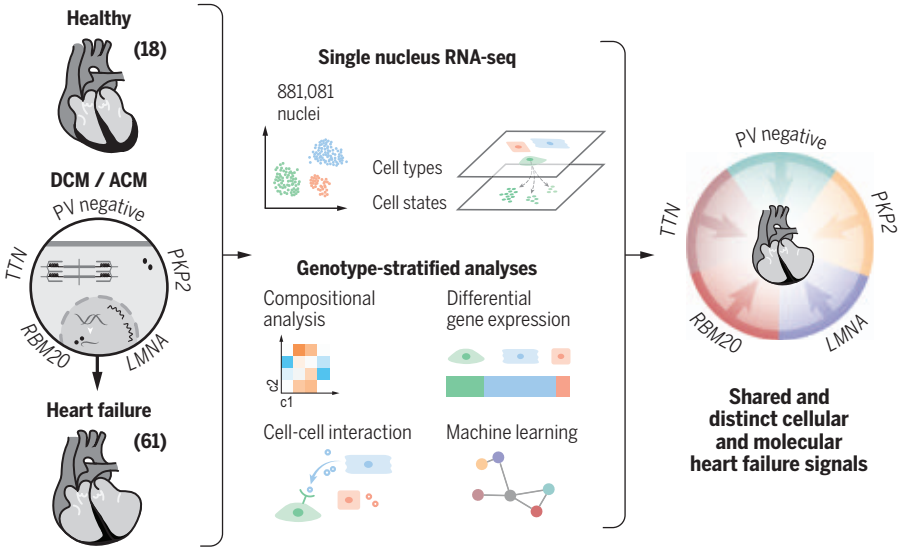
RATIONALE: A systematic identification of shared and distinct molecules and pathways involved in heart failure is lacking, and knowledge of these fundamental data could propel the development of more effective treatments. To enable these discoveries, we performed snRNAseq of explanted ventricular tissues from 18 healthy donors and 61 heart failure patients. By focusing analyses on multiple samples with pathogenic variants in DCM genes (*LMNA*, *RBM20*, and *TTN*), ACM genes (*PKP2*), or pathogenic variant–negative (PV negative) samples, we characterized genotype-stratified and common heart failure responses.

RESULTS: From 881,081 nuclei isolated from left and right diseased and healthy ventricles, we identified 10 major cell types and 71 distinct transcriptional states. DCM and ACM tissues showed significant depletion of cardiomyocytes and increased endothelial and immune cells. Fibrosis was expanded in disease hearts, but, unexpectedly, fibroblasts were not increased, and instead showed altered transcriptional states that indicated activated remodeling of the extracellular matrix. Genotype-stratified analyses identified multiple transcriptional changes shared only among the hearts harboring pathogenic variants or distinctive for individual and subsets of DCM and ACM genotypes. We validated many of these by single-molecule fluorescent in situ hybridization.

Through analyses of receptor and ligand expression across all cells, we observed changes in intercellular signaling and communications, such as increased endothelin signaling in *LMNA* hearts, tumor necrosis factor in *PKP2* hearts, and others. We also identified specific cardiac cell lineages expressing genes with common polymorphisms that were identified in validated association studies of DCM.

Because our findings indicated genotype-enriched transcripts and cell states, we harnessed machine learning to develop a graph attention network for the multinomial classification of genotypes. This network showed remarkably high prediction of the genotypes for each cardiac sample, thereby reinforcing our conclusion that genotypes activate very specific heart failure pathways.

CONCLUSION: snRNAseq of human ventricular samples illuminated cell types and states, molecular signals, and intercellular communications that characterize DCM and ACM. The cellular and molecular architectures that induce heart failure are both shared and distinct across genotypes. These data provide candidate therapeutic targets for future research and interventional opportunities to improve and personalize treatments for cardiomyopathies and heart failure. ■



Genotype-stratified analyses of heart failure at the single-nuclei level. The transcriptomes of 881,081 nuclei from 61 heart failure patients were profiled and compared with the transcriptional signatures of 18 healthy controls. Genotype-stratified analyses of cell types and cell state compositions, differential gene expression, cell-cell interactions, and machine learning illuminated the shared and distinct transcriptional signatures resulting from pathogenic variants in DCM and ACM.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: eric.lindberg@mdc-berlin.de (E.L.L.); seidman@genetics.med.harvard.edu (J.G.S.); nhuebner@mdc-berlin.de (N.H.); cseidman@genetics.med.harvard.edu (C.E.S.)
†These authors contributed equally to this work.
Cite this article as D. Reichart *et al.*, *Science* **377**, eabo1984 (2022). DOI: 10.1126/science.abo1984

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abo1984

RESEARCH ARTICLE

HEART DISEASE

Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies

Daniel Reichart[1,2,3]†, Eric L. Lindberg[4]†*, Henrike Maatz[4,5]†, Antonio M. A. Miranda[6,7], Anissa Viveiros[8,9], Nikolay Shvetsov[4], Anna Gärtner[10], Emily R. Nadelmann[1], Michael Lee[6], Kazumasa Kanemaru[11], Jorge Ruiz-Orera[4], Viktoria Strohmenger[1,12], Daniel M. DeLaughter[1,13], Giannino Patone[4], Hao Zhang[8,9], Andrew Woehler[14], Christoph Lippert[15,16], Yuri Kim[1,2], Eleonora Adami[4], Joshua M. Gorham[1], Sam N. Barnett[6], Kemar Brown[1,17], Rachel J. Buchan[6,18], Rasheda A. Chowdhury[6], Chrystalla Constantinou[6], James Cranley[11], Leanne E. Felkin[6,18], Henrik Fox[19], Ahla Ghauri[20], Jan Gummert[19], Masatoshi Kanda[4,21], Ruoyan Li[11], Lukas Mach[6,18], Barbara McDonough[2,13], Sara Samari[6], Farnoush Shahriaran[22], Clarence Yapp[23], Caroline Stanasiuk[10], Pantazis I. Theotokis[6,24], Fabian J. Theis[22], Antoon van den Bogaerdt[25], Hiroko Wakimoto[1], James S. Ware[6,18,24], Catherine L. Worth[4], Paul J. R. Barton[6,18,24], Young-Ae Lee[20,26], Sarah A. Teichmann[11,27], Hendrik Milting[10]†, Michela Noseda[6,7]†, Gavin Y. Oudit[8,9]†, Matthias Heinig[22,28,29]†, Jonathan G. Seidman[1]†*, Norbert Hubner[4,5,30]†*, Christine E. Seidman[1,2,12]†*

Pathogenic variants in genes that cause dilated cardiomyopathy (DCM) and arrhythmogenic cardiomyopathy (ACM) convey high risks for the development of heart failure through unknown mechanisms. Using single-nucleus RNA sequencing, we characterized the transcriptome of 880,000 nuclei from 18 control and 61 failing, nonischemic human hearts with pathogenic variants in DCM and ACM genes or idiopathic disease. We performed genotype-stratified analyses of the ventricular cell lineages and transcriptional states. The resultant DCM and ACM ventricular cell atlas demonstrated distinct right and left ventricular responses, highlighting genotype-associated pathways, intercellular interactions, and differential gene expression at single-cell resolution. Together, these data illuminate both shared and distinct cellular and molecular architectures of human heart failure and suggest candidate therapeutic targets.

Dilated cardiomyopathy (DCM), a prevalent disorder occurring in one in 250 individuals, is characterized by left ventricular (LV) dilatation, cardiomyocyte loss with fibrotic replacement, and impaired contractility (*1*). Arrhythmogenic cardiomyopathy (ACM) similarly incites ventricular dysfunction, often with more prominent right ventricular (RV) involvement, high arrhythmogenic burden, and fibrofatty accumulations (*2*). Both disorders can arise from genetic causes (*1*, *2*). DCM genes encode proteins involved in contractility [titin (*TTN*), troponin T (*TNNT2*), troponin C (*TNNC1*), tropomyosin (*TPM1*), and filamin C (*FLNC*)] that regulate cardiac splicing [RNA-binding motif protein (*RBM20*)] or calcium sequestration [phospholamban (*PLN*)] and maintain cytoskeletal [desmin (*DES*)] or nuclear [lamin A/C (*LMNA*)] integrity. ACM genes often encode desmosome proteins, including plakophilin-2 (*PKP2*) and desmoplakin (*DSP*). The cardiomyocyte-specific expression and damaging effects of pathogenic variants (PVs) in many DCM and ACM genes propel the development of arrhythmias and heart failure, a highly morbid condition affecting 23 million individuals worldwide (*3*).

We hypothesized that PVs in different genes would evoke distinct single-cell molecular phenotypes. To address this, we studied the molecular signals underlying heart failure pathogenesis using single-nucleus RNA sequencing (snRNAseq) of human hearts with advanced DCM and ACM compared with nonfailing donor (control) hearts. We revealed differences in the cellular landscape and transcriptional changes between several DCM and ACM genotypes. By leveraging machine-learning approaches, we illuminated genotype-specific molecular responses, as validated by reconstructing the underlying PVs using snRNA-seq data.

Results

Study cohort

We studied LV and RV tissues (Fig. 1A) obtained before any mechanical support in 61 patients and 18 controls (tables S1 and S2), including 12 controls previously reported (*4*). Thirty-eight independent PVs were identified in three DCM genes (*LMNA*, $n = 12$; *RMB20*, $n = 8$; *TTN*, $n = 12$) and in the ACM gene *PKP2* ($n = 6$), whereas no PVs were detected in eight DCM patients (PVneg) (Fig. 1B). Analyses were performed for these five genotypes individually ($n = 46$), for aggregated DCM genotypes (*LMNA*, *RBM20*, *TTN*, and PVneg), ACM (*PKP2*), and controls. Additionally, we generated data from 15 PVs across *PLN*, *BAG3*, *DES*, *FLNC*, *FKTN*, *TNNC1*, *TNNT2*, *TMP1*, and *DSP* (tables S1 and S2). Because there were few recurrent PVs, these genes were excluded in downstream analyses except where indicated.

Males predominated among patients (60%) and controls (72%) (fig. S1A). The mean age of patients was 48.4 ± 4.3 years exclusive of *RBM20* (mean age 32.9 ± 14.6 years). Clinical manifestations indicated similar LV dysfunction in *LMNA*, *TTN*, and PVneg patients, but greater LV dilatation and reduced systolic contraction in *RBM20* and preserved LV with reduced RV function in *PKP2* patients. *LMNA*

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Cardiovascular Division, Brigham and Women's Hospital, Boston, MA 02115, USA. [3]Department of Medicine I, University Hospital, LMU Munich, 80336 Munich, Germany. [4]Cardiovascular and Metabolic Sciences, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany. [5]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, 10785 Berlin, Germany. [6]National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK. [7]British Heart Foundation Centre for Research Excellence and Centre for Regenerative Medicine, Imperial College London, London WC2R 2LS, UK. [8]Division of Cardiology, Department of Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta T6G 2R3, Canada. [9]Mazankowski Alberta Heart Institute, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta T6G 2R3, Canada. [10]Erich and Hanna Klessmann Institute, Heart and Diabetes Center NRW, University Hospital of the Ruhr-University Bochum, 32545 Bad Oeynhausen, Germany. [11]Cellular Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton CB10 1SA, UK. [12]Walter-Brendel-Centre of Experimental Medicine, Ludwig-Maximilian University of Munich, 81377 Munich, Germany. [13]Howard Hughes Medical Institute, Bethesda, MD 20815, USA. [14]Systems Biology Imaging Platform, Berlin Institute for Medical Systems Biology (BIMSB), Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), 10115 Berlin, Germany. [15]Digital Health-Machine Learning group, Hasso Plattner Institute for Digital Engineering, University of Potsdam, 14482 Potsdam, Germany. [16]Hasso Plattner Institute for Digital Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [17]Cardiac Unit, Massachusetts General Hospital, Boston, MA 02114, USA. [18]Royal Brompton and Harefield Hospitals, Guy's and St. Thomas' NHS Foundation Trust, London SW3 6NR, UK. [19]Heart and Diabetes Center NRW, Clinic for Thoracic and Cardiovascular Surgery, University Hospital of the Ruhr-University, 32545 Bad Oeynhausen, Germany. [20]Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany. [21]Department of Rheumatology and Clinical Immunology, Sapporo Medical University School of Medicine, Sapporo 060-8556, Japan. [22]Computational Health Center, Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), 85764 Neuherberg, Germany. [23]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115, USA. [24]MRC London Institute of Medical Sciences, Imperial College London, London W12 0NN, UK. [25]ETB-Bislife Foundation, 2300 AH Leiden, Netherlands. [26]Clinic for Pediatric Allergy, Experimental and Clinical Research Center, Charité-Universitätsmedizin Berlin, 13125 Berlin, Germany. [27]Department of Physics, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK. [28]Department of Informatics, Technische Universitaet Muenchen (TUM), 85748 Munich, Germany. [29]DZHK (German Centre for Cardiovascular Research), Munich Heart Association, Partner Site Munich, 10785 Berlin, Germany. [30]Charité-Universitätsmedizin Berlin, 10117 Berlin, Germany.
***Corresponding author. Email: eric.lindberg@mdc-berlin.de (E.L.L.); seidman@genetics.med.harvard.edu (J.G.S.); nhuebner@mdc-berlin.de (N.H.); cseidman@genetics.med.harvard.edu (C.E.S.)**
†These authors contributed equally to this work.

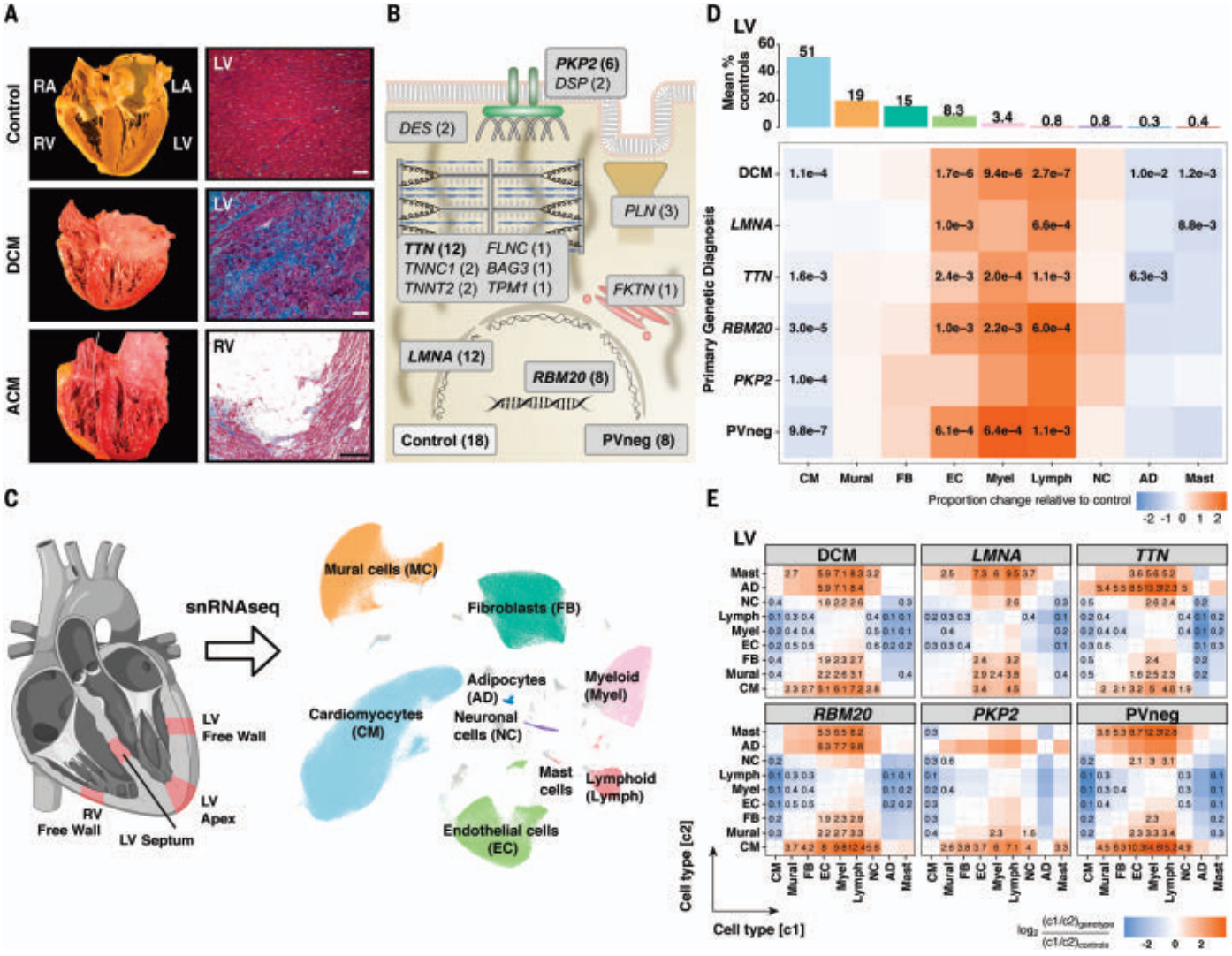


Fig. 1. PVs and unexplained causes of DCM and ACM alter cardiac morphology, histopathology, and cellular compositions. (**A**) Comparisons of normal cardiac anatomy and histology with DCM, demonstrated LV dilation with fibrosis, and ACM-revealed RV dilation with fibrofatty degeneration (Masson trichrome staining; scale bar, 10 μm). (**B**) Schematic depiction of the functions of DCM and ACM genes with PVs (number indicates unique genotypes, bold denotes six or more patients) in studied tissues. (**C**) Single nuclei isolated from transmural LV (free wall, apex, or septum) and RV sections were processed using 10X Chromium 3′ chemistry. UMAP embedding of 881,081 nuclei delineated 10 cell types and unassigned populations (gray). (**D**) Upper panel: Mean abundance (%) of cell types in control LVs. Lower panel: Proportional changes of cell types in specified genotypes or aggregated across DCM genotypes. Proportional changes are scaled by color: Red indicates an increase and blue a decrease in disease versus control. *P* values are indicated for significant proportional changes (FDR ≤ 0.05). (**E**) Pairwise cell type abundance ratios in specified genotypes or aggregated DCM genotypes in LVs relative to controls. Color scale, FDR, and significance are as in (D).

and *TTN* patients received more pacemaker and/or resynchronization therapies than those with other genotypes (table S1).

Disease-associated compositional changes of cell types

Nuclei were isolated as described previously (*4*) from full thickness LV free wall, apex, septum, and RV free wall (Fig. 1C). We compared ~500,000 high-quality nuclei from patients' ventricular tissues and ~380,000 nuclei from controls (fig. S1, B to E). After preprocessing and quality control filtering, data were integrated across samples using Harmony before constructing manifolds using Uniform Manifold Approximation and Projections (UMAPs). Clustering identified 10 major cardiac cell types encompassing ventricular cardiomyocytes (CMs), fibroblasts (FBs), adipocytes (ADs), pericytes and smooth muscle cells (mural, MCs), endothelial cells (ECs), myeloid and lymphoid (immune cells), neuronal (NCs), and mast cells (Fig. 1C) with 71 distinct cell states. States of the same cell type shared a transcriptional profile but also expressed distinct genes, which implied biological differences.

Cell type abundances in sample replicates were highly correlated (Pearson coefficient 0.74 to 0.99) (fig. S2, A and B). Cell composition, states, and transcript counts across the free wall, apex, and septum showed high similarities, and therefore are reported grouped together as LV (fig. S2, C and D).

Using center log ratio–transformed abundance of cell types, we considered the effects of sex on LV and RV cell compositions in DCM (n = 10 female, n = 29 male) and control (n = 7 female, n = 11 male) hearts (fig. S2E). Myeloid cells showed a modest sex-specific difference [false discovery rate (FDR) = 0.016]. Only *LMNA* tissues (from n = 7 males and n = 5 females) were sufficient for genotype-specific, sex-associated

analyses. Female LV myeloid cells were increased (FDR = 0.05), paralleling the findings in DCM versus controls. In addition, RV endothelial cells showed a modest female-specific increase between *LMNA* patients and controls (FDR = 0.048).

The proportions of LV nuclei across the genotypes demonstrated depletion of CMs except in *LMNA*, and increased ECs and immune cells except in *PKP2* (Fig. 1D and fig. S3). In RVs (fig. S4), CMs were depleted in the DCM subgroup except for *TTN*, whereas ECs were increased in *LMNA*, *TTN*, and *RBM20*, and immune cells were not changed. FBs were not increased in LVs and RVs (Fig. 1D) despite histopathological findings of fibrosis, which implied the acquisition of a secretory rather than a proliferative phenotype.

Individual-level abundances for cell types (LV, table S3; RV, table S4) and cell states (table S5) are provided for controls and patients. Different cell type abundances in disease compared with control LVs remained significant using a linear model adjusting for age. Pairwise cell type ratios in disease LVs compared with controls confirmed loss of CMs and showed accompanying increased ECs, lymphoid and myeloid cells, and altered ratios (FB and mural cells compared with CMs) with quantitative, genotype-specific differences (Fig. 1E). *RBM20* and PVneg, respectively, showed greatest shifts in the EC:CM (8- and 10.3-fold), myeloid:CM (9.8- and 14.6-fold), and lymphoid:CM (12.4- and 15.2-fold) ratios. By contrast, pairwise ratios of all cell types compared with CMs were modest or unchanged in *LMNA*.

Genotypes diversify cell types and states

Cardiomyocytes

Disease and control ventricles exhibited the previously described cardiomyocyte states, vCM1.0 to vCM5 (*4*), and four new states, vCM1.1, vCM1.2, vCM1.3, and vCM3.1 (Fig. 2, A and B, and table S6). Across all vCM cell states (fig. S5), differentially expressed genes (DEGs) in disease connoted increased stress (e.g., *NPPB*) and apoptosis (tables S7 to S12 and fig. S6A). Although these findings implied late-stage transcriptional convergence, 20 to 40% of DEGs were genotype specific (fig. S5). Only PVneg reduced *MYH6* expression, increasing *MYH7:MYH6* ratios (fig. S6B). Conversely, only vCMs with PVs down-regulated *SMYD1* (Fig. 2C; fig. S6, C to E; table S13), a cardioprotective muscle–specific histone methyltransferase that regulates sarcomere assembly and mitochondrial energetics (*5*, *6*), and *ADRB1* (β1-adrenergic receptor; fig. S6C), which is indicative of adrenergic activation and is therapeutically targeted by β-blockade medicines (*7*, *8*). Genotype-selective responses included up-regulation of *FNIP2* (Fig. 2D), inhibiting AMP-activated protein kinase activity and oxidative metabolism (*9*–*11*) in *LMNA* and *PKP2*, and *CPEB4*, an RNA-binding protein that activates glycolysis and stimulates fibrosis (*12*) in *LMNA*, *RBM20*, and *PKP2* (fig. S6F).

Proportional differences in vCM states (figs. S6G, S7, and S8) varied among genotypes and controls and were more prominent in RVs than LVs, perhaps reflecting hemodynamic differences. PVs generally increased vCM1.1 and vCM1.2, with accompanying decreased vCM3.1 in *LMNA* and *TTN* (Fig. 2B and figs. S7A and S8, A to D). State-enriched DEGs for vCM1.1 included *MYO18B*, which is required for sarcomere formation (*13*); *XPR1*, which regulates phosphate homeostasis (*14*); and *IGF1R* and *ROR1*, which are involved in cell survival (*8*). vCM1.2 had heightened expression of electrophysiologic genes including

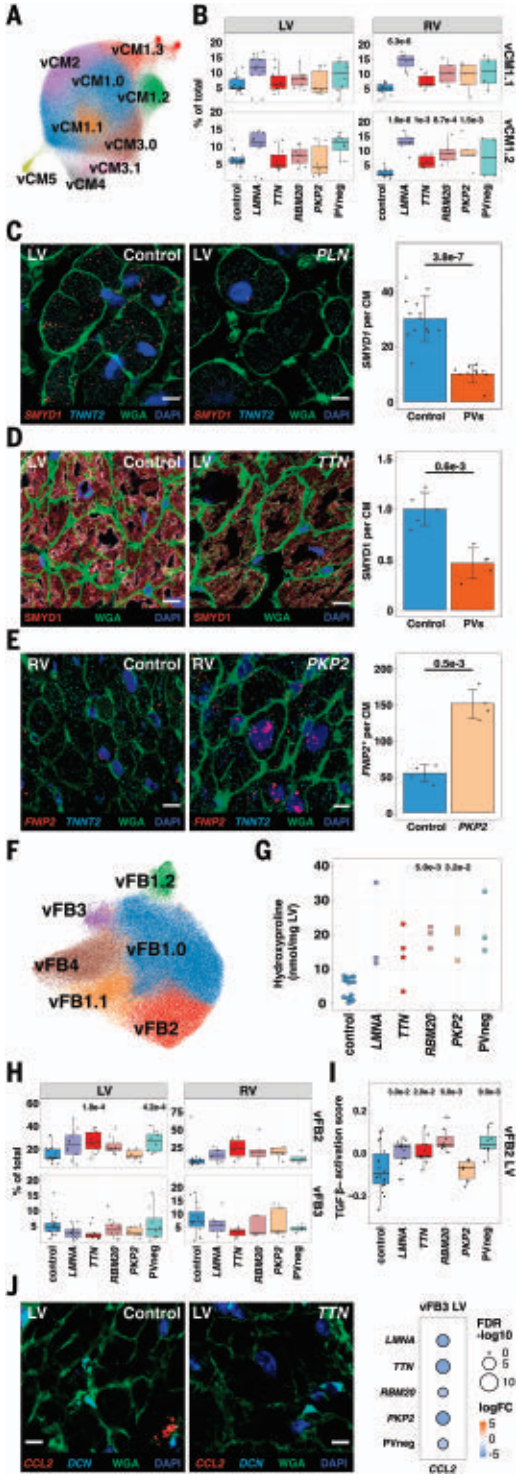


Fig. 2. Cardiomyocytes and fibroblast states in control, DCM, and ACM ventricles. (**A**) UMAP depicting CM states in all tissues. (**B**) Control and disease LV and RV abundance analyses for vCM1.1 (upper panel) and vCM1.2 (lower panel). (**C**) Single-molecule RNA fluorescent in situ hybridization exemplifies decreased *SMYD1* (red) expression in CMs (identified by *TNNT2* transcripts, cyan) within a DCM heart with a PV in *PLN* (phospholamban). Cell boundaries are stained with WGA (green); nuclei are stained with DAPI (blue). Scale bar, 10 μm. Quantified expression (spots per CM) and *p*-values from four independent control and disease LVs with PVs were assessed. (**D**) Immunohistochemistry validated decreased SMYD1 (red) protein in CMs (identified by troponin T staining, fig. S6E) in TTN LV section. Cell boundaries are stained with wheat germ agglutinin (WGA; green); nuclei are stained with 4′,6-diamidino-2-phenylindole (DAPI; blue). Scale bar, 10 μm. Quantified protein levels (intensity per CM) and *P* values were assessed from five independent control and DCM LVs with PVs. (**E**) Single-molecule RNA fluorescent in situ hybridization demonstrated increased expression of *FNIP2* (red). CMs, nuclei, and cell boundaries are labeled as in (C). Scale bar, 10 μm. Quantified expression of *FNIP2* (spots per CM and H-score) and *P* values reflect analyses of two independent control and *PKP2* samples. (**F**) UMAP depicting FB states. (**G**) Hydroxyproline assay (HPA) quantified cardiac collagen content for each genotype. (**H**) Control and diseased LV and RV abundance analyses for vFB2 (upper panel) and vFB3 (lower panel). (**I**) Pathway score of TGFβ activation in LV vFB2. (**J**) Single-molecule RNA fluorescent in situ hybridization showing decreased expression of *CCL2* (red) in vFB3 (DCN, cyan) in disease compared with controls. Cell boundaries are stained with WGA (green); nuclei are stained with DAPI (blue). Scale bar, 10 μm. Dot plot illustrating the log2 fold change (log2FC) and significance [−log10(FDR)] of *CCL2* expression in LV vFB3 across genotypes.

PCDH7, which enables intercellular contacts (*15*, *16*) but when overexpressed impedes synaptic currents (*17*); *CAMK2D*, which modulates excitation-contraction coupling (*18*); and *ANK2*, which harbors arrhythmogenic PVs (*19*, *20*). vCM1.3 was modestly increased in *RBM20*, associated with the largest numbers of DEGs in all genotypes, and enriched for *BMP* receptors [*BMPR1A*, *BMPR1B* (*21*,*22*); fig. S7, B and D) and *GPC5*, which regulates BMP signaling (*23*). vCM3.1, which was reduced in *LMNA* and *TTN*, expressed early-adaptive transcriptional regulators of stress responses (*ATF3* and *ATF4*) and cardioprotection (*NR4A3;* fig. S7E) (*24*). Only PVneg depleted vCM2 and was enriched for *SH3RF2*, an anti-apoptosis regulator of the *JNK* pathway (*25*). Because DCM and ACM genes are highly and often selectively expressed in CMs, these disease-associated vCM states and DEGs defined intrinsic responses to PVs as well as cell-cell responses to CM death, and depressed contractile performance and arrhythmogenicity.

Fibroblasts

We identified four previously characterized FB states (*4*) and two new ones, vFB1.1 and vFB1.2 (Fig. 2E and table S14). Across all vFB states, up-regulated DEGs included genes involved in extracellular matrix (ECM) remodeling (figs. S9 to S11 and tables S15 to S21). *LMNA*, *TTN*, and *PKP2* increased fibrogenic signaling receptor, and *EGFR* and *PKP2* also increased *AGTR1* (fig. S10A), which enables EGFR transactivation (*26*, *27*). The expression of profibrotic *TGFβ2* was universally increased (fig. S10B). DCM hearts were enriched for *PCOLCE2*, promoting insoluble collagen formation (*28*), and *LMNA*, *TTN*, and PVneg down-regulated metalloproteinase inhibitors *TIMP1* and *TIMP3* (fig. S10C). Collagen genes showed genotype-specific expression. *COL4A1* and *COL4A2* were up-regulated in *LMNA*, *TTN*, and *PKP2*, whereas *COL4A5* and *COL24A1* were enriched in PVneg. Thus, although ECM up-regulation and collagen deposition (Fig. 2F and fig. S10, D to F) were shared features, genotype and chamber influenced composition and organization.

FB state abundance was also altered across genotypes (figs. S11A and S12, A to D). vFB1.1 and vFB1.2 expressed canonical vFB1.0 genes, but vFB1.1 also up-regulated *APOD*, *APOE*, and *APOO*, which typify lipogenic fibroblasts (*29*), and *CST3* (fig. S13A), which is involved in matrix remodeling. vFB1.2-enriched genes are related to actin filament assembly (*DAAM1*) and chondrocyte differentiation (*GPC6*) (fig. S13B). vFB2 expressed prominent profibrotic genes, including *TGFβ* targets (increased in DCM LVs) and fibrogenic *IL11* (*30*) (highest in *RBM20* LVs) (Fig. 2H and fig. S14, A to C). vFB2 was increased in *TTN* and PVneg and modestly in other DCM LVs (Fig. 2, G and H, and fig. S11). vFB3, which was diminished in hearts with PVs, expressed proinflammatory cytokine genes (*CCL2*) and genes related to *OSM* signaling (*IL6ST*, *OSMR*, and oncostatin M-target genes) (Fig. 2I and fig. S14D). The resultant increased ratio of vFB2:vFB3 and altered DEGs suggested dysregulation of fibroblast-to-macrophage interactions that would promote deleterious ECM remodeling, particularly in *TTN* LV and RV (fig. S12, C and D).

Smooth muscle cells and pericytes

Three previously described pericyte cell states (PC1, PC2, and PC3) and three smooth muscle cell states (SMC1.1, SMC1.2, and SMC2; Fig. 3A and table S22) were identified. DEGs across all states (fig. S15 and tables S23 to S28) showed up-regulation of the sodium channel *SCN3A*, with unknown vascular functions, and the noncoding antisense *ADAMTS9-AS2*, with concordant down-regulation of *ADAMTS9*, a metalloprotease involved in ECM remodeling (*31*, *32*) (Fig. 3B and fig. S16A). Overall, the genes with dysregulated expression indicated that disease evoked signals to synthesize specific ECM and integrin components.

DEGs in PC states included down-regulation of two central signaling receptors, *NOTCH3* and *PDGFRB*, in *PKP2* and *TTN* PC1 (Fig. 3B). *NOTCH3* is required for SMC maturation and deficiency causes pericyte dysfunction and arteriovenous malformations. Notch signaling regulates *PDGFRB*, which is necessary for angiogenesis and PC recruitment (*33*–*37*).

DEGs in SMCs subdivided the previously identified canonical SMC1 (*4*) into two states: SMC1.1, which had higher expression of *ACTA2*, *MYH11*, and *TAGLN*, and SMC1.2, which strongly expressed *ITGA8*, required for maintenance of SMC contractile phenotype (*38*, *39*), and *ATP10A*, suggesting vascular stiffness and increasing diastolic dysfunction (Fig. 3B). Methylation of the *ATP10A* locus in SMC decreases with age and atherosclerosis (*40*). SMC1.2 was enriched in *LMNA*, *TTN*, and PVneg RVs (figs. S17 and S18). SMC2 expressed high levels of genes involved in collagen and elastic fiber formation (*ELN* and *LAMA2*) and *MYH10*, demarcating dedifferentiated SMCs with secretory properties (figs. S16B and S17C). *LMNA* and *PKP2* SMC2 up-regulated *SLIT3*, a stimulator of fibroblast activity (*41*), and ECM synthesis and collagen formation genes (Fig. 3, B and C). Collectively, disease remodeling of MC indicated modulation of PDGF and NOTCH signaling receptors and synthesis of selective ECM and integrin components.

Endothelial cells

Characterization of ECs identified seven previously described cell states (*4*): EC1, EC2, EC5 to EC8 and mesothelial cells) (Fig. 3D, figs. S19 to S21, and table S29). Shared DEGs occurred across all and between genotypes (tables S30 to S35), with clear RV and LV differences (fig. S19). Dysregulation of genes encoding factors involved in EC fate, blood vessel development, and angiogenesis (*NOTCH4*, *FLT1*, *FGFR1*, and *RGCC*) (*42*–*44*) in disease LVs indicated that vascular remodeling was a common heart failure feature.

EC composition in DCM RVs and LVs increased EC5 compared with controls (fig. S21A), whereas genotype-specific cell state ratios in LVs generally decreased EC1 relative to EC2, EC5, and EC6 (Fig. 3F and fig. S21C). Assessment of selective gene enrichment scores for proliferating cells, endothelial-to-mesenchymal transformation, and cell death showed no difference in disease compared with controls (fig. S22 and table S36).

EC7 had the most DEGs and the highest proportion of unique DEGs across all genotypes. Although initially defined as being atrial-enriched (*4*), EC7 expressed endocardial-enriched genes [*SMOC1* (*45*), *NPR3* (*46*), and *POSTN* (*47*)] (fig. S20, B and C), which prompted a revised annotation to endocardial cells. Up-regulated genes in EC7 from DCM LVs and *PKP2* RVs (Fig. 3G) encoded secreted proteins involved in myocardial stress adaptation (*NRG1*) (*48*), CM force production (*EDN1*) (*49*, *50*), and endocardial expansion during development or after cardiac injury (*BMP6*) (*51*,*52*). *BMP6* was selectively up-regulated in the endocardium of DCM LVs and ACM RVs (Fig. 3E). Conversely, *INHBA*, a TGF-β superfamily member involved in atrioventricular canal development (*53*), and cell adhesion molecule *OPCML* were down-regulated. In *PKP2*, an unconventional myosin promoting cell adhesion (*MYO10*), and *POSTN* were up-regulated in both ventricles. These data highlighted the involvement of the endocardium in chamber-specific remodeling of cardiomyopathies.

Myeloid cells

We classified 14 subclusters of myeloid cells comprising macrophages (MPs), monocytes (MOs), conventional dendritic cells (cDC1 and cDC2), and proliferating myeloid cells (Fig. 4A, figs. S23 to S27, and tables S37 to S45). Distinct gene sets were unique to each genotype and were particularly pronounced in PVneg and *PKP2* in LVs (fig. S24).

Although disease increased myeloid cells, the proportions of proliferating myeloids (Fig. 4, B and C, and figs. S26 and S27) were consistently lower compared with controls, implying monocyte infiltration. The tissue-resident MPs *LYVE1*^{high}MHCII^{low} and *LYVE1*^{low}MHCII^{high} (fig. S25C) were the most abundant myeloid cells (fig. S27, A and B). MP *LYVE1*^{low}MHCII^{high} were increased in the RV of *LMNA*, with similar trends in other genotypes (fig. S27B). Proportions of MP^{OSM} were modestly decreased with down-regulation of *OSM* in *TTN* ventricles (fig. S28A), which would attenuate the

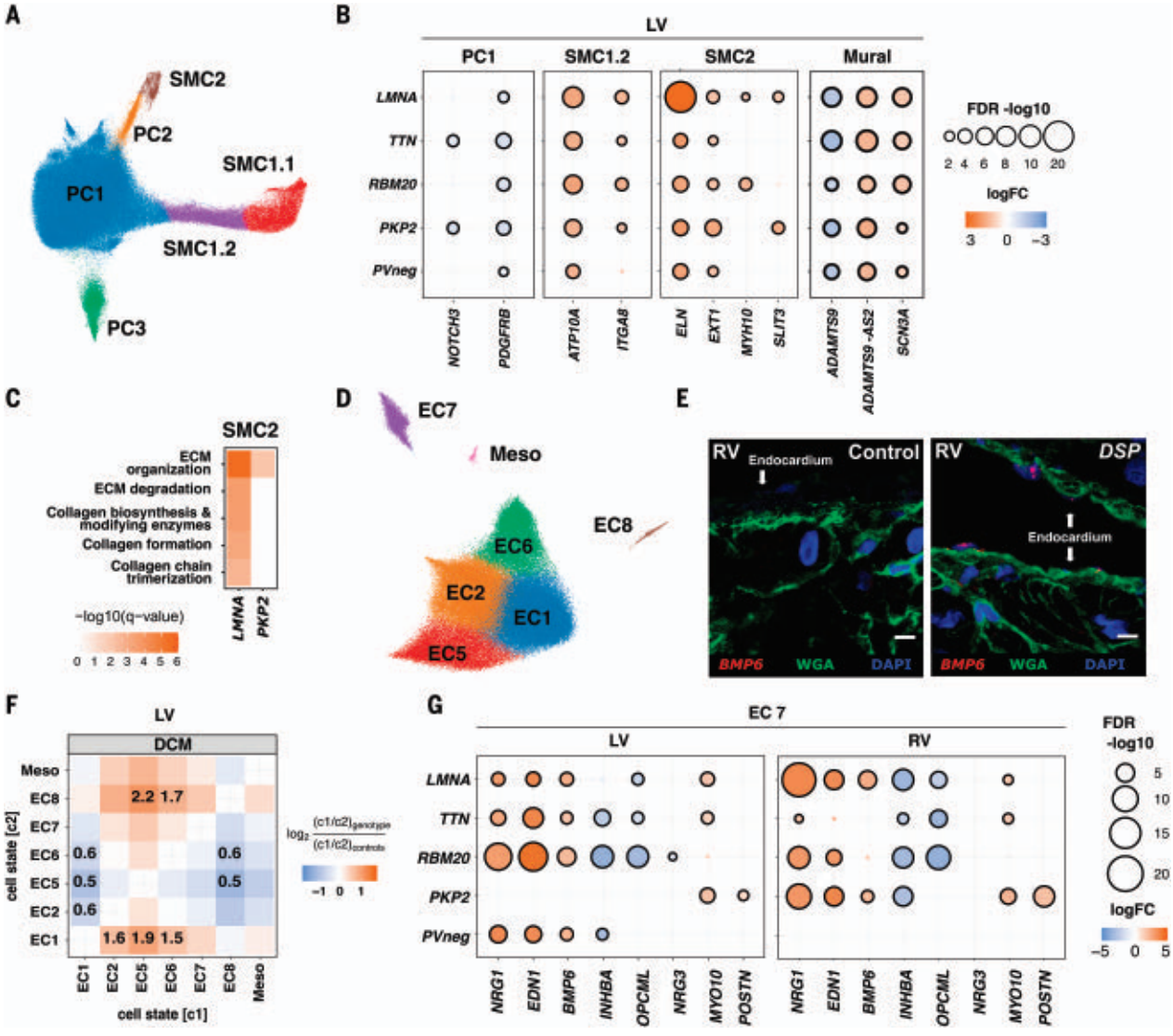


Fig. 3. Mural and endothelial cell states in control, DCM, and ACM hearts. (**A**) UMAP depicting PC and SMC states in all tissues. (**B**) Dot plots illustrating the levels (fold-change; logFC) and significance [−log10(FDR)] of selected DEGs in LV PC1, SMC1.2, SMC2, and MC (PC and SMC) across genotypes. (**C**) KEGG pathway analysis of DEGs in SMC2 among genotypes with one or more enriched pathways. Color intensity denotes enrichment significance [−log10(FDR)]. (**D**) UMAP depicting EC states in all tissues. (**E**) Single-molecule RNA fluorescent in situ hybridization identified *BMP6* (red) expression in the RV endocardium from a heart with a *DSP* (desmoplakin) PV compared with control RV. Cell boundaries are stained with WGA (green); nuclei are stained with DAPI (blue). Scale bar, 10 μm. (**F**) Pairwise cell state abundance ratios in DCM LVs relative to controls. Proportional changes are scaled by color: Red indicates an increase and blue a decrease in disease versus control. *P* values are indicated for significant proportional changes (FDR < 0.05). (**G**) Dot plots illustrating LV and RV levels (fold-change; logFC) and significance [−log10(FDR)] of selected DEGs in EC7 across genotypes. Dot size and color are as defined in (B).

MP^{OSM}-vFB3 signaling axis (*4*), and mirrored the decreased OSM pathway activation score observed in vFB3 (fig. S14B).

Prominent antigen-presenting activities were evident in MO^{VCAN}, MO^{CD16}, cDC1, and cDC2, and also occurred in MP^{FOLR2}, akin to tumor-associated MPs (*54*). Across antigen-presenting MPs, *RBM20* LVs showed the highest presentation of antigens based on MHCII genes (fig. S28, B and C) and more abundant cDC2 compared with the other genotypes (fig. S27). *PKP2* LVs up-regulated MP^{ISG} (fig. S27) with interferon-stimulated genes (*55*), perhaps contributing to inflammatory *PKP2* phenotypes (*56*).

Lymphoids

We classified 15 lymphoid cell states, including T and natural killer (NK) cell subsets, innate lymphoid cells (ILCs), B cells, plasma cells, and proliferating lymphoids (Fig. 4D, figs. S29 to S32, and table S46). Our experimental design did not enrich for immune cells, and few (<40) genotype-specific DEGs were identified (fig. S29). Proliferating lymphoids were rare, and their abundance was unchanged in disease (figs. S30 and 32, A to D, and tables S47 to S52).

The cardiac complexity of $CD4^+$ and $CD8^+$ T cells included naive ($CD4T^{naive}$), activated ($CD4T^{act}$), regulatory ($CD4T^{reg}$), cytotoxic ($CD8T^{cytox}$), transitional ($CD8T^{trans}$), terminal effector ($CD8T^{te}$), and effector memory ($CD8T^{em}$) cells (fig. S31B) (*57*). We detected increased $CD4T^{act}$ only within DCM samples (fig. S32, A to D). However, DEGs indicated lymphocyte activation (cytokines *IFNG*, *CCL3*, and *CCL4* and signaling molecules *CBLB*, *FYN*, and *TXNIP*), and maturation (cell surface

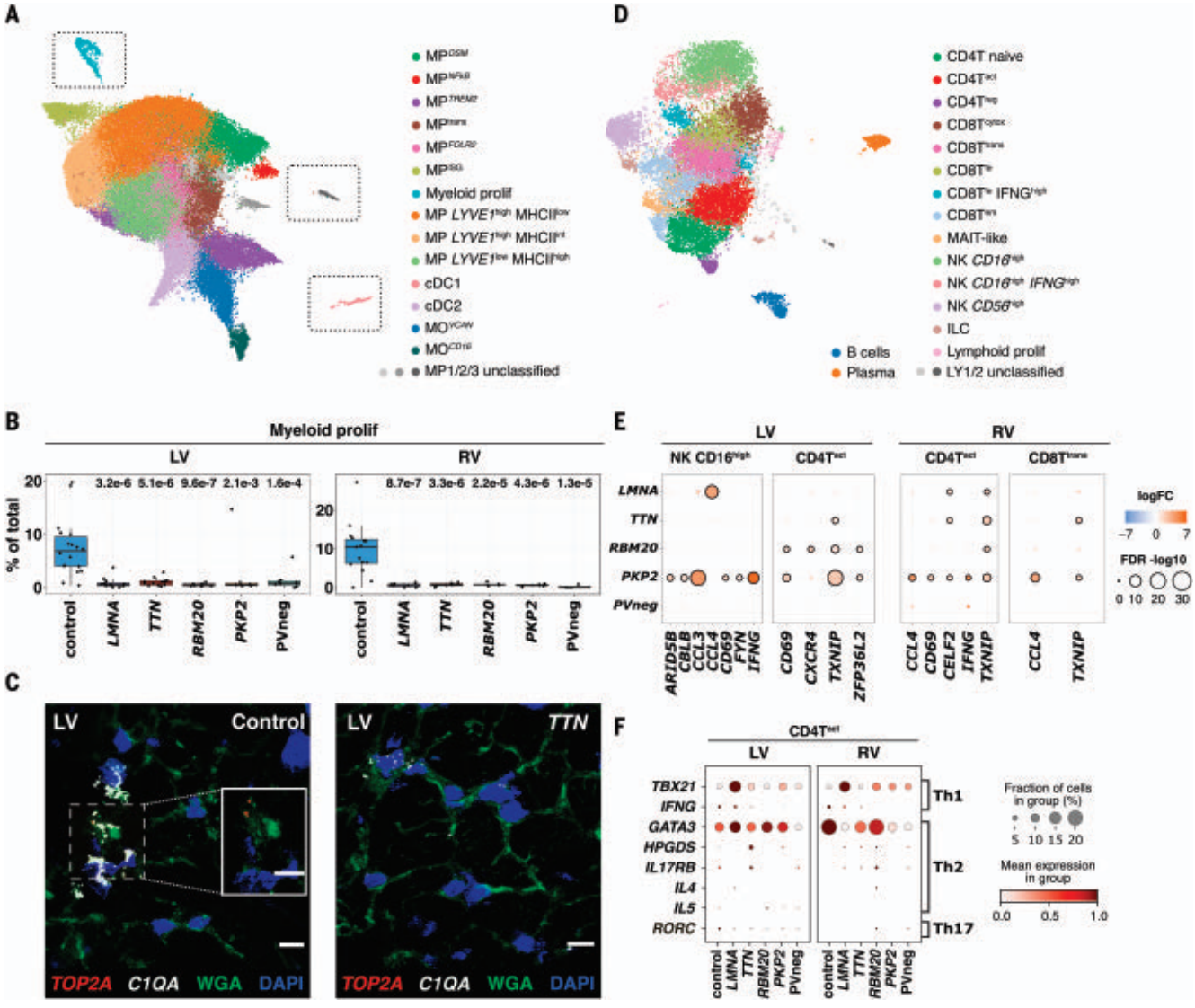


Fig. 4. Immune cell states in control, DCM, and ACM hearts. (**A**) UMAP depicting myeloid states in all tissues. Unclassified MP1, MP2, and MP3 require future characterization. Gray boxes enclosing proliferating (prolif) MPs, unclassified MPs, and cDC1s indicate that these were manually moved toward other states for ease of representation. The unmodified UMAP is in fig. S23. (**B**) Myeloid proliferation had higher abundance (% total myeloids) in controls versus disease. *P* values indicate significant differences in abundances. (**C**) Single-molecule RNA fluorescent in situ hybridization validated increased expression of *TOP2A* (red) and *C1QA* (white) in controls versus disease. Cell boundaries are stained with WGA (green); nuclei are stained with DAPI (blue). Scale bar, 10 μm. (**D**) UMAP depicting lymphoid cell states in all tissues. Unclassified LY1 and LY2 require future characterization. (**E**) Dot plots showing the level of fold-change (logFC) and significance [−log10(FDR)] of selected genes in LV NK CD16^{hi}, LV and RV CD4T^{act}, and RV CD8T^{trans} across genotypes. (**F**) Dot plots highlighting gene expression of the Th1, Th2, and Th17 signatures in CD4T^{act} cells. Dot size indicates the fraction (%) of expressing cells; color indicates the mean expression.

receptors *CD69* and *CXCR4*), particularly in *PKP2* NK^{-} cells, CD4^{+} T cells, and CD8^{+} T cells (Fig. 4E).

CD4^{+} helper T cells are critical drivers in the pathogenesis of cardiomyopathy and myocarditis (*58*, *59*). In *LMNA* CD4T^{act} cells, we observed up-regulation of *TBX21*, which is important for Th1 polarization. Conversely, the Th2-polarization transcription factor *GATA3* was down-regulated in PVneg (Fig. 4F).

Neuronal cells

Analysis of cardiac NCs was limited by rarity of this cell type (Fig. 5A, fig. S33, and tables S53 to S57). Across all NC states, DEGs indicated increased *NFATC2* in LVs with PVs and *PKP2* RVs, and genotype-selective enrichment of *LRRK2*, an activator of a neurotoxic cascade (*60*). Other up-regulated DEGs function in proteoglycan synthesis for neuronal myelination and axon regeneration (*XYLT1* and *HS3ST4*), and a complement inhibitor (*SUSD4*) that affects neural function and morphology (*61*, *62*). Genotype-specific DEGs were highest in *PKP2*.

We identified previously described NC states (*4*) and three new ones (Fig. 5, B and C, and figs. S34 and S35) that were genotype and chamber selective: NC1.1 in *LMNA* and *TTN* LVs; NC1.2 in *TTN* and *RBM20* LVs and in *LMNA*, *TTN*, and *PKP2* RVs. NC1.1 was distinguished by the highest up-regulation of *NFATC2*. NC1.2 up-regulated genes associated with electrocardiogram intervals (*SLC35F1* and *AJAP1*; Fig. 5D), *IGFBP5*, involved in neuronal apoptosis and autophagy, and the ion channel and heart rate regulator *KCNK1* (*63*–*65*). NC1.3 was enriched for the neuromodulator receptor *GALR1* and phosphodiesterases *PDE10A* and *PDE3B*, which participate in neuroprotection and signaling (*66*–*68*). Dysregulated expression of genes involved in stress responses and

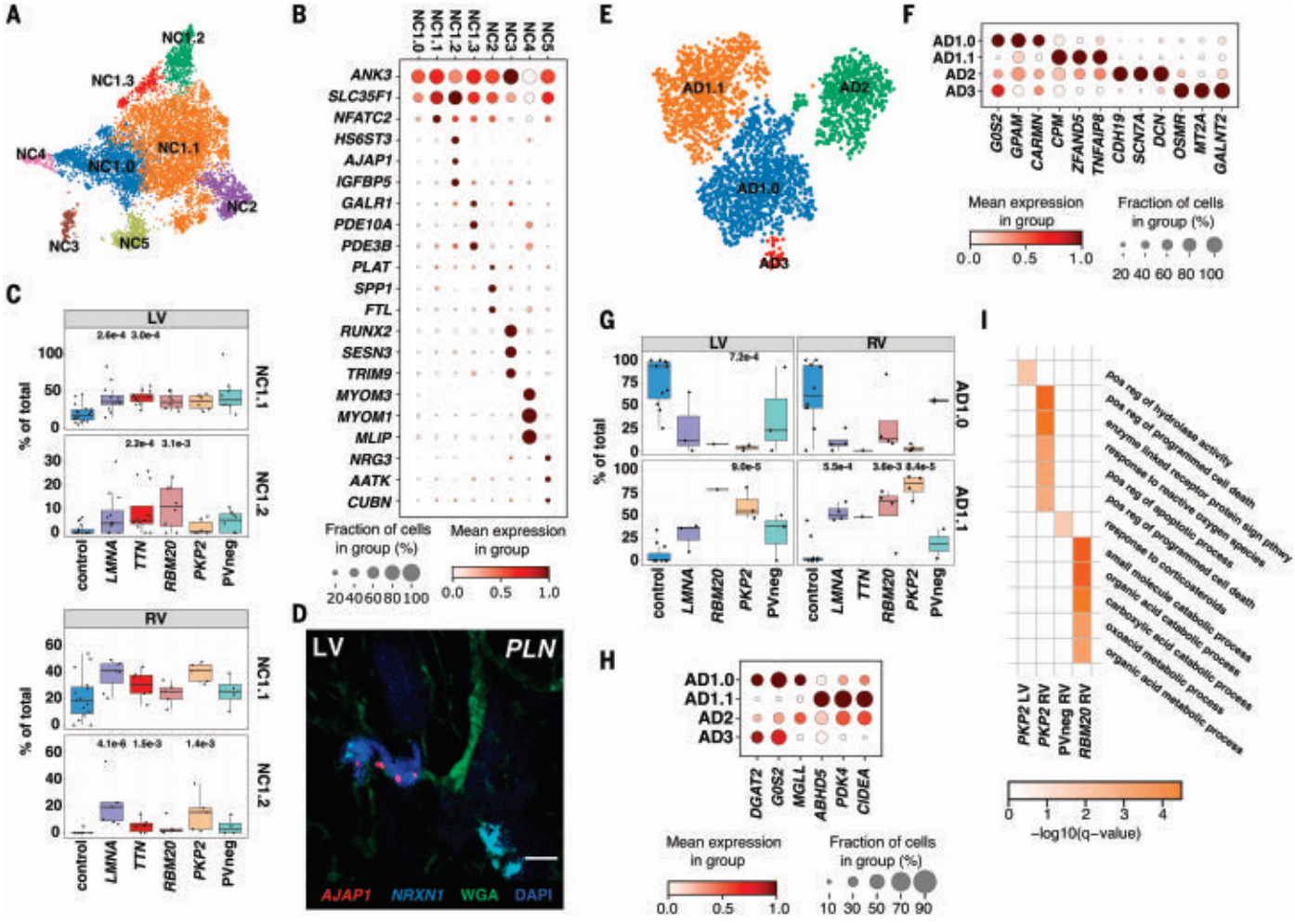


Fig. 5. Neuronal and adipocyte cell states in control, DCM, and ACM hearts. (**A**) UMAP depicting NC states in all tissues. (**B**) Dot plot highlighting the top marker genes for NC states. (**C**) LV and RV abundance analyses of NC1.1 and NC1.2 in controls versus disease. *P* values are indicated for significant proportional changes (FDR ≤ 0.05). (**D**) Single-molecule RNA fluorescent in situ hybridization showing colocalization of *AJAP1* (red) and *NRXN1* (cyan) in disease as exemplified in a DCM LV with a PV in *PLN* (phospholamban) demarcating the NC1.2 state. Cell boundaries are stained with WGA (green); nuclei are stained with DAPI (blue). Scale bar, 10 μm. (**E**) UMAP depicting AD states in all tissues. (**F**) Dot plot highlighting the top marker genes for AD states. (**G**) LV and RV abundance analyses demonstrated decreased AD1.0 and increased AD1.1 in disease versus control. (**H**) Dot plot of DEGs showing expression differences between AD states. (**I**) Heatmap of significantly enriched Gene Ontology Biological Processes terms based on up-regulated genes in disease versus control ADs. Dot size indicates the fraction (%) of expressing cells; color indicates the mean expression level. *P* values indicate significant differences (FDR ≤ 0.05).

electrophysiology may account for characteristic life-threatening arrhythmias in DCM and PVs in *PKP2* (*1*, *2*).

Adipocytes

Similar to NCs, limited adipocytes were captured. We identified three adipocyte states previously found in controls: canonical AD1.0, expressing lipid metabolism genes; stromal AD2, expressing ECM genes; and immune AD3, expressing *OSMR* and cytokine-responsive genes (Fig. 5, E and F, fig. S36, and tables S60 to S64). Although AD3 was not detected in disease, a fourth identified state, AD1.1, was almost exclusive to diseased hearts (Fig. 5G and fig. S37). Compositional analysis identified increased proportions of AD1.1 in *PKP2* LVs and RVs concurrent with decreased LV proportions of AD1.0. *LMNA* and *RBM20* RVs, unlike LVs, also increased proportions of AD1.1 (fig. S38, A to D).

DEGs between AD1.1 and AD1.0 revealed changes in fatty acid metabolism pathways (Fig. 5H). AD1.1 showed down-regulation of *DGAT2*, encoding a triglyceride-forming enzyme (*69*), and *G0S2* and *MGLL*, encoding lipolysis regulators (*70*). Conversely *ABHD5*, a positive lipolysis regulator (*70*, *71*); *PDK4*, a kinase promoting the shift from glucose to fatty acid metabolism; and *CIDEA*, a regulator of adipose tissue energy expenditure, were up-regulated. *PKP2* RVs, which typically display fibrofatty replacement, also showed an enrichment of Gene Ontology biological processes for apoptosis and cell death (Fig. 5I). These data implied genotype-specific state transitions or replacement of canonical adipocytes in DCM and ACM.

Differential expression of GWAS genes in cardiomyopathies

Genome-wide association studies (GWAS) have identified common genetic variants associated with DCM. We selected candidate genes from 15 previously identified DCM loci (table S67) and examined expression across cell types and genotypes (fig. S39). Overall, GWAS genes were more often DEGs in LV and RV than expected by chance in snRNAseq data [LV: odds ratio (OR) = 7.0, $P = 0.0007$; RV: OR = 6.1, $P = 0.0009$, one-sided Fisher's exact test). Multiple genes showed cell-type-specific expressions, with the majority highly enriched in CM (*ALPK3*, *BAG3*, *FHOD3*, *FLNC*, *HSPB7*, *MLIP*, *SLC6A6*,

SMARCB1, *SVIL*, and *TTN*). Among these, we observed both genotype- and chamber-specific expression differences. The heat shock protein *HSBP7* was reduced in TTN LVs and in *LMNA* and *PKP2* RVs. *SLC6A6*, a taurine and amino acid transporter with cardioprotective effects, was increased in both LV and RV across all genotypes except for *PKP2* and PVneg. *CDKN1A*, a cell cycle regulator and modulator of apoptosis (*72*), was increased in LMNA CMs from LVs but not RVs (*72*).

GWAS genes that were more broadly expressed across cell types included *MTSS1*, encoding a putative actin-cytoskeletal interactor. *MTSS1* showed highest and unchanged expression in myeloids and was widely increased in mural cells as well as in FBs in *LMNA*, *TTN*, and *PKP2* (fig. S39), suggesting influences beyond direct effects on CM function (*73*, *74*). *SLC39A8*, a lowly expressed cardiac solute carrier, was unchanged in CMs but increased in *LMNA* and *TTN* LV ECs. We suggest that cell-specific expression changes of GWAS genes may improve interpretation of their biologic effects.

Predicted and altered cell-cell interactions across genotypes

By examining the expression of genes encoding for receptors and ligands, we inferred intercellular signaling and communication (*75*). We initially quantified the probability of cell-cell interactions and compared signaling between cell states, and then aggregated information to produce cell-specific and across-all-cell-types data for each genotype relative to controls. This sequential approach accounted for differential abundances of cell states.

We detected aberrant intercellular signaling across disease (Fig. 6A and fig. S40), including up-regulation of the BMP, FN1, collagen, EGF, IGF, and TGF pathways that promote fibrosis. Signaling dependent on VEGF, NOTCH, and ANGPT was also increased in disease, implying vascular remodeling. Genotype-selective increases in intercellular signaling pathways were also identified in LVs (Fig. 6B and fig. S40, A and B): EDN in *LMNA*, the proinflammatory IL6 in *TTN*, BAFF/LIGHT (denoting TNF signaling) in *RBM20*, pro-inflammatory CCL and TNF in *PKP2*, and the immune modulator BTLA in PVneg. Some of these intercellular signaling pathways were similarly dysregulated in RV, but chamber-specific changes were also observed (fig. S40, C and D, and fig. S41).

We also identified genotype-specific differences in the cells sending and receiving signals

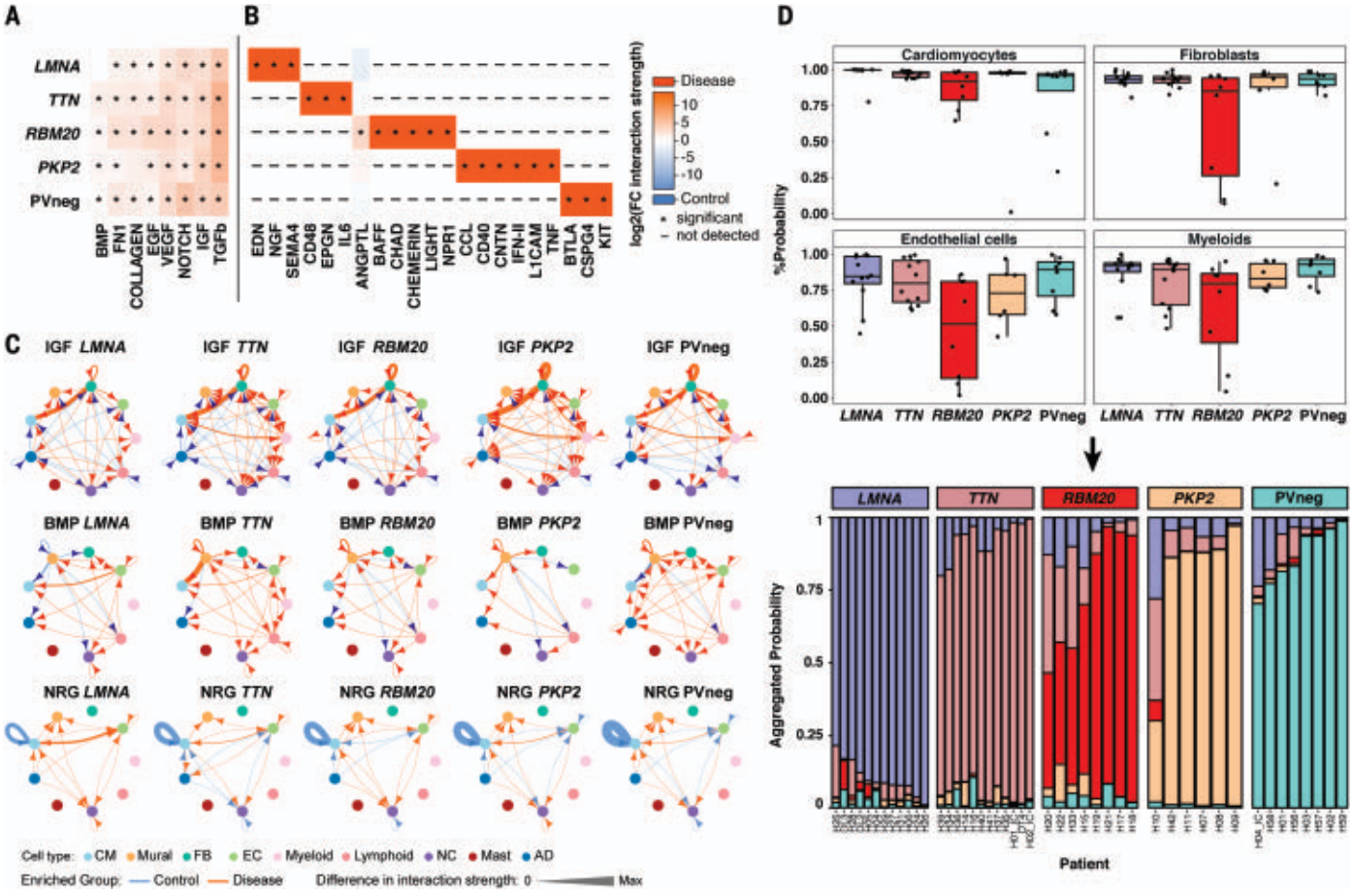


Fig. 6. Altered cell-cell interactions and recognition of genotype-specific transcriptional responses. Heatmaps depict shared (**A**) and unique (**B**) signaling pathways in LVs, with significantly different expression in genotypes compared with controls. Signaling pathways are defined in the CellChat database (*75*). Changes in interaction strength (log2FC; table S68), scaled by color intensity (red, increased; blue, decreased). Asterisks denote statistical significance (adjusted *P* value ≤ 0.05); dashes indicate that expression was not detected in control or disease. (**C**) Circle plots of significant (adjusted *P* value ≤ 0.05) cell-cell communication depicting differentially regulated IGF, BMP, and NRG pathways and interactions in disease LVs. The line thickness denotes interaction strength of signals from sending and receiving cells; color is scaled from zero to maximum in disease versus controls (orange, increased; blue, decreased). Arrows indicate directionality. (**D**) Top: Genotype prediction probability from GAT per cell type. Bottom: Stacked bar plots representing the likelihood (% aggregated probability) of individual patient genotypes by GAT prediction. Most established genotypes were predicted with high probability, with lower prediction probability only in H10 (*PKP2*), H20 (*RBM20*), H22 (*RBM20*), and H33 (*RBM20*).

in disease pathways (Fig. 6C). For example, IGF signaling (fig. S42), which is crucial in cell growth and CM hypertrophy, showed increased FB autocrine (highest in *PKP2*) and paracrine FB to CM signaling, paralleling findings in experimental models (*76*, *77*). In addition, IGF signaling from myeloid cells to CMs only occurred in *TTN*, *PKP2*, and PVneg, an interaction that might promote muscle repair (*78*).

The source of BMP signaling changed in a genotype-specific manner (Fig. 6C and fig. S43). BMP signaling from MC to CM was increased in *TTN* LVs and RVs, but was down-regulated in PVneg and *LMNA* LVs and other RVs (figs. S41 and S43). BMP signaling in ECs originated solely from EC7, likely depending on *BMP6* up-regulation (Fig. 3, E and G). EDN signaling from EC7 to CM and MC was highly genotype selective, occurring in *LMNA* LV and *PKP2* RV (fig. S44).

NRG signaling (comprising *NRG1-3* and *ERBB* receptors) showed multifaceted changes. Disease LVs markedly attenuated autocrine NRG signaling in CMs, whereas EC and CM signaling was up-regulated (highest in *LMNA*) in all genotypes except *RBM20* (Fig. 6C and fig. S45). Additionally, NRG3-ERBB4 interactions identified in controls shifted in disease to NRG1-ERBB4 and NRG1-ERBB3 in a genotype-specific manner, consistent with changes in *NRG1/3* expression in EC7 (Fig. 3G and fig. S46) (*79*). This predicted NRG/ERBB shift may provide compensatory responses to adverse remodeling in cardiomyopathies (*48*).

Graph attention networks recognized genotype-specific expression patterns

We applied machine-learning approaches to snRNAseq data to further advance the recognition of cell- and genotype-specific transcriptional patterns. Cell-specific neighborhood graphs showed more connectivities among single-nuclei transcriptomes from PVneg hearts and hearts with PVs in the same gene compared with PVs from different genes (fig. S47). Subsequently, we generated a graph attention network (GAT) for multinomial classification of genotypes trained on four major informative LV cell types: CMs, FBs, ECs, and myeloids (fig. S47, A and B). The GAT predicted *LMNA*, *TTN*, *RBM20*, *PKP2*, and PVneg genotypes with high accuracy. Among LV samples, the genotype prediction accuracy differed by cell type: CMs, 0.93; FBs, 0.92; myeloids, 0.85; and ECs, 0.79 (Fig. 6D, fig. S48, and tables S69 to S70; corresponding RV data are shown in fig. S47C). Aggregation of genotype predictions obtained from these four LV cell types strengthened the correct prediction of genotypes, resulting in a high confidence model (Fig. 6D). Three (H10, H22, and H33) of the four lower prediction probabilities occurred in samples with both a primary and secondary PV, assigned as such by prior genotype and clinical phenotype review (table S1). Moreover, because this machine-learning model independently confirmed a genotype- and cell-type-specific transcriptional signature, we concluded that these snRNAseq datasets accurately described the molecular responses to PVs and unexplained causes of DCM and ACM.

Discussion

Our analyses of snRNAseq of LV and RV samples illuminated the cell types and states, molecular signals, and predicted intercellular communications that characterized DCM and ACM. Compared with control hearts, we identified differences at multiple levels, including changes in the proportions of cell types and states, additional cell states, and differential gene expression, substantially expanding earlier insights achieved by bulk tissue analyses. Across all genotypes, diseased hearts demonstrated some common dissimilarities from control hearts, often with graded differences between LVs and RVs. Despite studying hearts from patients with advanced disease who received diverse therapies, congruent transcriptional signatures emerged for different PVs within the same gene and varied between genotypes.

The differences between genotype groups and controls reflected differences in mean expression and not differences in variance. This was true for each genotype group, including PVneg. Transcriptional signatures were complex, diversified both by the proportions of canonical and stressed cell states and by differentially expressed genes within the same states. Although interrogation of these datasets provides ongoing opportunities for discovery, our findings provided substantial evidence that genotype influenced pathological remodeling of the heart. These results upend a prevalent dogma that heart failure results from a common final pathway and can guide the future development of therapies with selective targets to enhance personalized medicine.

Despite anatomical and histopathological differences between DCM and ACM, we identified shared changes in the cellular composition of ventricular tissues, albeit with DCM LV features largely mirrored in ACM RVs. Cardiomyopathies were depleted in CMs, whereas EC and immune cell populations were increased. FBs did not expand, but increased states that augmented ECM gene expression and collagen deposition. Based on these changes and cytokine profiles (TGFβ activation, increased IGF signaling, decreased *CCL2* expression; Fig. 2, H and I, and 6C), we predicted cell-cell interactions and key molecules that are appropriate for mechanistic studies to causally link differential expression with adverse cardiac remodeling.

Disease CMs exhibited loss of the canonical state vCM1.0 and the emergence of genotype-enriched cell states with DEGs. Many of these responses are associated with stress-induced contractile, metabolic, and electrophysiologic properties that are prominent clinical manifestations of some genotypes. For example, attenuated expression of *SMYD1*, a critical organizer of sarcomere structure and epigenetic and metabolic remodeling, occurred only in CMs with PVs and was unaltered in PVneg samples. With its global impact on myocardial function, dysregulation of *SMYD1* might contribute to earlier presentation and poorer outcomes of cardiomyopathy patients with versus without PVs (*80*). *LMNA* CMs had the greatest expansion of vCM1.2, enriched for genes encoding Ca^{2+} regulators and molecules with electrophysiologic functions, whereas *RMB20* and *LMNA* genotypes had the highest expression of *MYL4*, a sarcomere protein associated with atrial fibrillation. These data suggested molecular mechanisms whereby particular genotypes convey increased risks for arrhythmias and sudden cardiac death in patients.

Shifts in FB states explained the paradoxically increased fibrosis in cardiomyopathies, without expansion of overall FB populations. DCM LVs and ACM RVs showed increased proportions of vFB2, enriched for genes that modulate ECM composition, turnover, stiffness, and fibrotic scarring (*81*), and reciprocally decreased proportions of vFB3 that express transcripts suppressing fibrosis. Fibrogenic genes within vFB2 were highly expressed in *RBM20*, a genotype within our cohort with the poorest ventricular function and youngest age for heart failure diagnosis and cardiac transplantation. Strategies to manipulate proteins encoded by these genes may attenuate the prominent myocardial fibrosis that characterizes cardiomyopathies.

Unexpectedly, mural cells (PCs and SMCs) showed no increase compared with ECs across cardiomyopathies. DEG analysis suggested that these cells promoted vascular remodeling and dysfunction. PCs diminished *PDGFRB* expression and showed aberrant NOTCH signaling in vascular beds, whereas genotype-selective DEGs in SMCs up-regulated contractile genes, augmenting fiber formation. Together, these molecular signals may contribute to the microvascular dysfunction that occurs in cardiomyopathy patients (*82*) and adversely influences ventricular performance.

Among EC, EC7 demarcated the endocardium and had the most DEGs in disease compared with control. Little is known about this myocardial layer in heart failure pathogenesis. The endocardium forms through dynamic regulation of NOTCH, neuregulin (*83*), and BMP (*84*) signals. Pediatric heart diseases can exhibit pathological expansion of the endocardium (denoted endocardial fibroelastosis), which diminishes cardiac performance and is associated with progression to heart failure (*84*). Our data suggested similar dysregulation of

these molecular pathways in adult-onset cardiomyopathies with increased *BMP6* and *NRG1* in EC7 and decreased *INHBA* (*53*). These signals may cause pathological changes in endocardium in DCM and ACM hearts and contribute to myocardial dysfunction.

Diseased hearts had increased myeloid cells. Expansion of immune cell populations might arise from recruitment of circulating cells or proliferation of resident cells. In our samples, proliferating myeloid cells decreased, as was previously observed in other DCM studies (*85*). Myeloid recruitment through the CCR2/CCL2 axis, which is primarily mediated by vFB3, was decreased in all samples, with particularly dysregulated fibroblast-to-macrophage interactions in *TTN*. Although distinguishing between proliferation versus recruitment late in disease was problematic, previous studies indicated that peak cytokine expression precedes the emergence of heart failure (*86*, *87*). Analyses of earlier time points during disease progression will be important to discern these key signals.

We used machine-learning strategies to confirm and expand the conclusion that genotype-specific signals meaningfully contribute to disease pathogenesis. Using GAT to classify patients' genotypes from each cell type, we found that CMs, FBs, ECs, and myeloid cells provided the highest and most discriminatory information. Harnessing LV and RV data from these cell types, we independently predicted the established genotype of each patient with high accuracy. Moreover, among the four samples with the lowest genotype-predictive probability, three samples carried two PVs, indicating that the model detected subtle transcriptional differences with additional influences.

Although expanding machine-learning models to much larger datasets will undoubtedly improve accuracy, these early analyses supported the conclusion that PVs in different genes evoked cell-type- and state-specific responses that altered intercellular communications and promoted distinct disease pathways. We recognize that pathways may converge, but even in advanced disease, our data indicate that genotypes promoted specific transcriptional signals that likely contributed to distinct as well as common manifestations of genetic cardiomyopathies.

Future studies are needed to comprehensively define the molecular pathophysiology of cardiomyopathies and heart failure, including assessments of age, sex, and ancestral influences; other DCM and ACM genotypes; additional cardiac regions; and longitudinal analyses to identify initiating and secondary processes. We also expect that the deployment of strategies that upsample conduction system and other rare cell types and the incorporation of techniques to characterize the epigenome, proteome, and spatial relationships between cell types, states, and gene expression will also be highly informative. To promote these initiatives, we freely provide all datasets and an interactive platform (https://cellxgene.cziscience.com/collections/e75342a8-0f3b-4ec5-8ee1-245a23e0f7cb) with cell type and state annotations. We expect that these resources will advance mechanistic studies to improve treatment of cardiomyopathies and enable heart failure prevention strategies.

Methods summary

Detailed information on human subject studies, experimental methods, data access, codes, algorithms, and computational programs used in this manuscript is provided in the supplementary materials (*88*).

Human studies were performed using protocols that were reviewed and approved by the ethics boards of participating institutions. DCM and ACM ventricular samples were collected from genotyped patients undergoing mechanical support (n = 15) or heart transplantation (n = 31) and from deceased donors with nonfailing hearts as described previously. Nuclei from full-thickness LV and RV regions were isolated and processed for snRNAseq as described previously. Data were mapped to the human genome (GRCh38), processed to remove doublets and to identify nuclei that met high quality standards, and harmonized to remove batch effects. Manifolds were constructed using UMAPs for all and individual cell types.

Differential abundance analyses of cell types and states were performed using centered log ratio transformation including a linear model. Differential gene expression between disease and control tissues were deduced using a pseudobulk approach and EdgeR. Comparative analyses assessed cell type and cell state abundances and differential gene expression between disease samples and controls and between genotypes. Selected genes with differential expression were validated using single-molecule fluorescent in situ hybridization or quantitative immunohistochemistry.

We investigated cell-cell communication using CellChat. The expression of genes previously identified through GWAS of DCM was assessed in diseased and control tissues. We interrogated transcriptional datasets with machine-learning tools to generate cell type features that distinguished PVs within the same gene from PVs within different genes and used these data to generate a GAT. The accuracy for GAT analyses of randomly selected patient data to assign the correct, clinically assigned genotype was assessed.

REFERENCES AND NOTES

1. E. M. McNally, L. Mestroni, Dilated cardiomyopathy: Genetic determinants and mechanisms. *Circ. Res.* **121**, 731–748 (2017). doi: 10.1161/CIRCRESAHA.116.309396; pmid: 28912180
2. K. M. Austin *et al*., Molecular mechanisms of arrhythmogenic cardiomyopathy. *Nat. Rev. Cardiol.* **16**, 519–537 (2019). doi: 10.1038/s41569-019-0200-7; pmid: 31028357
3. A. L. Bui, T. B. Horwich, G. C. Fonarow, Epidemiology and risk profile of heart failure. *Nat. Rev. Cardiol.* **8**, 30–41 (2011). doi: 10.1038/nrcardio.2010.165; pmid: 21060326
4. M. Litviňuková *et al*., Cells of the adult human heart. *Nature* **588**, 466–472 (2020). doi: 10.1038/s41586-020-2797-4; pmid: 32971526
5. J. S. Warren *et al*., Histone methyltransferase Smyd1 regulates mitochondrial energetics in the heart. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E7871–E7880 (2018). doi: 10.1073/pnas.1800680115; pmid: 30061404
6. S. Franklin *et al*., The chromatin-binding protein Smyd1 restricts adult mammalian heart growth. *Am. J. Physiol. Heart Circ. Physiol.* **311**, H1234–H1247 (2016). doi: 10.1152/ajpheart.00235.2016; pmid: 27663768
7. S. F. Steinberg, Beta$_1$-adrenergic receptor regulation revisited. *Circ. Res.* **123**, 1199–1201 (2018). doi: 10.1161/CIRCRESAHA.118.313884; pmid: 30571467
8. M. R. Bristow, Mechanism of action of beta-blocking agents in heart failure. *Am. J. Cardiol.* **80** (11A), 26L–40L (1997). doi: 10.1016/S0002-9149(97)00846-1; pmid: 9412540
9. Y. Hasumi *et al*., Folliculin (Flcn) inactivation leads to murine cardiac hypertrophy through mTORC1 deregulation. *Hum. Mol. Genet.* **23**, 5706–5719 (2014). doi: 10.1093/hmg/ddu286; pmid: 24908670
10. H. Hasumi *et al*., Folliculin-interacting proteins Fnip1 and Fnip2 play critical roles in kidney tumor suppression in cooperation with Flcn. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1624–E1631 (2015). doi: 10.1073/pnas.1419502112; pmid: 25775561
11. K. Shen *et al*., Cryo-EM structure of the human FLCN-FNIP2-Rag-Ragulator complex. *Cell* **179**, 1319–1329.e8 (2019). doi: 10.1016/j.cell.2019.10.036; pmid: 31704029
12. M. Mejias *et al*., CPEB4 increases expression of PFKFB3 to induce glycolysis and activate mouse and human hepatic stellate cells, promoting liver fibrosis. *Gastroenterology* **159**, 273–288 (2020). doi: 10.1053/j.gastro.2020.03.008; pmid: 32169429
13. R. Ajima *et al*., Deficiency of Myo18B in mice results in embryonic lethality with cardiac myofibrillar aberrations. *Genes Cells* **13**, 987–999 (2008). doi: 10.1111/j.1365-2443.2008.01226.x; pmid: 18761673
14. M. Hueso *et al*., Silencing of CD40 in vivo reduces progression of experimental atherogenesis through an NF-κB/miR-125b axis and reveals new potential mediators in the pathogenesis of atherosclerosis. *Atherosclerosis* **255**, 80–89 (2016). doi: 10.1016/j.atherosclerosis.2016.11.002; pmid: 27835742
15. S.-Y. Kim, S. Yasuda, H. Tanaka, K. Yamagata, H. Kim, Non-clustered protocadherin. *Cell Adh. Migr.* **5**, 97–105 (2011). doi: 10.4161/cam.5.2.14374; pmid: 21173574
16. Y. Li *et al*., The N-cadherin interactome in primary cardiomyocytes as defined using quantitative proximity proteomics. *J. Cell Sci.* **132**, jcs.221606 (2019). doi: 10.1242/jcs.221606; pmid: 30630894
17. Y. Wang *et al*., PCDH7 interacts with GluN1 and regulates dendritic spine morphology and synaptic function. *Sci. Rep.* **10**, 10951 (2020). doi: 10.1038/s41598-020-67831-8; pmid: 32616769
18. P. D. Swaminathan, A. Purohit, T. J. Hund, M. E. Anderson, Calmodulin-dependent protein kinase II: Linking heart failure and arrhythmias. *Circ. Res.* **110**, 1661–1677 (2012). doi: 10.1161/CIRCRESAHA.111.243956; pmid: 22679140
19. L. A. Swayne *et al*., Novel variant in the *ANK2* membrane-binding domain is associated with ankyrin-B syndrome and structural heart disease in a first nations population with a high rate of long QT syndrome. *Circ. Cardiovasc. Genet.* **10**, e001537 (2017). doi: 10.1161/CIRCGENETICS.116.001537; pmid: 28196901
20. H. D. Devalla *et al*., TECRL, a new life-threatening inherited arrhythmia gene associated with overlapping clinical features of both LQTS and CPVT. *EMBO Mol. Med.* **8**, 1390–1408 (2016). doi: 10.15252/emmm.201505719; pmid: 27861123
21. V. Gaussin *et al*., Alk3/Bmpr1a receptor is required for development of the atrioventricular canal into valves and annulus fibrosus. *Circ. Res.* **97**, 219–226 (2005). doi: 10.1161/01.RES.0000177862.85474.63; pmid: 16037571
22. N. W. Morrell *et al*., Targeting BMP signalling in cardiovascular disease and anaemia. *Nat. Rev. Cardiol.* **13**, 106–120 (2016). doi: 10.1038/nrcardio.2015.156; pmid: 26461965
23. J. Filmus, M. Capurro, The role of glypicans in Hedgehog signaling. *Matrix Biol.* **35**, 248–252 (2014). doi: 10.1016/j.matbio.2013.12.007; pmid: 24412155

24. M. L. Schmitz, M. S. Shaban, B. V. Albert, A. Gökçen, M. Kracht, The crosstalk of endoplasmic reticulum (ER) stress pathways with NF-κB: Complex mechanisms relevant for cancer, inflammation and infection. *Biomedicines* **6**, 58 (2018). doi: 10.3390/biomedicines6020058; pmid: 29772680
25. M. Wilhelm *et al*., Sh3rf2/POSHER protein promotes cell survival by ring-mediated proteasomal degradation of the c-Jun N-terminal kinase scaffold POSH (Plenty of SH3s) protein. *J. Biol. Chem.* **287**, 2247–2256 (2012). doi: 10.1074/jbc.M111.269431; pmid: 22128169
26. R. Eckenstaler, J. Sandori, M. Gekle, R. A. Benndorf, Angiotensin II receptor type 1 - An update on structure, expression and pathology. *Biochem. Pharmacol.* **192**, 114673 (2021). doi: 10.1016/j.bcp.2021.114673; pmid: 34252409
27. H. Itabashi *et al*., Angiotensin II and epidermal growth factor receptor cross-talk mediated by a disintegrin and metalloprotease accelerates tumor cell proliferation of hepatocellular carcinoma cell lines. *Hepatol. Res.* **38**, 601–613 (2008). doi: 10.1111/j.1872-034X.2007.00304.x; pmid: 18452483
28. C. F. Baicu *et al*., Effects of the absence of procollagen C-endopeptidase enhancer-2 on myocardial collagen accumulation in chronic pressure overload. *Am. J. Physiol. Heart Circ. Physiol.* **303**, H234–H240 (2012). doi: 10.1152/ajpheart.00227.2012; pmid: 22610170
29. K. J. Travaglini *et al*., A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020). doi: 10.1038/s41586-020-2922-4; pmid: 33208946
30. S. Schafer *et al*., IL-11 is a crucial determinant of cardiovascular fibrosis. *Nature* **552**, 110–115 (2017). doi: 10.1038/nature24676; pmid: 29160304
31. S. Nandadasa, C. M. Nelson, S. S. Apte, ADAMTS9-mediated extracellular matrix dynamics regulates umbilical cord vascular smooth muscle differentiation and rotation. *Cell Rep.* **11**, 1519–1528 (2015). doi: 10.1016/j.celrep.2015.05.005; pmid: 26027930
32. B.-H. Koo *et al*., ADAMTS9 is a cell-autonomously acting, anti-angiogenic metalloprotease expressed by microvascular endothelial cells. *Am. J. Pathol.* **176**, 1494–1504 (2010). doi: 10.2353/ajpath.2010.090655; pmid: 20093484
33. T. Nadeem, W. Bogue, B. Bigit, H. Cuervo, Deficiency of Notch signaling in pericytes results in arteriovenous malformations. *JCI Insight* **5**, e125940 (2020). doi: 10.1172/jci.insight.125940; pmid: 33148887
34. P. U. Magnusson *et al*., Platelet-derived growth factor receptor-beta constitutive activity promotes angiogenesis in vivo and in vitro. *Arterioscler. Thromb. Vasc. Biol.* **27**, 2142–2149 (2007). doi: 10.1161/01.ATV.0000282198.60701.94; pmid: 17656670
35. S. Jin *et al*., Notch signaling regulates platelet-derived growth factor receptor-β expression in vascular smooth muscle cells. *Circ. Res.* **102**, 1483–1491 (2008). doi: 10.1161/CIRCRESAHA.107.167965; pmid: 18483410
36. V. Domenga *et al*., Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. *Genes Dev.* **18**, 2730–2735 (2004). doi: 10.1101/gad.308904; pmid: 15545631
37. A. H. Campos, W. Wang, M. J. Pollman, G. H. Gibbons, Determinants of Notch-3 receptor expression and signaling in vascular smooth muscle cells: Implications in cell-cycle regulation. *Circ. Res.* **91**, 999–1006 (2002). doi: 10.1161/01.RES.0000044944.99984.25; pmid: 12456485
38. I. Marek *et al*., Alpha8 integrin (Itga8) signalling attenuates chronic renal interstitial fibrosis by reducing fibroblast activation, not by interfering with regulation of cell turnover. *PLOS ONE* **11**, e0150471 (2016). doi: 10.1371/journal.pone.0150471; pmid: 26938996
39. C. F. Hung, C. L. Wilson, Y.-H. Chow, L. M. Schnapp, Role of integrin alpha8 in murine model of lung fibrosis. *PLOS ONE* **13**, e0197937 (2018). doi: 10.1371/journal.pone.0197937; pmid: 29813125
40. J. J. Connelly *et al*., Epigenetic regulation of COL15A1 in smooth muscle cell replicative aging and atherosclerosis. *Hum. Mol. Genet.* **22**, 5107–5120 (2013). doi: 10.1093/hmg/ddt365; pmid: 23912340
41. L. Gong *et al*., SLIT3 deficiency attenuates pressure overload-induced cardiac fibrosis and remodeling. *JCI Insight* **5**, e136852 (2020). doi: 10.1172/jci.insight.136852; pmid: 32644051
42. R. Sewduth, M. M. Santoro, "Decoding" angiogenesis: New facets controlling endothelial cell behavior. *Front. Physiol.* **7**, 306 (2016). doi: 10.3389/fphys.2016.00306; pmid: 27493632
43. S. L. House *et al*., Endothelial fibroblast growth factor receptor signaling is required for vascular remodeling following cardiac ischemia-reperfusion injury. *Am. J. Physiol. Heart Circ. Physiol.* **310**, H559–H571 (2016). doi: 10.1152/ajpheart.00758.2015; pmid: 26747503
44. X.-B. Cui, S.-Y. Chen, Response gene to complement 32 in vascular diseases. *Front. Cardiovasc. Med.* **5**, 128 (2018). doi: 10.3389/fcvm.2018.00128; pmid: 30280101
45. Y. Xiao *et al*., Hippo signaling plays an essential role in cell state transitions during cardiac fibroblast development. *Dev. Cell* **45**, 153–169.e6 (2018). doi: 10.1016/j.devcel.2018.03.019; pmid: 29689192
46. J. Tang *et al*., Genetic fate mapping defines the vascular potential of endocardial cells in the adult heart. *Circ. Res.* **122**, 984–993 (2018). doi: 10.1161/CIRCRESAHA.117.312354; pmid: 29374073
47. R. A. Norris *et al*., Periostin regulates atrioventricular valve maturation. *Dev. Biol.* **316**, 200–213 (2008). doi: 10.1016/j.ydbio.2008.01.003; pmid: 18313657
48. G. W. De Keulenaer *et al*., Mechanisms of the multitasking endothelial protein NRG-1 as a compensatory factor during chronic heart failure. *Circ. Heart Fail.* **12**, e006288 (2019). doi: 10.1161/CIRCHEARTFAILURE.119.006288; pmid: 31607147
49. X.-S. Zhao, W. Pan, R. Bekeredjian, R. V. Shohet, Endogenous endothelin-1 is required for cardiomyocyte survival in vivo. *Circulation* **114**, 830–837 (2006). doi: 10.1161/CIRCULATIONAHA.105.577288; pmid: 16908774
50. C. K. Hathaway *et al*., Endothelin-1 critically influences cardiac function via superoxide-MMP9 cascade. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5141–5146 (2015). doi: 10.1073/pnas.1504557112; pmid: 25848038
51. R. Y. Kim, E. J. Robertson, M. J. Solloway, Bmp6 and Bmp7 are required for cushion formation and septation in the developing mouse heart. *Dev. Biol.* **235**, 449–466 (2001). doi: 10.1006/dbio.2001.0284; pmid: 11437450
52. Y. Fang *et al*., Tbx20 induction promotes zebrafish heart regeneration by inducing cardiomyocyte dedifferentiation and endocardial expansion. *Front. Cell Dev. Biol.* **8**, 738 (2020). doi: 10.3389/fcell.2020.00738; pmid: 32850848
53. A. C. Y. Chang *et al*., Notch initiates the endothelial-to-mesenchymal transition in the atrioventricular canal through autocrine activation of soluble guanylyl cyclase. *Dev. Cell* **21**, 288–300 (2011). doi: 10.1016/j.devcel.2011.06.022; pmid: 21839921
54. A. Puig-Kröger *et al*., Folate receptor beta is expressed by tumor-associated macrophages and constitutes a marker for M2 anti-inflammatory/regulatory macrophages. *Cancer Res.* **69**, 9395–9403 (2009). doi: 10.1158/0008-5472.CAN-09-2050; pmid: 19951991
55. S. A. Dick *et al*., Self-renewing resident cardiac macrophages limit adverse remodeling following myocardial infarction. *Nat. Immunol.* **20**, 29–39 (2019). doi: 10.1038/s41590-018-0272-2; pmid: 30538339
56. M. Pérez-Hernández *et al*., Transcriptomic coupling of PKP2 with inflammatory and immune pathways endogenous to adult cardiac myocytes. *Front. Physiol.* **11**, 623190 (2021). doi: 10.3389/fphys.2020.623190; pmid: 33536940
57. A. M. van der Leun, D. S. Thommen, T. N. Schumacher, $CD8^+$ T cell states in human cancer: Insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020). doi: 10.1038/s41568-019-0235-4; pmid: 32024970
58. J. G. Barin, D. Čiháková, Control of inflammatory heart disease by CD4+ T cells. *Ann. N. Y. Acad. Sci.* **1285**, 80–96 (2013). doi: 10.1111/nyas.12134; pmid: 23692566
59. D. Vdovenko, U. Eriksson, Regulatory role of $CD4^+$ T cells in myocarditis. *J. Immunol. Res.* **2018**, 4396351 (2018). doi: 10.1155/2018/4396351; pmid: 30035131
60. C. Kim *et al*., LRRK2 mediates microglial neurotoxicity via NFATc2 in rodent models of synucleinopathies. *Sci. Transl. Med.* **12**, eaay0399 (2020). doi: 10.1126/scitranslmed.aay0399; pmid: 33055242
61. N. B. Schwartz, M. S. Domowicz, Proteoglycans in brain development and pathogenesis. *FEBS Lett.* **592**, 3791–3805 (2018). doi: 10.1002/1873-3468.13026; pmid: 29513405
62. H. Zhu *et al*., The complement regulator Susd4 influences nervous-system function and neuronal morphology in mice. *iScience* **23**, 100957 (2020). doi: 10.1016/j.isci.2020.100957; pmid: 32179479
63. R. Méndez-Giráldez *et al*., GWAS of the electrocardiographic QT interval in Hispanics/Latinos generalizes previously identified loci and identifies population-specific signals. *Sci. Rep.* **7**, 17075 (2017). doi: 10.1038/s41598-017-17136-0; pmid: 29213071
64. A. H. Christensen *et al*., The two-pore domain potassium channel, TWIK-1, has a role in the regulation of heart rate and atrial size. *J. Mol. Cell. Cardiol.* **97**, 24–35 (2016). doi: 10.1016/j.yjmcc.2016.04.006; pmid: 27103460
65. X. Xu *et al*., Methamphetamine exposure triggers apoptosis and autophagy in neuronal cells by activating the C/EBPβ-related signaling pathway. *FASEB J.* **32**, 6737–6759 (2018). doi: 10.1096/fj.201701460RRR
66. P. E. Schauwecker, Galanin receptor 1 deletion exacerbates hippocampal neuronal loss after systemic kainate administration in mice. *PLOS ONE* **5**, e15657 (2010). doi: 10.1371/journal.pone.0015657; pmid: 21179451
67. V. Beaumont *et al*., Phosphodiesterase 10A inhibition improves cortico-basal ganglia function in huntington's disease models. *Neuron* **92**, 1220–1237 (2016). doi: 10.1016/j.neuron.2016.10.064; pmid: 27916455
68. Y. Mitome-Mishima *et al*., Differences in phosphodiesterase 3A and 3B expression after ischemic insult. *Neurosci. Res.* **75**, 340–348 (2013). doi: 10.1016/j.neures.2013.02.006; pmid: 23471014
69. C. Chitraju, T. C. Walther, R. V. Farese Jr., The triglyceride synthesis enzymes DGAT1 and DGAT2 have distinct and overlapping functions in adipocytes. *J. Lipid Res.* **60**, 1112–1120 (2019). doi: 10.1194/jlr.M093112; pmid: 30936184
70. M. Schweiger *et al*., G0/G1 switch gene-2 regulates human adipocyte lipolysis by affecting activity and localization of adipose triglyceride lipase. *J. Lipid Res.* **53**, 2307–2317 (2012). doi: 10.1194/jlr.M027409; pmid: 22891293
71. M. A. Sanders, H. Zhang, L. Mladenovic, Y. Y. Tseng, J. G. Granneman, Molecular basis of ABHD5 lipolysis activation. *Sci. Rep.* **7**, 42589 (2017). doi: 10.1038/srep42589; pmid: 28211464
72. K. Onoue *et al*., Cardiomyocyte proliferative capacity is restricted in mice with *Lmna* mutation. *Front. Cardiovasc. Med.* **8**, 639148 (2021). doi: 10.3389/fcvm.2021.639148; pmid: 34250035
73. M. P. Morley *et al*., Cardioprotective effects of MTSS1 enhancer variants. *Circulation* **139**, 2073–2076 (2019). doi: 10.1161/CIRCULATIONAHA.118.037939; pmid: 31070942
74. H. V. Meyer *et al*., Genetic and functional insights into the fractal structure of the heart. *Nature* **584**, 589–594 (2020). doi: 10.1038/s41586-020-2635-8; pmid: 32814899
75. S. Jin *et al*., Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021). doi: 10.1038/s41467-021-21246-9; pmid: 33597522
76. N. Takeda *et al*., Cardiac fibroblasts are essential for the adaptive response of the murine heart to pressure overload. *J. Clin. Invest.* **120**, 254–265 (2010). doi: 10.1172/JCI40295; pmid: 20038803
77. T. Horio *et al*., Production and autocrine/paracrine effects of endogenous insulin-like growth factor-1 in rat cardiac fibroblasts. *Regul. Pept.* **124**, 65–72 (2005). doi: 10.1016/j.regpep.2004.06.029; pmid: 15544842
78. J. Tonkin *et al*., Monocyte/macrophage-derived IGF-1 orchestrates murine skeletal muscle regeneration and modulates autocrine polarization. *Mol. Ther.* **23**, 1189–1200 (2015). doi: 10.1038/mt.2015.66; pmid: 25896247
79. O. Odiete, M. F. Hill, D. B. Sawyer, Neuregulin in cardiovascular development and disease. *Circ. Res.* **111**, 1376–1385 (2012). doi: 10.1161/CIRCRESAHA.112.267286; pmid: 23104879
80. L. Escobar-Lopez *et al*., Association of genetic variants with outcomes in patients with nonischemic dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **78**, 1682–1699 (2021). doi: 10.1016/j.jacc.2021.08.039; pmid: 34674813
81. P. Kong, P. Christia, N. G. Frangogiannis, The pathogenesis of cardiac fibrosis. *Cell. Mol. Life Sci.* **71**, 549–574 (2014). doi: 10.1007/s00018-013-1349-6; pmid: 23649149
82. A. Gulati *et al*., Microvascular dysfunction in dilated cardiomyopathy: A quantitative stress perfusion cardiovascular magnetic resonance study. *JACC Cardiovasc. Imaging* **12**, 1699–1708 (2019). doi: 10.1016/j.jcmg.2018.10.032; pmid: 30660522
83. G. del Monte-Nieto *et al*., Control of cardiac jelly dynamics by NOTCH1 and NRG1 defines the building plan for trabeculation. *Nature* **557**, 439–445 (2018). doi: 10.1038/s41586-018-0110-6; pmid: 29743679
84. X. Xu *et al*., Endocardial fibroelastosis is caused by aberrant endothelial to mesenchymal transition. *Circ. Res.* **116**, 857–866 (2015). doi: 10.1161/CIRCRESAHA.116.305629; pmid: 25587097
85. R. O. Ramirez Flores *et al*., Consensus transcriptional landscape of human end-stage heart failure. *J. Am. Heart Assoc.* **10**, e019667 (2021). doi: 10.1161/JAHA.120.019667; pmid: 33787284
86. X. Liao *et al*., Distinct roles of resident and nonresident macrophages in nonischemic cardiomyopathy. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4661–E4669 (2018). doi: 10.1073/pnas.1720065115; pmid: 29712858

87. O. Dewald *et al*., CCL2/Monocyte Chemoattractant Protein-1 regulates inflammatory responses critical to healing myocardial infarcts. *Circ. Res.* **96**, 881–889 (2005). doi: 10.1161/01.RES.0000163017.13772.3a; pmid: 15774854
88. The supplementary materials are available online.
89. E. Lindberg *et al*., Code for: "Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies"heiniglab/DCM_heart_cell_atlas: v1.0 release," Zenodo, Version 1 (2022); .doi: 10.5281/zenodo.6720185

ACKNOWLEDGMENTS

The authors extend their heartfelt gratitude to the patients and relatives who provided cardiac tissues for research and to the biobanks who supported these efforts: Biobank of the Heart and Diabetes Center NRW, University Hospital of the Ruhr-University Bochum, the HELP Program (Mazankowski Alberta Heart Institute) and HOPE program (University of Alberta Hospital), the Cardiovascular Research Centre Biobank at the Royal Brompton and Harefield hospitals and NHS Blood and Transplant, and the Brigham and Women's Cardiovascular Tissue Bank. We also thank H.-P. Rahn of the Flow Cytometry Technology Platform, the Systems Biology Imaging Platform (both MDC), T. Borodina of the BIH/MDC Genomics Technology Platform, and S. Schmidt (MDC) for technical help in single-nuclei isolation and library preparation; the Facility for Imaging by Light Microscopy and S. Rothery (Imperial College London) and the LMS/NIHR Imperial Biomedical Research Centre Flow Cytometry Facility for support; L. Lawrence from the Research Histology Facility at Imperial College, the Biopolymers Laboratory, P. Montero Llopis of the Microscopy Resources on the North Quad (MicRoN) core (Harvard Medical School), and G. Plummer at the Faculty of Medicine & Dentistry Cell Imaging Core (University of Alberta) for technical support; D. Keinmüller for advice on machine learning; and S. Bartlett (Brigham and Women's Hospital and Harvard Medical School) for administrative assistance. **Funding:** This project has been made possible in part by the Chan Zuckerberg Foundation (2019-202666) to N.H., M.H., M.N., C.E.S., J.G.S., and S.A.T., the Leducq Foundation (16CVD03) to N.H., C.E.S., and J.G.S.; the British Heart Foundation and Deutsches Zentrum für Herz-Kreislauf-Forschung (BHF/DZHK: SP/19/1/34461) to N. H., M.N., and S.A.T.; the Medical Faculty of the Ruhr-University, FoRUM (F842R-2015) to H.Mi., A.G. & J.G.; the Deutsche Forschungsgemeinschaft to H.Mi. (1164,6-1, and 1146,2-2), to N.H. (SFB-1470 – B03) and to D.R. (Research Fellowship); the Wellcome Trust (107469/Z/15/Z; 200990/A/16/Z) to J.S.W. and to S.A.T. (206194, 108413/A/15/D), the Medical Research Council (UK), the National Institute for Health Research (NIHR) Royal Brompton Cardiovascular Biomedical Research Unit and NIHR Imperial College Biomedical Research Centre to J.S.W. and P.J.B.; the National Institutes of Health (HL080494 and HL084553) to C.E.S. and J.G.S and to Y.K. (KL2-TR002542); the Engineering Research Centers Program of the National Science Foundation (NSF Cooperative Agreement EEC-1647837) to C.E.S. and J.G.S.; the John S. LaDue Fellowship of Harvard Medical School to Y.K. and K.B.; the Wellcome Trust Clinical PhD Fellowship to J.C.; the Overseas Research Fellowship of the Takeda Science Foundation to K.K., the Stanley Sarnoff Foundation Fellowship to E.R.N.; Boehringer Ingelheim Fonds to V.S.; the Brigham and Women's Hospital Khoury Innovation Award to Y.K.; the Ludwig Institute for Cancer Research grant, National Cancer Institute grant (U54-CA225088) and CCSP Supplemental Award to C.Y.; the British Heart Foundation (RE/18/4/34215) to J.S.W; the ERC-advanced grant under the European Union Horizon 2020 Research and Innovation Program (AdG788970) to N.H.; the University of Alberta Hospital Foundation (UHF), the Canadian Institute of Health Research (CIHR), the Heart and Stroke Foundation (HSF), and the Canada Research Chair in Heart Failure to G.Y.O.; the Choudhrie Family Fund and Howard Hughes Medical Institute to C.E.S. **Author contributions:** Conceptualization: M.H., N.H., H.Ma., H.Mi., M.N., G.Y.O., C.E.S. J.G.S.; Methodology: M.H., N.H., E.L.L., H.Ma., H.Mi, M.N., G.Y.O., D.R., C.E.S., J.G.S.; Patient recruitment, clinical data analyses, tissue procurement, patient genotyping, and biobanking: P.J.R.B., R.A.C., A.Gä., L.M., H.Mi., Y.K., B.M., M.N., G.Y.O., C.E.S., P.I.T., A.v.d.B., J.S.W.; Sample processing (tissue preparation, nuclei isolation, library prep, sequencing): R.J.B., D.R.D, L.E.F., J.M.G., A.Gh., E.L.L., Y.A.L., L.M., D.R., G.P., V.S., J.G.S, A.V., H.W. H.Z.; Data analysis: CM: D.R., C.E.S., J.G.S., FB: E.A., N.H., M.K., E.L.L., H.Ma., EC: A.M.A.M., M.N., C.L.W., Myeloid: N.H., M.K., E.L.L., H.Ma., Lymphoid: J.C., K.K., R.L., S.A.T., MC: N.H. M.K., E.L.L., H.Ma., J.R.-O., NC: E.R.N. D.R., C.E.S., J.G.S., Adipocytes: M.L., M.N., Compositional: M.H., F.S., F.J.T., J.G.S., Data interpretation: All histology, immunostaining, single-molecule in situ hybridization, microscopy: K.B., S.N.B., E.L.L., A.M.A.M., E.R.N., D.R., V.S., A.V., H.W., A.W., H.Z.; GWAS: P.J.R.B., M.H.; CellChat: M.L., A.M.A.M., M.N.; GAT: E.L.L., C.L., N.S.; Online data hosting: M.H.; Writing: E.A., N.H., E.L.L., M.L., A.M.A.M., H.Ma., E.R.N., D.R., J.R.-O., N.S., C.E.S., J.G.S.; Review and editing: E.A., P.J.R.B., N.H., E.L.L., H.Ma., H.Mi., M.N., G.Y.O., D.R., C.E.S. J.G.S.; Supervision: P.J.R.B., M.H., N.H., H.Ma., H.Mi., M.N., G.Y.O., C.E.S., J.G.S, S.A.T., F.T.; Funding: P.J.R.B., M.H., N.H., H.Mi., M.N., G.Y.O., C.E.S. J.G.S., S.A.T., J.S.W. **Competing interests:** The following authors report competing interests. Consultancy fees: J.S.W. (MyoKardia Inc, and Foresite Labs); C.E.S. and J.G.S. (Maze, BridgeBio, Bristol Myers Squibb); S.A.T. (Genentech and Roche). Scientific advisory boards: S.A.T. (Biogen, GlaxoSmithKline, Qiagen and Foresite Labs). Board of directors: C.E.S. (Burroughs Wellcome Fund U.S. and Merck). Co-founder and equity holder: S.A.T. (Transition Bio). J.S.W. acknowledges research support from MyoKardia, Bristol Myers Squibb, and Pfizer (formerly Array BioPharma). These companies were not involved in any aspect of the study's experimental design, execution, and analyses or in the preparation of the manuscript. D.R., E.L.L., H.M., A.M.A.M., A.V., N.S., A.G., E.R.N., M.L., K.K., J.R.-O., V.S., D.M.D., G.P., H.Z., A.W., C.L., Y.K., E.A., J.M.G., S.N.G., K.B., R.J.B., R.A.C., C.C., J.C., L.E.F., H.F., A.G., J.G., M.K., R.L., L.M., B.M., S.S., F.S., C.Y., C.S., P.I.T., F.J.T., A.v.d.B., H.W., C.L.W., P.J.R.B., Y.A.L., H.M., M.N., G.Y.O., M.H., and N.H. declare no competing interests. **Data and materials availability:** All sequence data generated and analyzed in this study have been deposited as BAM files at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001006374 (https://ega-archive.org/studies/EGAS00001006374). Further information about EGA can be found on https://ega-archive.org, "The European Genome-phenome Archive in 2021" (https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkab1059/6430505). Processed data can be analyzed using the Cellxgene tool website: https://cellxgene.cziscience.com/collections/e75342a8-0f3b-4ec5-8ee1-245a23e0f7cb/private and are available in h5ad format to download. Metadata sheets and patient information are available in tables S1 and S2. All code used to generate the figures in this publication are available on GitHub (https://github.com/heiniglab/DCM_heart_cell_atlas). Code used to reproduce the presented analyses is indexed on Zenodo (*89*). All scripts run on jupyter notebooks are available as ".ipynb" files, and scripts executed in command line are available as .txts or .sh files. R scripts are available as ".R". Anaconda environments are available as yml files containing information on package versions. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by the Howard Hughes Medical Institute, a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo1984
Materials and Methods
Figs. S1 to S53
Tables S1 to S71
References (*90*–*129*)
MDAR Reproducibility Checklist

Submitted 20 January 2022; accepted 1 July 2022
10.1126/science.abo1984

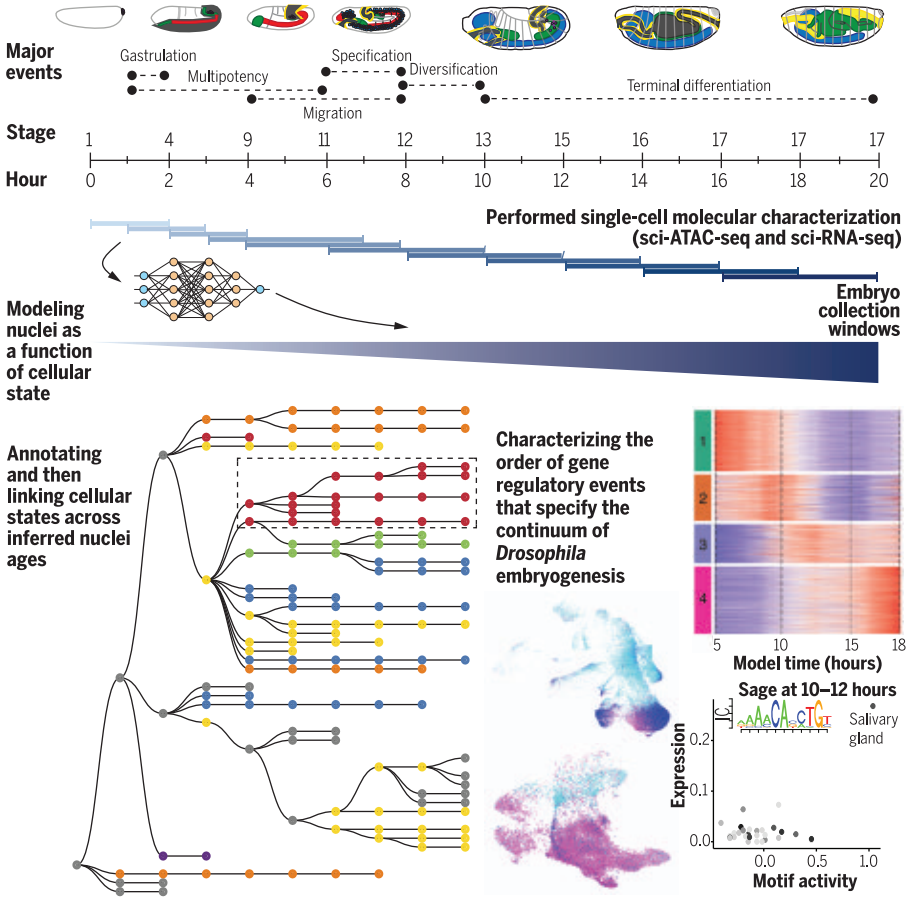
RESEARCH ARTICLE SUMMARY

DEVELOPMENT

The continuum of *Drosophila* embryonic development at single-cell resolution

Diego Calderon†, Ronnie Blecher-Gonen†, Xingfan Huang†, Stefano Secchia†, James Kentro, Riza M. Daza, Beth Martin, Alessandro Dulja, Christoph Schaub, Cole Trapnell, Erica Larschan, Kate M. O'Connor-Giles, Eileen E. M. Furlong*, Jay Shendure*

INTRODUCTION: Single-cell technologies are a powerful means of studying metazoan development, enabling comprehensive surveys of cellular diversity at profiled time points and shedding light on the dynamics of regulatory element activity and gene expression changes during the in vivo emergence of each cell type. However, nearly all such whole-embryo atlases of embryogenesis remain limited by sampling density—i.e., the number of discrete time points at which individual embryos are harvested and cells or nuclei are collected. Given the rapidity with which molecular and cellular programs unfold, this limits the resolution at which regulatory transitions can be characterized. For example, in the mouse, there are typically 6 to 24 hours between sampled embryonic time points—gaps within which massive molecular and morphological changes take place.



Characterizing the continuum of *Drosophila* embryogenesis. We collected staged *Drosophila* embryos from overlapping time windows across the first 20 hours of embryogenesis. Then we extracted nuclei and performed single-cell RNA sequencing (RNA-seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq) profiling using combinatorial indexing (sci-RNA-seq and sci-ATAC-seq) to comprehensively map expressed genes and putatively active regulatory elements. We applied machine learning to infer a continuum of nuclear ages that is synchronized across unfolding lineages in absolute time. The continuous nuclear age predictions were used to annotate and then link cellular states at nonoverlapping 2-hour intervals, as well as to explore transcriptional regulatory dynamics across major cell lineages of embryonic development at fine-scale temporal resolution.

RATIONALE: To construct an ungapped representation of embryogenesis in vivo, we would ideally sample embryos continuously. Although this is not practical for most model organisms, it is potentially possible in *Drosophila melanogaster*, where collections of timed and yet somewhat asynchronous embryos are easy to obtain, such that, at least in principle, one can achieve arbitrarily high temporal resolution. *Drosophila* could therefore serve as a test case to develop a framework for the inference of continuous regulatory and cellular trajectories of in vivo embryogenesis. Because *Drosophila* is a preeminent model organism that has yielded many advances in the biological and biomedical sciences, obtaining a single-cell atlas of *Drosophila* embryogenesis is also an important goal in itself. This includes its embryonic development, where the use of this model in conjunction with powerful genetic tools has transformed our understanding of the mechanisms by which developmental complexity is achieved, in addition to uncovering many general principles of both genetic and epigenetic gene regulation.

RESULTS: We profiled chromatin accessibility in almost 1 million nuclei and gene expression in half a million nuclei from eleven overlapping windows spanning the entirety of embryogenesis (0 to 20 hours). To exploit the developmental asynchronicity of embryos from each collection window, we applied deep neural network–based predictive modeling to more-precisely predict the developmental age of each nucleus within the dataset, resulting in continuous, multimodal views of molecular and cellular transitions in absolute time. With these data, the dynamics of enhancer usage and gene expression can be explored within and across lineages at the scale of minutes, including for precise transitions like zygotic genome activation.

CONCLUSION: This *Drosophila* embryonic atlas broadly informs the orchestration of cellular states during the most dynamic stages in the life cycle of metazoan organisms. The inclusion of predicted nuclear ages will facilitate the exploration of the precise time points at which genes become active in distinct tissues as well as how chromatin is remodeled across time. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: furlong@embl.de (E.E.M.F.); shendure@uw.edu (J.S.)
†These authors contributed equally to this work.
Cite this article as D. Calderon *et al.*, *Science* 377, eabn5800 (2022). DOI: 10.1126/science.abn5800

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abn5800

RESEARCH ARTICLE

DEVELOPMENT

The continuum of *Drosophila* embryonic development at single-cell resolution

Diego Calderon[1]†, Ronnie Blecher-Gonen[1,2]†, Xingfan Huang[1,3]†, Stefano Secchia[4]†, James Kentro[5], Riza M. Daza[1], Beth Martin[1], Alessandro Dulja[4], Christoph Schaub[4], Cole Trapnell[1,6,7], Erica Larschan[5], Kate M. O'Connor-Giles[8], Eileen E. M. Furlong[4]*, Jay Shendure[1,6,7,9]*

Drosophila melanogaster is a powerful, long-standing model for metazoan development and gene regulation. We profiled chromatin accessibility in almost 1 million and gene expression in half a million nuclei from overlapping windows spanning the entirety of embryogenesis. Leveraging developmental asynchronicity within embryo collections, we applied deep neural networks to infer the age of each nucleus, resulting in continuous, multimodal views of molecular and cellular transitions in absolute time. We identify cell lineages; infer their developmental relationships; and link dynamic changes in enhancer usage, transcription factor (TF) expression, and the accessibility of TFs' cognate motifs. With these data, the dynamics of enhancer usage and gene expression can be explored within and across lineages at the scale of minutes, including for precise transitions like zygotic genome activation.

Single-cell technologies are a powerful means of studying metazoan development, shedding light on the emergence of cellular diversity and the dynamics of gene regulation. However, nearly all such atlases of embryogenesis are limited in terms of the number of discrete time points and cells sampled per time point. Given the rapidity with which molecular and cellular programs unfold, this limits the resolution at which regulatory transitions can be characterized.

To more completely represent development, embryos would ideally be sampled continuously. Although impractical for most model organisms, it is feasible in *Drosophila*, where collections of timed and yet somewhat asynchronous embryos are easy to obtain, such that, in principle, one can achieve arbitrarily high temporal resolution. This sharply contrasts with mice, for which there are typically 6 to 24 hours between sampled time points, gaps within which massive molecular and morphological changes take place (*1*–*4*). Although sampling gaps can be computationally filled through the continuum of cell states represented in single embryos (*4*, *5*), the asynchronous ages of *Drosophila* embryos within staged collections present an opportunity for more bona fide continuity—e.g., with seconds or minutes separating the developmental ages of consecutive embryos rather than hours or days. Moreover, because *Drosophila melanogaster* is a preeminent model organism that has yielded many discoveries and general principles of metazoan development and gene regulation, obtaining a single-cell atlas of *Drosophila* embryogenesis is an important goal in itself.

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [2]The Crown Genomics Institute of the Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot, Israel. [3]Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA 98195, USA. [4]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. [5]Molecular Biology, Cell Biology, and Biochemistry Graduate Program, Brown University, Providence, RI 02912, USA. [6]Brotman Baty Institute for Precision Medicine, University of Washington, Seattle, WA 98195, USA. [7]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA 98195, USA. [8]Department of Neuroscience and Carney Institute for Brain Science, Brown University, Providence, RI 02912, USA. [9]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.
***Corresponding author. Email: furlong@embl.de (E.E.M.F.); shendure@uw.edu (J.S.)**
†These authors contributed equally to this work.

Results

We set out to measure chromatin accessibility and gene expression from individual nuclei spanning a continuum of *D. melanogaster* embryogenesis. Staged embryos were collected in 11 overlapping time windows, collectively 0 to 20 hours, covering the entirety of embryogenesis at 25°C. Overlapping 2-hour collections were used to capture the rapid transitions during early stages, followed by overlapping 4-hour collections from 3 hours onward (Fig. 1A). From each collection, samples were split and separately processed for assay for transposase-accessible chromatin using sequencing (ATAC-seq) or RNA sequencing (RNA-seq). Although we hereafter refer to cells, all data were generated from nuclei. Single-cell profiling was conducted using three-level combinatorial indexing (sci-ATAC-seq3 and sci-RNA-seq3) with minor modifications (*1*, *6*).

Sci-ATAC-seq3 and sci-RNA-seq3 libraries were sequenced to generate 30 billion and 6.8 billion raw reads, respectively (fig. S1). After deduplication and application of quality filters, we obtained chromatin accessibility profiles for 976,460 cells [single-cell ATAC (scATAC): median 5206 nonduplicate reads per cell] and gene expression profiles for 547,805 cells [single-cell RNA (scRNA): median 399 unique molecular identifiers (UMIs) and 274 genes detected per cell]. Although our scRNA data have fewer UMIs per nucleus than previously obtained from *Drosophila* embryos (*7*), we profiled many more nuclei spanning many more stages of embryogenesis and complemented this with scATAC with a high number of unique reads per nucleus. Given the small size of the *Drosophila* embryo, such deep "shotgun cellular coverage" should effectively sample all tissue types during embryogenesis. The data did not appear to be confounded by batch effects (fig. S2, A to G).

For both data modalities, integrating and visualizing single-cell profiles across all time points resulted in branching structures going from early to late stages, consistent with increasing complexity (Fig. 1, B and C). From the scATAC data, we identified 110,185 regions exhibiting accessibility at some point during embryogenesis. Collectively, these candidate regulatory elements cover 30.4 Mb (22%) of *Drosophila* euchromatin (dm6) and include 85% of known embryonic enhancers, based on overlap with nearly 5000 curated enhancers confirmed in transgenic embryos (Fig. 1D) (*8*–*10*). This, together with the high coverage of both bulk deoxyribonuclease (DNase) I hypersensitive site (DHS) peaks (87%) and scATAC-derived peaks (98%) from 2 to 12 hours (*11*, *12*), supports the comprehensiveness of this compendium. Similar results were obtained computing overlaps on a per-base rather than per-element basis (fig. S2H). We additionally uncovered more than 40,000 distal accessible regions not identified in these previous studies (Fig. 1D) that are enriched for enhancer-associated histone marks, suggesting that they are previously uncharacterized developmental enhancers (fig. S2I). The compendium also recovered 94% of 8008 extensively validated mesodermal cis-regulatory modules (*13*) and 96% of nearly 1 million chromatin immunoprecipitation (ChIP)–defined binding sites across 233 transcription factors (TFs) (*14*) (fig. S2J).

In exploring these data, we identified thousands of genomic regions and transcripts whose accessibility and expression levels, respectively, were strongly correlated with the progression of developmental time (Fig. 1, E and F). Notably, not all of these correlations were cell type specific (fig. S3). The presence of such time-dependent elements and transcripts suggests that a dynamic process is unfolding across development, at least some aspects of which are cell type specific, whereas other aspects appear general to germ layers or the entire organism. We reasoned that we could leverage these correlations to build a model to predict absolute developmental age of any

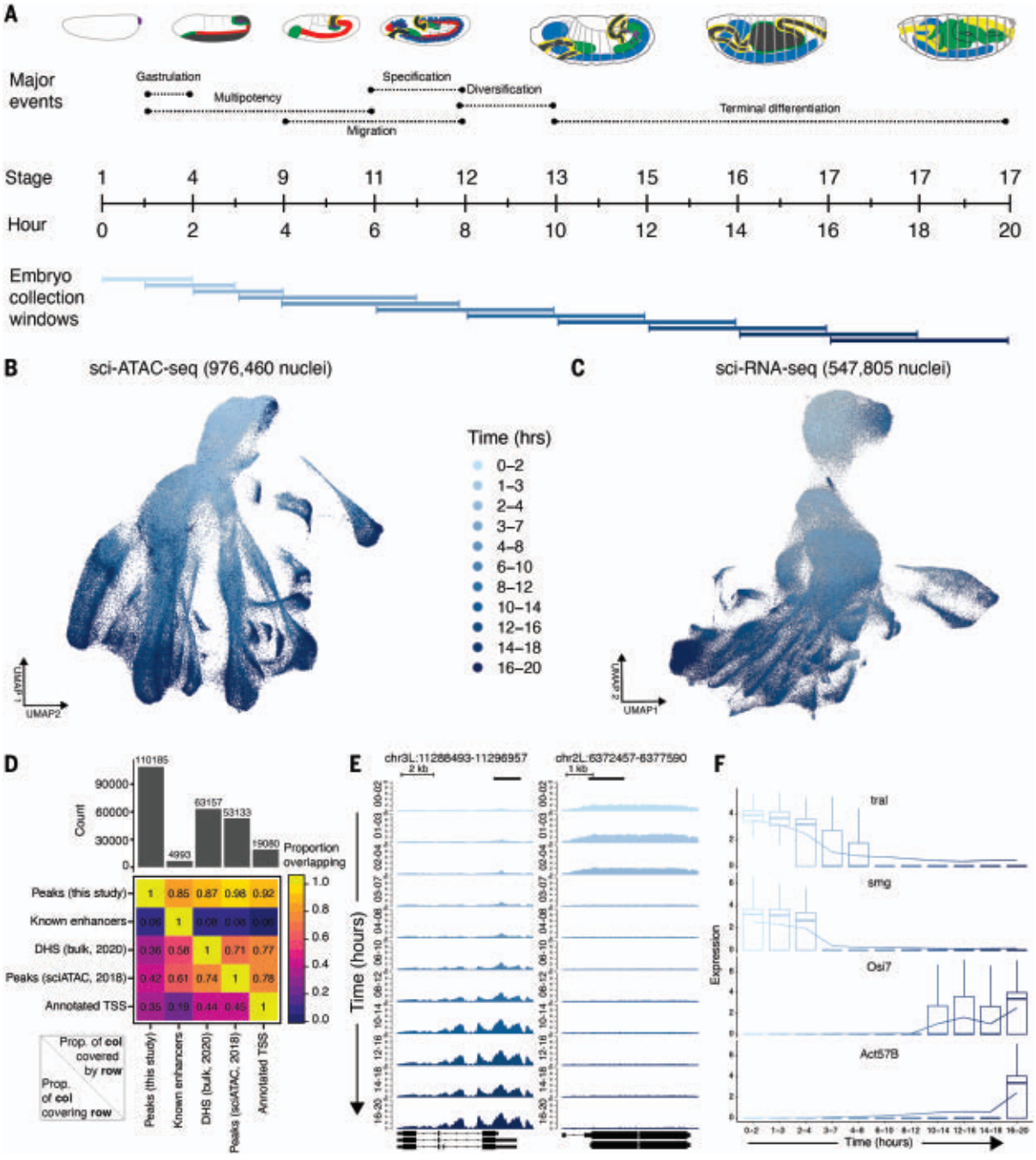


Fig. 1. Single-cell profiling of chromatin accessibility and gene expression throughout *Drosophila* embryogenesis. (**A**) Eleven overlapping collection windows that collectively cover embryogenesis. (**B**) UMAP visualization of cell-x-peak matrix of evenly time-subsampled sci-ATAC-seq nuclei that passed QC. (**C**) Same as (B), but for sci-RNA-seq. (**D**) Heatmap showing proportion of our scATAC peaks overlapping ~5000 curated enhancers (*8*–*10*), bulk DHS peaks from 2 to 12 hours (*11*), scATAC peaks from 2 to 12 hours (*12*), or annotated TSSs (*49*). (**E**) Chromatin accessibility, normalized by counts per million reads, across representative regions exhibiting time dependence across 11 collection windows. (**F**) Gene expression of representative genes exhibiting time dependence across 11 collection windows. Read counts were normalized, multiplied by a scale-factor, log-transformed after the addition of a pseudocount, and averaged across all cells within each window.

given nucleus with greater temporal resolution than our 2- to 4-hour collection windows.

Predicting the absolute age of individual nuclei

In these data, the precise developmental age of each sampled nucleus is unknown—only the 2- to 4-hour collection window from which it derived. To estimate the age of each nucleus with greater precision, we fit a series of models using either the scATAC or scRNA data as input and predicting the center hour of the collection window from which any given nucleus was obtained (Fig. 2A). Specifically, we split a subset of each dataset, evenly subsampled with respect to time, into 11 partitions, 10 of which were used as training data to fit either a lasso linear (LL) model or a neural network (NN)–based model with 10-fold cross-validation across various test parameters. After selecting the highest performing parameterization, the NN-based models markedly

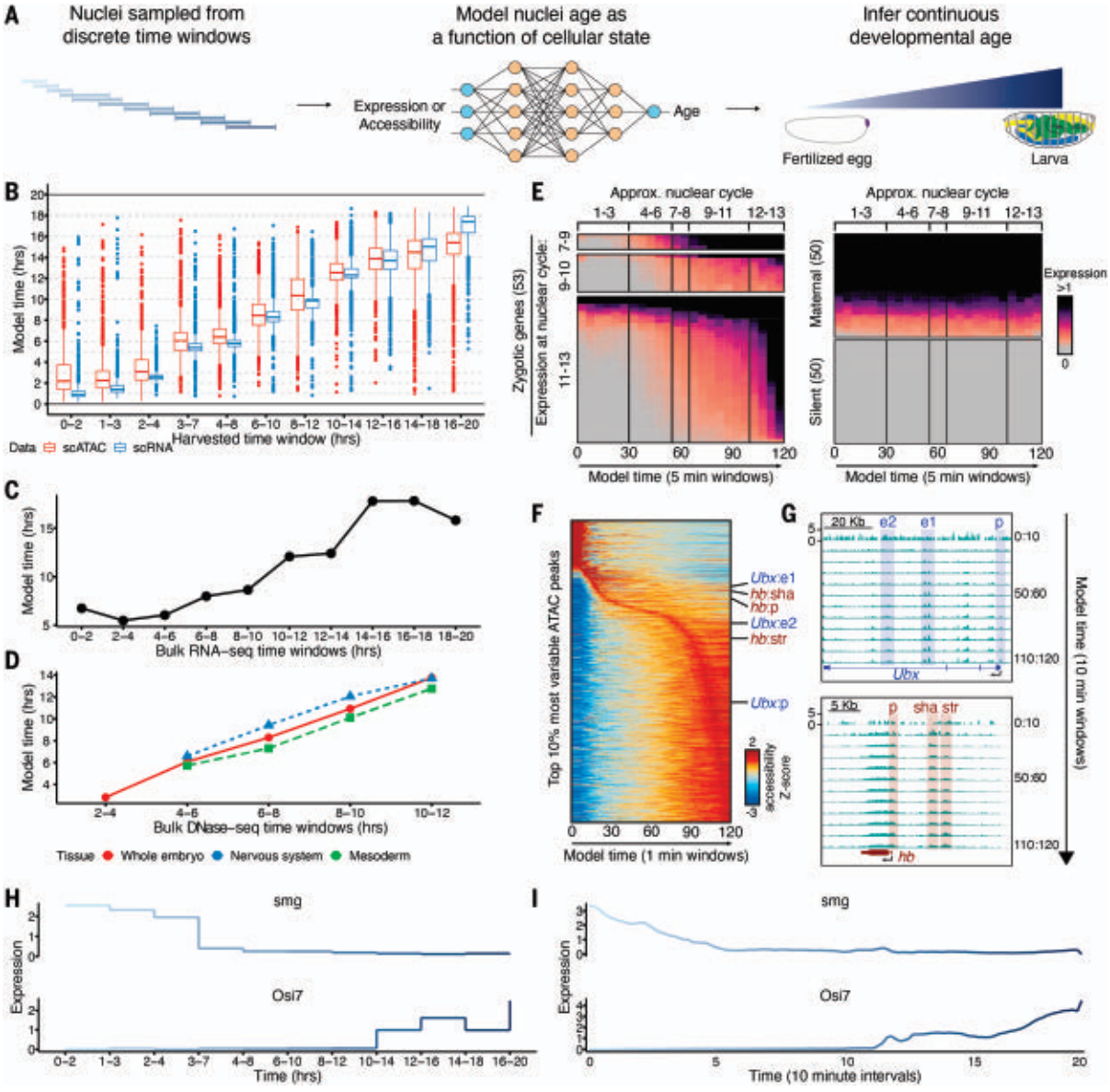


Fig. 2. Inferring developmental age from cellular state. (**A**) We fit a NN-based model that uses either gene expression or chromatin accessibility to predict the center hour of the time window from which each nucleus was sampled. The inferred nuclear ages make up a continuum. (**B**) NN model–predicted developmental ages (*y* axis) of test set nuclei, equally sampled from discrete time windows (*x* axis) and not included in model training. (**C**) NN model–predicted developmental ages (*y* axis) of bulk RNA-seq samples (*15*) collected from 2-hour windows (*x* axis). (**D**) NN model–predicted developmental ages (*y* axis) of bulk DNase-seq samples from either whole-embryo or purified tissues collected from 2-hour windows (*x* axis). (**E**) Expression of zygotic (left), maternal (top right), or silent (bottom right) genes in nuclei from predicted age windows in 5-min increments across 0 to 2 hours of development. (**F**) Accessibility of most variable scATAC peaks from predicted age windows in 1-min increments across 0 to 2 hours of development. Labels indicate regions illustrated in (G). (**G**) Examples of cis-regulatory regions known to exhibit dynamic accessibility in early embryos (*17*). (**H** and **I**) Examples of time-associated genes, with expression values averaged across all nuclei from indicated collection windows (H) or from predicted age windows in 10-min increments (I).

outperformed LL models for both data types in predicting the developmental age of nuclei within the held-out 11th partition [for NN versus LL, mean squared error (MSE): ATAC = 5.26 versus 8.8, RNA = 2.54 versus 4.72; proportion correct: ATAC = 0.67 versus 0.53, RNA = 0.87 versus 0.65]. We therefore moved forward with NN-based nuclear age predictions for the remainder of this study (Fig. 2B and fig. S4). Notably, the scRNA-based model was slightly more accurate than the scATAC-based model, likely leading to slightly older age predictions during early collection windows and slightly younger age predictions during late collection windows for scATAC ages compared with scRNA ages.

To further assess accuracy, we applied the scRNA-derived models to a bulk RNA-seq time course of staged embryos in 2-hour intervals (*15*) and found high concordance between predicted and actual developmental age (Fig. 2C). The scATAC-derived models were similarly able to order a time course of bulk DNase sequencing (DNase-seq) data from either whole embryos or specific fluorescence-activated cell sorting (FACS)–purified lineages (*11*) (Fig. 2D). To assess predicted ages at much finer time scales (minutes rather than hours), we focused on genes whose expression is activated at specific nuclear cycles during zygotic genome

activation (ZGA) (*16*). Genes turning on during ZGA were dynamically up-regulated in association with predicted nuclear ages (scRNA-based; 5-min increments), whereas maternal and silent genes were not (Fig. 2E). Early dynamically accessible enhancers and promoters could similarly be predicted (scATAC-based; 1-min increments) (Fig. 2F), opening in the same order as previously observed by bulk ATAC-seq of hand-picked embryos at 3-min intervals (Fig. 2G) (*17*). To further illustrate the value of this framework, we note that pseudobulk profiles corresponding to collection windows lead to piecewise expression dynamics (Fig. 2H). By contrast, pseudobulk profiles based on model-predicted ages yield more continuous dynamics (Fig. 2I).

Although there are similarities between the goal of our approach and the concept of pseudotime (*18*), a key advantage of inferred age is that, both in training and prediction, cells are anchored to absolute time, which enables more interpretable ordering of cellular processes as well as their synchronization across lineages. One concern is that contamination with embryos whose developmental age falls outside the collection window will have exaggerated confounding effects on early time points because older embryos contain vastly more nuclei. Consistent with this, our model predicted that 2.8% of the ~80,000 scRNA-profiled cells from 0 to 2 hours were at least 4 hours in developmental age. These older cells represent the majority of a discrete cluster in uniform manifold approximation and projection (UMAP) space (fig. S5A). Similar contamination is also observed with scATAC profiles from this early time window (12.7% of ~20,000 cells; fig. S5, B to D). Clustering and visualizing only the cells inferred to be 0 to 2 hours in age eliminates this developmentally advanced cluster (fig. S5E).

Annotation and inference of diversifying developmental trajectories

To systematically track the emergence and diversification of developmental trajectories, we used inferred ages to separately process and cluster cells from a series of 2-hour nonoverlapping time windows. Clusters were then annotated by leveraging stage-matched information on gene expression from thousands of in situ hybridizations spanning embryogenesis as well as extensive enhancer activity data (*12*, *19*, *20*) (Fig. 3, A and B).

Notably, the last few hours of the time course had reduced numbers of inferred cells (e.g., after 18 hours, 61% fewer than would be expected under uniform sampling) and fewer identified clusters (fig. S6A). We suspect that this may be the result of edge effects of the model because we also observe reduced numbers of inferred cells for the first several hours, although there they have less effect because the data from early time points lack extensive structure. For this reason, we excluded cells with an inferred age of >18 hours from this set of analyses.

Here, we use cell state to mean an annotated cluster at a given time window. Altogether, we identified 171 cell states in sci-ATAC-seq data and 268 in sci-RNA-seq data across the nine time windows, each of which received one of 38 cell type annotations for ATAC or one of 54 cell type annotations for RNA (tables S1 and S2 and Fig. 3, A and B). Across time windows, we identified an average of 109 marker genes and 2469 marker accessible regions per cluster (tables S3 and S4).

The early stages of *Drosophila* embryogenesis, represented by our 0- to 2-hour time window, include 13 rapid nuclear divisions within a syncytium that generates 6000 nuclei, regulated by maternal genes. At ~2 hours and 20 min after fertilization, cellularization occurs and the zygotic genome is activated (*21*), followed by gastrulation to generate the three germ layers. Our single-cell data recapitulate these events, where the earliest time window (0 to 2 hours) has two large clusters annotated as maternal or unknown. At 2 to 4 hours, the maternal cluster is no longer present, and instead, pole cells and anlage clusters appear. A notable expansion in the diversity of cell types follows across 6 to 10 hours, matching expectations for when the major lineages in each germ layer are specified (Fig. 3, A and B).

To follow the emergence and diversification of cell lineages, we systematically linked cell clusters across developmental time, applying similar methods as in earlier studies (*3*, *22*) to coembeddings of cells from adjacent nonoverlapping, inferred time windows (fig. S6, B and C). For cells of each state derived from the "child" time window, we calculated the median proportion of nearest neighbors from the "parent" window that were derived from each potential parental cell state and treated this as the weight of the corresponding edge. The maximum edge weights >0.2 were retained, resulting in acyclic, directed graphs, independently generated from scRNA and scATAC data (Fig. 3, C and D). Although these procedures were generated independently of our cell cluster annotations at each time window, they overwhelmingly yielded internally consistent results. For example, muscle clusters in one time window connect to muscle clusters in the next time window, and the same is true for other major lineages (e.g., central nervous system, peripheral nervous system, etc.) as embryogenesis proceeds. We note that some paths seem to terminate prematurely, potentially because of drastic increases in cell number in later embryogenesis, which were not matched by corresponding increases in our sampling, or because of unknown technical or biological factors. More generally, because these are inferences based on cellular state rather than lineage tracing, they may be prone to certain kinds of error (*3*).

To illustrate the potential of these data to facilitate exploration of specific lineages at finer resolution, we reanalyzed 59,012 cells annotated as neuroectoderm using scRNA data from 6 to 18 hours (Fig. 3E and fig. S7A). This revealed 20 subclusters, including a large group of early cells corresponding to the brain primordium and neural progenitors that express regulators of neurogenesis, such as *Notch* (*N*) and *Delta* (*Dl*), and neuroblast temporal TFs, such as *miranda* (*mira*) and *castor* (*cas*). Two additional neural progenitor clusters correspond to sensory progenitors, whereas immature neurons express low levels of both neural progenitor and pan-synaptic genes, including *cacophony* (*cac*) and *synaptotagmin 1* (*syt1*). Mature neurons are marked by higher levels of pan- and subtype-specific synaptic genes coupled with low or no expression of earlier developmental genes. Finally, midline cells, consisting of both neurons and glia cluster together, become evident at 6 to 8 hours; using the midline TF *single minded* (*sim*) and glial immunoglobulin family member *wrapper* as markers, we can follow them forward in time as they mature (fig. S7B). We can also follow the maturation of sensory neural progenitors, marked by *shaven* (*sv*), from 6 to 16 hours (fig. S7B).

To further explore neuronal diversity, we reclustered 6703 mature neurons, revealing 11 neuronal subtypes, which we manually curated (Fig. 3F). Among these, we identify four clearly separable sensory cell clusters. There are two types of *Drosophila* sensory neurons based on dendritic morphology: type I sensilla, which include both external sensory (ES) neurons and internal chordotonal (Ch) neurons, and type II multidendritic (MD) neurons. We can clearly distinguish MD neurons on the basis of expression of genes, such as dendritic arbor reduction 1 (*dar1*), which promotes their characteristic branching dendrites, and the pseudouridine synthase RluA-1, which was recently identified as a marker of MD neurons (*23*) (Fig. 3, F and G). Consistent with their nociceptive role, this cluster also specifically expresses the mechanical nociception degenerin/epithelial sodium channel subunits *pickpocket* (*ppk*) and *ppk26*. Mechanosensory ES neurons are specified by the TF *hamlet* (*ham*), which is specifically expressed in the middle sensory cluster (Fig. 3, F and G) (*24*). The adjacent cluster, likely Ch sensory neurons, is identified by expression of the mechanosensitive nonselective cation channel subunit *no mechanoreceptor potential C* (*nompC*) as well as fate-determinant *Rfx* and a number of as-yet uncharacterized genes specific to this cluster (*25*, *26*) (Fig. 3, F and G). The final sensory cluster likely corresponds to

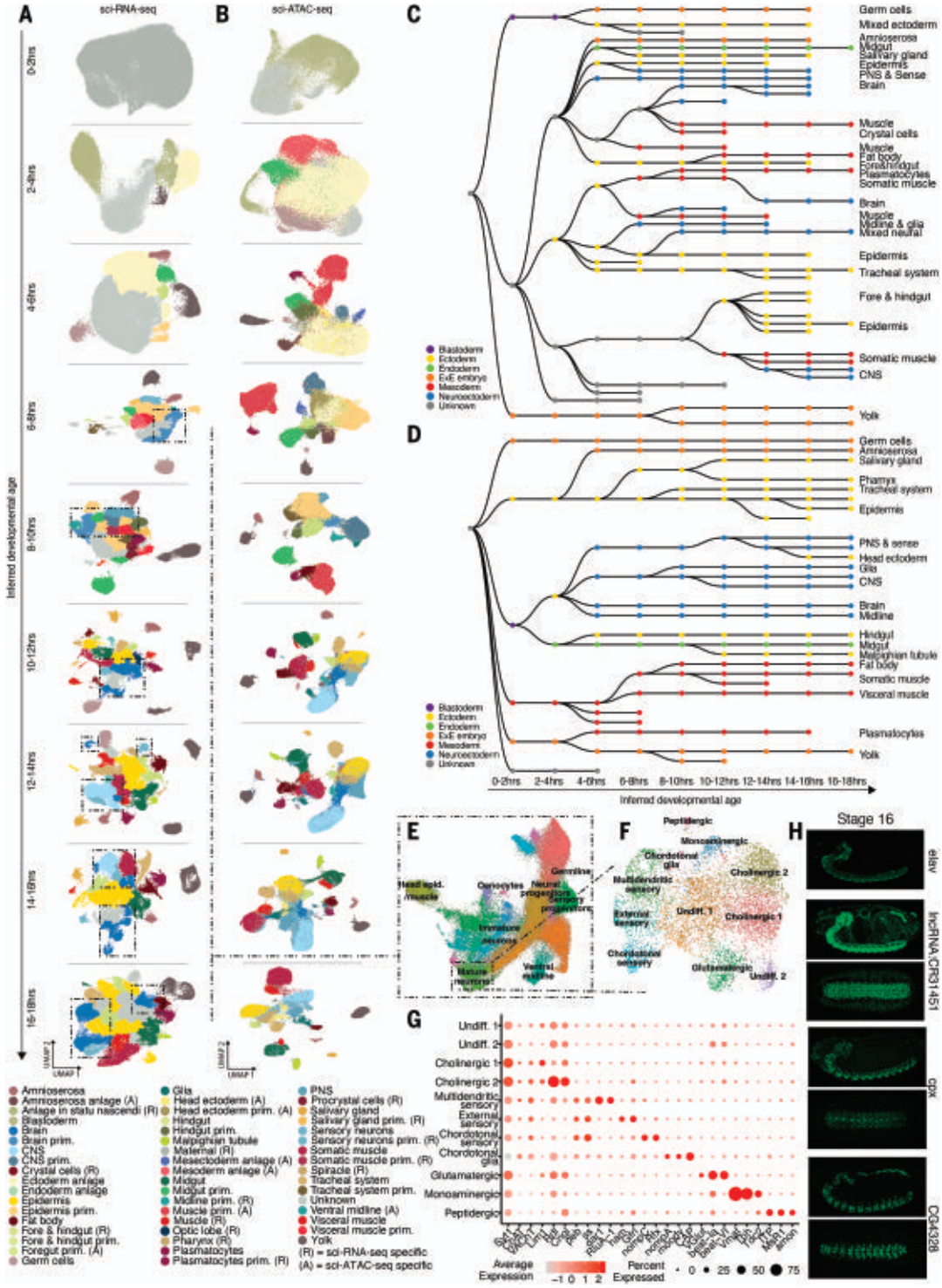


Fig. 3. Annotation of diversifying developmental trajectories. (**A**) UMAP visualization of nonoverlapping, inferred 2-hour time windows for scRNA clusters colored by cell state annotation. Dashed boxes highlight neuroectodermal clusters. (**B**) Same as (A), but for scATAC data. PNS, peripheral nervous system; CNS, central nervous system. (**C**) ScRNA-based acyclic directed graph representation of clusters linked through nonoverlapping time windows. (**D**) Same as (C), but from scATAC data. (**E**) UMAP of scRNA data for ~60,000 annotated neuroectodermal cells—i.e., cell states highlighted in (A) with dashed boxes, colored by cluster. (**F**) UMAP of ~6000 mature neurons, colored by cluster. The chordotonal glia cluster includes Ch and ES organ glial-like support cells. (**G**) Dot plot showing marker gene expression for annotated clusters in (F). (**H**) In situ hybridization of stage 16 embryos, showing the expression of lncRNA *CR31451*, *cpx*, and *CG4328* in the nervous system. A tissue marker (*elav*) is provided in the top panel. A lateral and ventral embryo view is shown for each gene.

Ch glial-like support cells based on the expression of glial markers, including *moody*, and *Cbl-associated protein* (*CAP*) and *nompA*, which promote the development and function of Ch support cells, respectively (Fig. 3, F and G). On the basis of vesicular neurotransmitter transporter expression, we also identify two clusters of central cholinergic neurons, a glutamatergic cluster that likely includes motor neurons, and monoaminergic neurons (Fig. 3, F and G). Finally, peptidergic neurons cluster separately and were identified on the basis of the expression of neuropeptides [*ion transport peptide* (*ITP*)], enzymes involved in their synthesis [*amontillado* (*amon*)], and receptors [*myosuppressin receptor 1* (*MsR1*)] (Fig. 3, F and G).

We validated the expression of uncharacterized long noncoding RNA (lncRNA) CR31451 as enriched in mature neurons as well as two genes, *complexin* (*cpx*) and *CG4328*, identified in our analysis as enriched in the monoaminergic cluster, which includes midline neurons (Fig. 3H). This neuronal subtype enrichment is unexpected for *cpx*, which encodes a presynaptic regulator of synaptic vesicle release, and may point to additional requirements for Cpx in midline monoaminergic neurons. In the course

of exploring these fine neuronal subtypes, we also made an unexpected finding regarding *elav*, a classic marker gene for neurons. Specifically, we noticed lower-level expression of *elav* in clusters annotated as visceral muscle. Performing double fluorescent in situ hybridization with a visceral muscle–specific marker gene (*biniou*) confirmed this unexpected finding (fig. S7C) and raises the possibility of a potential previously unknown role of this well-studied gene.

This deeper exploration of the neuroectoderm, validating and extending years of research from many groups, illustrates the depth of information that can be obtained from these data. We additionally performed a more detailed annotation of nonmyogenic mesoderm (supplementary note 1). A full exploration of all lineages represented in these data will require a community-wide effort by tissue experts (as done in this study for neuronal diversity).

In addition to delineating developmental trajectories, these data can also capture spatial differences arising during developmental patterning. Previous bulk ATAC-seq on embryo halves has shown variability in the accessibility of enhancers along the anterior-posterior (A-P) axis of the blastoderm embryo (*27*). Using label transfer to map anterior or posterior identities from a previous blastoderm dataset (*12*) onto our 2- to 4-hour data, we computed a positional accessibility skew score for validated enhancers with strict A-P activity (*27*). This indicates that accessibility of most A-P enhancers is skewed in the expected anterior or posterior cell group (fig. S7D), recapitulating the bulk data (*27*). Notably, we also identify differences among enhancers of the same gene. For example, in the *eve* locus, the stripe 1 enhancer has a much stronger skew for anterior accessibility compared with stripe 2, as has also been previously reported (*27*). Our single-cell data thus capture the biological variability in enhancer accessibility along the A-P axis, extending previous observations. We similarly could transfer labels from our sci-RNA-seq clusters to spatial coordinates from a spatial enhanced resolution omics sequencing (Stereo-seq)–based spatial study of *Drosophila* embryos at 14 to 16 hours and 16 to 18 hours of development (*28*). Using the assigned annotations of tissues from the spatial study, we observe a correspondence with our cluster annotations, which again suggests the spatial-relevant variability present in these data (fig. S7E).

Tracing dynamic gene modules across development

To further leverage continuous views of unfolding trajectories, we next explored the gene regulatory modules active in germ layer–specific development. We focused on the mesoderm and its derivatives as a complex, well-characterized system that we and others have studied previously (*11*, *13*, *29*, *30*). For this, we selected all cells corresponding to mesoderm-derived cell states, collectively 51,338 (scRNA) and 200,907 (scATAC) profiles across 4 to 20 hours and 2 to 20 hours of inferred developmental age, respectively (Fig. 4, A and B).

Focusing first on RNA, we selected the top 2000 most variable genes. After normalizing expression values to be comparable across time, we used dynamic time warp clustering to group genes into four clusters with distinct temporal regulation (Fig. 4C, fig. S8A, and table S6). These clusters define broad successive waves of gene expression during mesoderm development (Fig. 4D) and notably exhibit similarly ordered waves of chromatin accessibility (fig. S8, B and D, and supplementary note 2). Gene pathway enrichment suggests different functional roles for each cluster (fig. S8C). Cluster 1 genes (n = 571) are highly expressed from the beginning of mesoderm development (directly after gastrulation; 4 to 9 hours); are enriched for TFs ($P = 1.4 \times 10^{-6}$); and likely represent a mixture of genes involved in progenitor cells, mesoderm development, and transcriptional activation (Fig. 4D and fig. S8C). Cluster 2 genes (n = 433) peak at ~9 to 11 hours, during the subdivision of the mesoderm into different muscle primordia and their subsequent specification. This cluster is enriched for genes involved in mesoderm development, including myoblast fusion and myotube differentiation, while losing enrichment for stem cell and self-renewal terms (Fig. 4D and fig. S8C). By contrast, cluster 3 genes (n = 365) initiate expression at ~10 hours and steadily increase to the end of embryogenesis, whereas cluster 4 genes (n = 631) only switch on at ~15 hours, during muscle terminal differentiation. The last cluster lacks enrichment for TFs and rather includes genes involved in myofibril assembly and muscle assembly and maintenance as well as essential contractile proteins for differentiated muscle (Fig. 4D and fig. S8C). We validated the spatiotemporal expression of five poorly characterized genes by in situ hybridization, confirming that they are expressed in the mesoderm or muscle at the inferred time window (Fig. 4E).

The temporal and cell type–specific nature of these expression signatures for both the downstream effector molecules and their upstream regulators should provide the resolution to order genes into putative regulatory hierarchies. For example, several genes with essential roles in muscle differentiation, such as myosin heavy chain (*Mhc*), are present in clusters 3 and 4. Mhc protein plays a critical role in providing muscle-contractile force. Our scRNA data show increasing *Mhc* expression along the muscle lineages in cells with later embryonic ages (Fig. 4, A and F), matching the expression pattern of *Mhc*. Concomitantly, there is a gradual increase in open chromatin at characterized *Mhc* enhancers at later stages along multiple muscle trajectories (Fig. 4G).

Before the expression of *Mhc* and other muscle differentiation genes, we observe transient expression of mesoderm-associated TFs (cluster 2; Fig. 4C). One example is *Kahuli* (*Kah*), a TF associated with muscle development, which has peak expression at 10 hours of embryogenesis (cluster 2; Fig. 4, C, D, F, and G). To investigate the relationship between open chromatin and gene expression, we computed gene activity scores, defined as the sum of sci-ATAC-seq reads in the gene body and the 2 kb flanking the transcription start site (TSS). The gene activity scores for both *Mhc* and *Kah* recapitulate their sequential temporal patterns of expression, with *Kah*'s activity signature appearing earlier along the mesodermal trajectories compared with that of *Mhc* (Fig. 4, F and G). To determine the extent to which we could map the exact ordering of accessibility and expression changes, we overlaid the scaled expression values and gene activity scores averaged across bins with equal numbers of cells (Fig. 4G). Notably for *Kah*, gene expression temporally follows the trajectory of the corresponding gene activity score based on open chromatin, suggesting an ordering where first the gene body becomes accessible followed by accumulating levels of the corresponding transcript; however, this was not the case for *Mhc*, for which expression and accessibility increased in tandem (Fig. 4G). Kah binds to several characterized *Mhc* enhancers near the gene's promoter, as observed in bulk ChIP sequencing (ChIP-seq) data (*14*), which suggests a regulatory link between Kah and *Mhc* expression (Fig. 4H).

To extend this analysis more globally, we searched for TF motifs enriched in putative enhancers (mesoderm-specific scATAC peaks 1 to 10 kb upstream of the TSS) of genes belonging to each of the four scRNA mesoderm expression clusters. This identified 458 TF motif–to-cluster enrichments ($q < 1 \times 10^{-3}$ and presence in >1% of target peaks; table S7) corresponding with 152 unique TFs. Of these, 31 are TFs whose expression changes along mesoderm differentiation and are thus included in the expression-based clustering (table S7). These 31 include many TFs essential for mesoderm development, including a number of direct target genes of the master regulator Twist (the functional ortholog of MyoD) at the beginning of mesoderm development (e.g., *hb*, *en*, *Ubx*, and *pb*), and concordantly expressed in the first temporal cluster. These factors have many functions, including setting up the segmentation of the mesoderm, regulating the expression of somatic muscle identity genes, establishing midgut constrictions in the visceral mesoderm, and heart patterning. Other examples from the second and third temporal

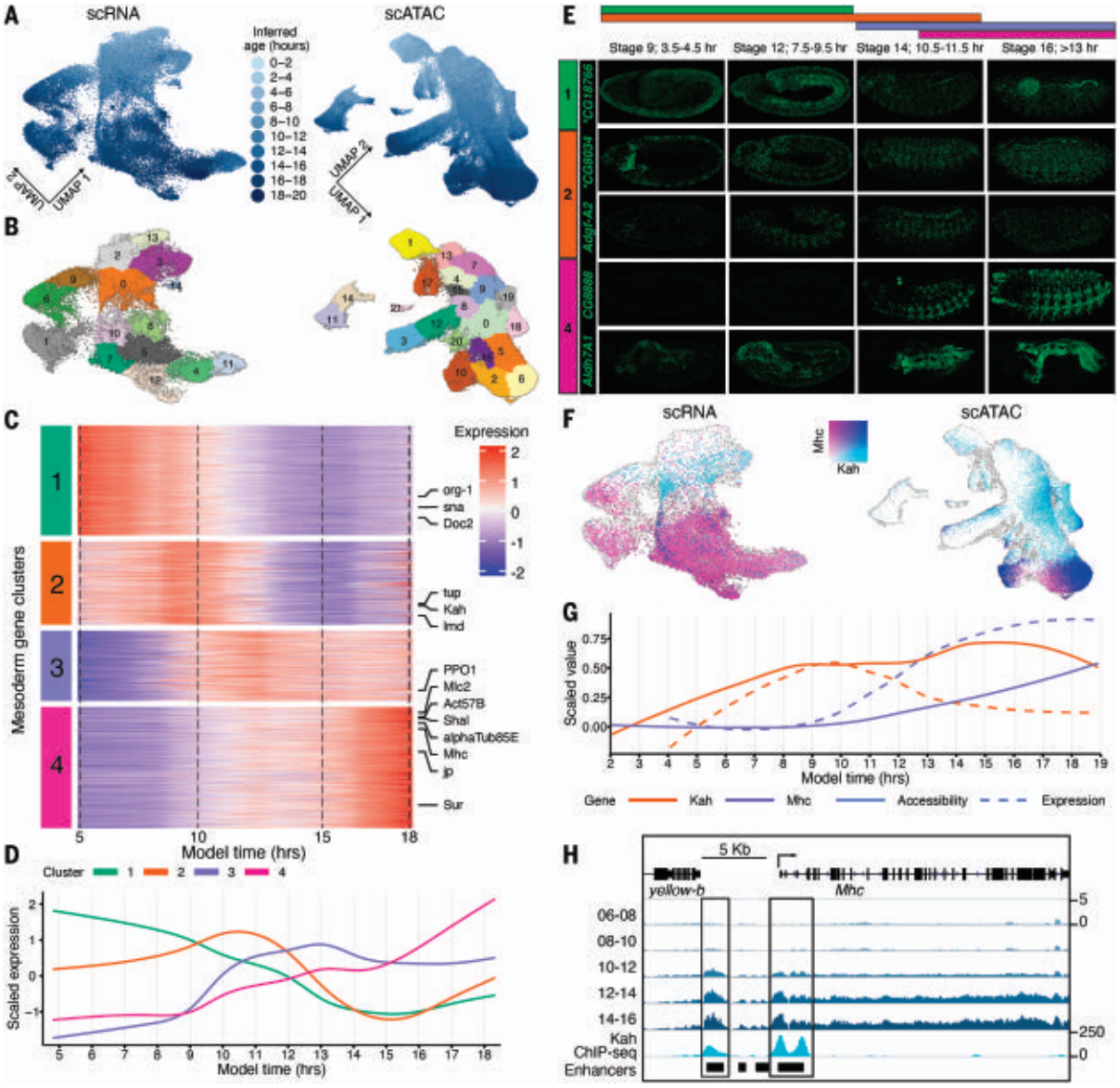


Fig. 4. Dynamic regulation of mesoderm-specific gene modules. (**A**) UMAP of scRNA (left) or scATAC (right) data for all mesodermal cells, colored by inferred developmental age. (**B**) Same as (A), but colored as reprocessed leiden-based clusters. (**C**) Normalized expression of mesoderm genes across inferred developmental time. (**D**) Average expression of the gene modules across inferred time. (**E**) In situ hybridization experiments validating temporal expression of selected genes with predicted expression in mesoderm and muscle (asterisks indicate see supplementary note 3). (**F**) Same as (A), but expression of *Kah* (cyan) and *Mhc* (purple) is overlaid. Points from cells that express both *Kah* and *Mhc* are colored dark blue. (**G**) Comparison of gene activity score (solid line) and gene expression (dashed line) over the continuum of inferred developmental age for *Kah* (cluster 2) and *Mhc* (cluster 3) in mesoderm-annotated cells. Gene activity scores and expression were binned into 100 equal partitions by inferred age, averaged, and scaled to 0 to 1 with min-max values. (**H**) Chromatin accessibility profile surrounding *Mhc* for pseudobulk mesoderm cells from 6 to 16 hours inferred time in 2-hour increments, along with Kah ChIP-seq generated from 0- to 16-hour whole embryos (*14*).

clusters are genes required for cell fate specification of somatic muscle founder cells (e.g., *Six4* and *ap*) and heart development (e.g., *tup* and *Lim3*).

We note that this approach may miss the contribution of important TFs that were not variably expressed in mesoderm. In particular, if a TF is variably expressed and has corresponding variability in motif activity, this TF is likely active. However, this does not imply that all expressed TFs are active (e.g., there may be coactivators or posttranslational modifications that are required). This caveat notwithstanding, these analyses highlight the potential for further discovery of coregulated gene modules related to distinct germ layers or cell types.

Nominating stage– and cell type–specific TF regulators

We next investigated whether we could leverage the diversity of cell states across embryogenesis to infer which TFs drive specific programs of cell type differentiation. For this, we used all scATAC clusters at all time points (in contrast to the scRNA-focused cluster analysis above) and searched for differential enrichment of TF position weight matrices (PWMs) within each cluster's open chromatin regions.

We first characterized enrichments across clusters from the 10- to 12-hour time window based on predicted time (Fig. 5A). Encouragingly, hierarchical clustering of the enrichment profiles of all associated PWMs grouped each cluster roughly by germ layer (this was also observed in other time windows; fig. S9A). The

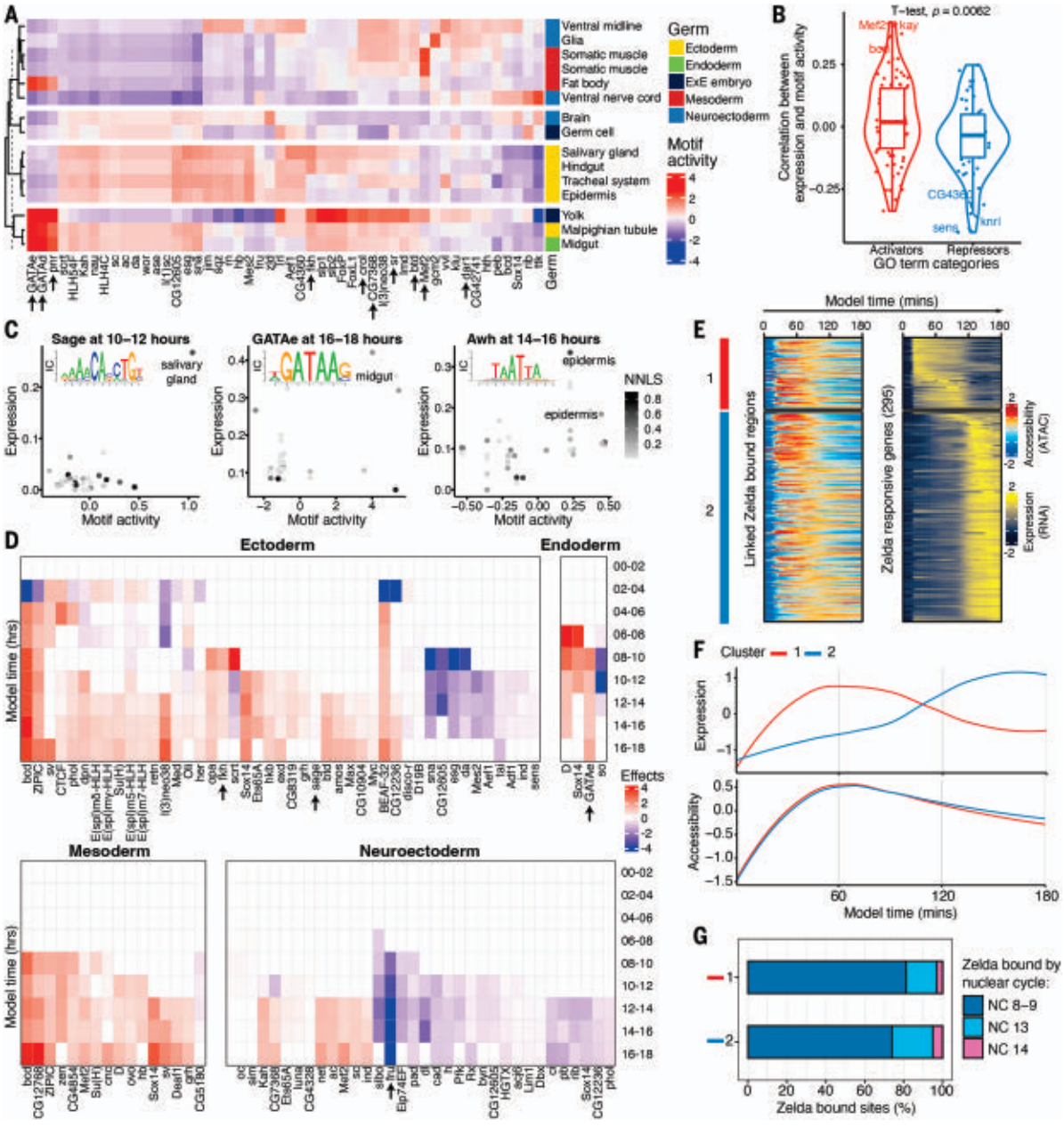


Fig. 5. Integration of scRNA and scATAC data to identify TFs with potential regulatory roles across differentiating tissues and developmental time. (**A**) Heatmap with averaged chromatin accessibility differences associated with the 50 most variable TF-specific motifs from all cells in annotated ATAC-seq clusters from 10 to 12 hours. Arrows indicate TFs discussed in the main text. (**B**) Correlation between expression and motif-associated accessibility grouped by expression activation- or repression-associated GO categories. TFs in GO pathways for gene activation are linked to increasing chromatin accessibility. (**C**) Comparison of gene expression (*y* axis) and motif-associated chromatin accessibility (*x* axis) across NNLS-linked clusters for the TFs Sage (left), GATAe (middle), and Awh (right). Each TF's corresponding PWM is inset in each plot, with the size of each base scaled by information content. (**D**) Heatmaps of estimated effects of gene expression at predicting motif-associated chromatin accessibility changes through time in different germ layers. Displayed TFs had three or more consecutive time windows with a significant ($P < 1 \times 10^{-3}$) and sign-consistent effect. Arrows indicate TFs discussed in the main text. (**E**) Heatmap of expression at Zelda-responsive genes (right) and aggregated chromatin accessibility (left) at their Zelda-bound cis-regulatory regions (*38*, *39*). Values were averaged in 1-min windows over 0 to 3 hours of development. The red and blue bars to the left indicate two temporal clusters of expression of Zelda-responsive genes. (**F**) Smoothed average expression and accessibility for the two Zelda temporal clusters from (E). (**G**) Proportion of accessible regions from (E) that are bound by Zelda in clusters 1 and 2 in ChIP-seq data (*39*) from different nuclear cycles (NCs).

nonmyogenic mesoderm (fat body) and myogenic mesoderm (somatic muscle) cluster together (Fig. 5A). Open chromatin regions in the myogenic clusters are enriched in motifs for many TFs known to play a role in muscle development, including Mef2 and Fork head (Fkh) TFs. The myogenic clusters also appear close to two neuronal clusters (Fig. 5A), which is driven by shared motif enrichment with neuroectoderm and glial cells, particularly many C2H2 zinc finger TFs, including Btd, CG7368, Crol, Sr, and Dar1. Many of these factors have known roles in neuronal development (e.g., Dar1), whereas Stripe (Sr) is essential for muscle tendon cell fate and muscle

attachment in the epidermis at late stages of embryogenesis (*31*).

Because members of the same family of TFs typically recognize similar motif sequences (e.g., GATAe, GATAd, and pnr), it is often difficult from motif analysis alone to pinpoint the responsible TF. To address this, we leveraged our scRNA data to identify the most likely active TF on the basis of its expression within the clusters among all factors that share the same motif binding pattern. First, we used a regression-based framework to integrate the scATAC and scRNA datasets and identify links between the different cell clusters (*1*, *6*). Specifically, we adopted a nonnegative least square (NNLS) matrix factorization approach that decomposes expression data as a mixture of components derived from proximal gene activity scores generated from the scATAC data. Despite possible temporal differences between accessibility and expression, NNLS identifies stronger links between clusters from the same 2-hour window compared with those from adjacent 2-hour windows (fig. S9B). We also inferred NNLS links in the opposite direction by decomposing proximal gene activity scores by gene expression associated with scRNA clusters. For each cluster of a given data type, the result of NNLS factorization is a mixture proportion of clusters from the other data type, where a higher value represents a stronger association between the scRNA and scATAC cluster (fig. S9, C to F, and table S8). This factor decomposition approach resulted in a strong linkage (NNLS-mixture coefficient of >0.1) of 120 cell state clusters present in the same inferred time windows, with most of the strongly linked clusters being from 4 to 6 hours onward. Upon manual inspection, many linked scATAC and scRNA clusters, which had been independently annotated, are from matching tissues. For example, from the 10- to 12-hour window, the epidermis cluster (cluster 0) in scATAC data was matched to the epidermis (cluster 3) in scRNA data. Altogether, of 21 ATAC clusters from the 10- to 12-hour window, 16 had a linked RNA annotation with a NNLS correlation value >0.1, of which 14 were between comparable tissue annotations.

These integrated scRNA and scATAC clusters, which span 0 to 18 hours of embryogenesis, enabled a more direct analysis of the role of specific TFs in different cell types' differentiation. We reasoned that active TFs should be more highly expressed in cell types for which they have a functional role, and their associated PWM should be more enriched or depleted in accessible regions when the TF is activating or repressing expression (*6*). In line with this, correlation values between motif-associated accessibility and gene expression were shifted toward more positive values for TFs annotated [by gene ontology (GO)] as activators and toward more negative values for annotated repressors (Fig. 5B and table S9), a trend also observed in human fetal tissues (*6*). This approach of linking TFs' cluster-specific expression and motif enrichments allowed us to nominate TFs as active at specific times in specific tissues (Fig. 5C). For example, this analysis predicts a specific role for Sage in salivary gland development, as the salivary gland is the only cell type exhibiting both high expression of the *sage* transcript and high accessibility of the Sage-associated PWM (Fig. 5C, top). This finding matches the essential role for *sage* in salivary gland development, as determined by genetic loss-of-function analysis (*32*). Similar predictions were made for GATAe in the midgut at 16 to 18 hours and Awh in the epidermis at 14 to 16 hours (Fig. 5C, middle and bottom), matching the functional role for both TFs in midgut endoderm (*33*) and epidermis (*34*, *35*) development, respectively.

To expand this analysis and systematically nominate TFs that potentially drive germ layer–specific differentiation programs, we fit a linear model that predicts a TF's motif-associated chromatin changes from an estimated effect of an interaction term that includes the expression level of the TF in a specific germ layer and time window. Our model's effect estimates can identify TFs with specific motif activity in particular germ layers and suggest time windows from which a TF initiates its activity. For example, the model refined the role of Sage as becoming active in the ectoderm germ layer specifically from 10 to 12 hours onward and the activity of GATAe initiating in the endoderm from 8 to 10 hours onward (Fig. 5D, top). Such a model encompassing germ layers across development time may also identify additional likely coactive TFs. For example, in addition to Sage, we found Fkh to be both coexpressed and coactive in the ectoderm—a TF reported to act together with Sage to activate salivary gland–specific genes (*36*).

This analysis also generated additional interesting findings for other time points and germ layers [e.g., Fruitless (Fru); supplementary note 4 and Fig. 5D]. Altogether, from eight high-level germ layer–associated tissue annotations and 316 TF motifs tested, we identified 1258 significant (Benjamini-Hochberg–corrected $P < 1 \times 10^{-3}$) TF-to-tissue relationships having both associated expression and chromatin activity at one or more of the nine time windows assessed. We note that in time windows with fewer clusters, the association effect estimates are susceptible to outliers and should be interpreted with caution. Notwithstanding this caveat, these putative assignments represent an extensive resource for future studies (table S10).

To demonstrate the potential of our approach to discover previously unknown putative roles for TFs, we selected four genes and validated whether they were expressed in the linked germ layer by fluorescent in situ hybridization. Although these genes were inferred to have effects in multiple germ layers, their function in either mesoderm (*CG5953* and *CG11617*) or neuroectodermal tissues (*Ets65A* and *CG12605*) was poorly characterized. We confirmed that these factors are in fact expressed in the tissue and time window predicted by our data (fig. S10), suggesting potential roles for these TFs in mesoderm and neuronal development.

To complement the NNLS, we applied a recently developed tool, FigR (*37*), to further facilitate gene regulatory network (GRN) reconstruction. Because multi-omic ATAC-RNA data from the same cell are required for this task, we first integrated our two independent assays for all cells from 10 to 12 hours using canonical correlation analysis (CCA), identifying the most likely ATAC-RNA cell pairs using geodesic distance–based pairing (*37*) within the common CCA space. Using these pairs as input for GRN inference with FigR, we linked ATAC peaks to their target genes based on peak-to-TSS accessibility correlation and then computed TF motif enrichments for the linked regions, which, together with the TF expression-accessibility correlation, allowed us to define hundreds of putative activators and repressors at this embryonic stage (fig. S11A). Ranking the TFs by their regulation score (fig. S11B) nominated many activators and repressors that we also identified in the NNLS analysis above, including *l(3)neo38*, *Lim3*, *lola*, *fkh*, and *fru* (Fig. 5D). Focusing on the targets of the regulatory networks across all cells at 10 to 12 hours, we found a large set of genes that appear to be extensively regulated (209 genes with >10 linked regulatory regions) (fig. S11C). We then used the inferred TF activities to explore the factors acting on these genes and their mode of regulation. For example, *tup*, a TF gene required for heart development, undergoes extensive self-regulation (highest motif-RNA correlation) besides being positively regulated by the pan-muscle TF Mef2 and repressed by Run and Opa (fig. S11D). Another top-ranking gene, *chinmo*, an essential neuronal TF, is activated by other nervous system TFs, such as Lim1 and Onecut, and is negatively regulated by Fru (fig. S11E), which we also identified as a neuroectoderm-specific repressor in our NNLS-based analysis (Fig. 5D and supplementary note 4).

Finally, we sought to exploit the fine-grained resolution of inferred nuclear ages to explore the dynamics of an early pioneer TF, Zelda, in regulating chromatin opening followed by transcription during ZGA. We recovered the expression of a set of genes that are Zelda dependent during ZGA (*38*) and, for each gene, aggregated accessibility at the linked Zelda-bound regions (*39*) in intervals of 1 min across 0 to 3 hours of embryogenesis (Fig. 5E). Clustering of gene expression identified two broad temporal clusters—a first group of early genes

and a second group whose expression increases later, after ~1.5 hours of embryogenesis. Notably, although accessibility at the Zelda-bound regions linked to the early cluster seems to mirror the temporal expression, regions linked to the late expression gene cluster gain accessibility much earlier, almost as early as the first cluster, which suggests that Zelda is opening these regions for future activation (Fig. 5F). To verify whether accessibility is reflective of Zelda binding, we retrieved Zelda occupancy by nuclear cycle (*39*), which confirmed that >70% of regions in both temporal clusters are already occupied by Zelda at nuclear cycle 8 to 9, regardless of the associated gene expression (Fig. 5G). Moreover, we found a partial Clamp TF motif match within the second temporal cluster (and no match for the first cluster of a TF that is also expressed), which corroborates its Zelda-paired role during ZGA (*40*). These results suggest that Zelda establishes chromatin accessibility at a large set of regulatory regions in the early embryo, independently of future gene expression, in agreement with its well-known role as a pioneer factor. In some cases, Zelda possibly also functions as the activator of gene expression (cluster 1), whereas in others it retains a pioneering role, and the gene's expression is induced by later TFs (cluster 2).

Discussion

This continuum of *Drosophila* embryogenesis builds on our previous work generating sci-ATAC-seq from three nonoverlapping time windows of embryogenesis (*12*) and complements other studies performed on specific tissues (*30*, *41*–*46*) as well as scRNA from entire embryos at one specific stage (*7*) or on dissected tissues from adults (*47*). Despite the growing use of single-cell assays to generate large-scale atlases, characterizing fine-scale dynamics of chromatin accessibility and gene expression across developmental time remains a challenge. The large number of cell types and even greater number of cell states and branch points during embryogenesis requires extensive cell sampling at continuous stages to capture regulatory transitions, especially for rare cell types. This is very difficult if not essentially impossible to obtain in most model organisms.

In this work, sampling embryo collections from overlapping 2- to 4-hour time windows, coupled with NN-based inference of more precise nuclear ages, enabled continuous representation of *Drosophila* embryonic development. Other studies have attempted a similar ordering of embryos by developmental time over a 2-day window of mouse development (*4*). However, because only dozens rather than thousands of mouse embryos can practically be sampled, reliable inference at the scale of hours or minutes is challenging. Similarly, cell age was inferred in *Caenorhabditis elegans* using an independent time series of bulk RNA-seq from whole embryos (*48*). However, relying on such whole-embryo bulk data to predict developmental age in single cells risks inaccurate aging of rare or transient cell types, especially for more complex organisms.

Computationally, our NN-based inference of developmental age bears some similarity to the concept of pseudotime. As originally proposed, pseudotime aims to serve as "a quantitative measure of progress through a biological process" (*18*). Analogously, our inferred developmental age tracks the progression of nuclei through development. However, the advantage of pairing an experimental design including overlapping yet tightly defined time windows with temporal ordering is that we can anchor inferred ages to fixed time points, which can potentially lead to a more accurate representation of developmental age for complex cellular trajectories. Put another way, inferred ages are interpretable as units of absolute time that are synchronized across all tissue trajectories. With such a continuum of cellular states, we can begin to infer cell type trajectories that more closely capture the continuous processes of cellular differentiation unfolding within a complex, developing multicellular organism.

There remain further possible improvements to our experimental framework. The alignment or anchoring to real time could be refined with sampling of more tightly staged windows. Multi-omic methods for characterizing multiple data types from the same nuclei may facilitate a joint model that can link paired gene expression and chromatin accessibility (and other modalities) to developmental age inference. There are cases where technical features of the data can lead to increased uncertainty of model predictions. For example, we found that cells annotated as germ cells, from the first collection time window, or with low read count were associated with greater prediction error (fig. S11F). Moving forward, we suggest caution for interpreting findings solely on the basis of inferred nuclear ages from clusters with these features.

The extensive scATAC data, with deep coverage across almost a million cells, likely captured most regulatory elements active during embryonic development and provides a comprehensive resource of potential enhancers for almost any cell type in the embryo. By contrast, our scRNA data had relatively low unique reads per cell and will likely miss some differentially expressed genes in specific cell types. As a result, some delicate analyses remain challenging. For example, we found transcriptional velocity estimates to be unstable with sparse scRNA data, although this issue was mitigated by constructing metacells before velocity analysis (fig. S11G), which may be useful for pursuing targeted questions. In scATAC data, we were able to distinguish XX versus XY nuclei from the proportion of chrX-mapped reads (fig. S11H); however, this was challenging for the scRNA data, again as a result of data sparsity. These shortcomings are to some degree compensated by the large number of cells profiled, as shown by our ability to recapitulate aspects of previously documented heterogeneity even for highly dynamic or restricted phenomena—e.g., ZGA (Fig. 2E).

Overall, this *Drosophila* embryonic atlas provides broad insights into the orchestration of cellular states during the most dynamic stages in the life cycle of the organism. Our results represent a rich resource for understanding precise time points at which genes become active in distinct tissues as well as how chromatin is remodeled across time. The annotation of cell types within these data is an ongoing process and one that is much more challenging at early and mid-stages of embryogenesis as compared with late time points or in adults with differentiated tissues. A comprehensive annotation of embryonic cell states will require a collective effort from the *Drosophila* community. To support these ongoing efforts, we provide information on expression and peaks from all clusters (Fig. 3, A to D) in addition to all intermediate and raw data for further exploration. Although larval stages remain insufficiently profiled, we hope that these data and methods, together with the recently released large-scale adult atlas (*47*), bring us closer to the community-wide goal of a multimodal *Drosophila* atlas spanning a continuum from zygote to adulthood.

Materials and methods summary

A detailed version of the materials and methods is provided in the supplementary materials. In brief, *D. melanogaster* embryos were acquired for each of 11 collection windows, and then each pool of embryos was divided, with each half being extracted and fixed for either sci-RNA-seq3 or sci-ATAC-seq3. Libraries were sequenced deeply, and the resulting reads were mapped to dm6 and then processed with a uniform processing pipeline that included quality control (QC) filters for low read depth or high proportions of reads mapping to the mitochondria or ribosomal genes and extensive doublet removal. Between the two data modalities, we obtained profiles for ~1.5 million nuclei, although unique read depth per nucleus was considerably lower for scRNA than scATAC data.

Using the center hour of the collection window, we used several machine learning approaches to fit a model that could infer the age of a nucleus with either gene expression or chromatin accessibility information. Both LL regression and neural networks were fitted to the same training data, with a held-out subset used for model validation and comparison. Given its consistently superior performance, we then relied on specific parameterizations of

NN model–inferred ages to reposition nuclei in time. To zoom into fine-scale time points, we binned data by small increments to explore the regulatory dynamics of ZGA. Then, using 2-hour adjacent windows of cells, we computed clusters of similar cells and performed extensive manual review to annotate each cluster's likely germ layer and cell type. We then used an iterative approach for constructing an acyclic tree of differentiation by identifying the likely precursor cluster for each cluster in a given time window.

Neuroectoderm was iteratively analyzed for deeper annotation of neuronal subtypes, whereas mesoderm was picked for analyses focused on identifying coregulated genes and accessible regions, which were then subjected to ontology and TF motif enrichment analysis. To connect scATAC cell clusters with scRNA cell clusters, we used a regression-based approach (NNLS). Such connections between ATAC and RNA clusters enabled a series of analyses, such as correlating expression with motif accessibility, applying GRN analysis pipelines, etc.

Several additional analyses were performed. We used probabilistic label transfer to map likely cluster annotations from these data to spatial information from patterned DNA nanoballs. We also found it is possible to infer the sex of cells from the proportion of chrX-mapped scATAC reads using a Gaussian mixture model to classify cells. Although RNA velocity was challenging to apply to sparse scRNA data, it yielded more sensible results when subsets of cells were first aggregated to metacells.

The expressions of several genes were verified by fluorescent in situ hybridization: specific neuronal genes active in identified clusters, unexpected coactivity of the *elav* with *binou*, genes active at specific mesoderm time points, and putative active TFs with less-characterized roles in tissue development.

Raw data are available through the Gene Expression Omnibus (GEO). Additional scripts and intermediate files, including bigwigs and a custom web app to visualize UMAPs, are available through our data-sharing website.

REFERENCES AND NOTES

1. J. Cao *et al*., The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019). doi: 10.1038/s41586-019-0969-x; pmid: 30787437
2. B. Pijuan-Sala *et al*., A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019). doi: 10.1038/s41586-019-0933-9; pmid: 30787436
3. C. Qiu *et al*., Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat. Genet.* **54**, 328–341 (2022). doi: 10.1038/s41588-022-01018-x; pmid: 35288709
4. M. Mittnenzweig *et al*., A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* **184**, 2825–2842.e22 (2021). doi: 10.1016/j.cell.2021.04.004; pmid: 33932341
5. G. La Manno *et al*., Molecular architecture of the developing mouse brain. *Nature* **596**, 92–96 (2021). doi: 10.1038/s41586-021-03775-x; pmid: 34321664
6. S. Domcke *et al*., A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020). doi: 10.1126/science.aba7612; pmid: 33184180
7. N. Karaiskos *et al*., The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017). doi: 10.1126/science.aan3235; pmid: 28860209
8. J. Rivera, S. V. E. Keränen, S. M. Gallo, M. S. Halfon, REDfly: The transcriptional regulatory element database for *Drosophila*. *Nucleic Acids Res.* **47**, D828–D834 (2019). doi: 10.1093/nar/gky957; pmid: 30329093
9. E. Z. Kvon *et al*., Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature* **512**, 91–95 (2014). doi: 10.1038/nature13395; pmid: 24896182
10. S. Bonn *et al*., Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat. Genet.* **44**, 148–156 (2012). doi: 10.1038/ng.1064; pmid: 22231485
11. J. P. Reddington *et al*., Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. *Dev. Cell* **55**, 648–664.e9 (2020). doi: 10.1016/j.devcel.2020.10.009; pmid: 33171098
12. D. A. Cusanovich *et al*., The *cis*-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018). doi: 10.1038/nature25981; pmid: 29539636
13. R. P. Zinzen, C. Girardot, J. Gagneur, M. Braun, E. E. M. Furlong, Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* **462**, 65–70 (2009). doi: 10.1038/nature08531; pmid: 19890324
14. M. M. Kudron *et al*., The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of *Drosophila* and *Caenorhabditis elegans* Transcription Factors. *Genetics* **208**, 937–949 (2018). doi: 10.1534/genetics.117.300657; pmid: 29284660
15. B. R. Graveley *et al*., The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011). doi: 10.1038/nature09715; pmid: 21179090
16. J. C. Kwasnieski, T. L. Orr-Weaver, D. P. Bartel, Early genome activation in *Drosophila* is extensive with an initial tendency for aborted transcripts and retained introns. *Genome Res.* **29**, 1188–1197 (2019). doi: 10.1101/gr.242164.118; pmid: 31235656
17. S. A. Blythe, E. F. Wieschaus, Establishment and maintenance of heritable chromatin structure during early *Drosophila* embryogenesis. *eLife* **5**, e20148 (2016). doi: 10.7554/eLife.20148; pmid: 27879204
18. C. Trapnell *et al*., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014). doi: 10.1038/nbt.2859; pmid: 24658644
19. A. S. Hammonds *et al*., Spatial expression of transcription factors in *Drosophila* embryonic organ development. *Genome Biol.* **14**, R140 (2013). doi: 10.1186/gb-2013-14-12-r140; pmid: 24359758
20. P. Tomancak *et al*., Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8**, R145 (2007). doi: 10.1186/gb-2007-8-7-r145; pmid: 17645804
21. J. D. Laver, A. J. Marsolais, C. A. Smibert, H. D. Lipshitz, in *Current Topics in Developmental Biology*, vol. 113, H. D. Lipshitz, Ed. (Academic Press, 2015), pp. 43–84.
22. J. A. Briggs *et al*., The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018). doi: 10.1126/science.aar5780; pmid: 29700227
23. W. Song, S. Ressl, W. D. Tracey, Loss of Pseudouridine Synthases in the RluA Family Causes Hypersensitive Nociception in *Drosophila*. *G3* **10**, 4425–4438 (2020). doi: 10.1534/g3.120.401767; pmid: 33028630
24. A. W. Moore, L. Y. Jan, Y. N. Jan, *hamlet*, a binary genetic switch between single- and multiple- dendrite neuron morphology. *Science* **297**, 1355–1358 (2002). doi: 10.1126/science.1072387; pmid: 12193790
25. R. Dubruille *et al*., *Drosophila* regulatory factor X is necessary for ciliated sensory neuron differentiation. *Development* **129**, 5487–5498 (2002). doi: 10.1242/dev.00148; pmid: 12403718
26. R. G. Walker, A. T. Willingham, C. S. Zuker, A *Drosophila* mechanosensory transduction channel. *Science* **287**, 2229–2234 (2000). doi: 10.1126/science.287.5461.2229; pmid: 10744543
27. J. E. Haines, M. B. Eisen, Patterns of chromatin accessibility along the anterior-posterior axis in the early *Drosophila* embryo. *PLOS Genet.* **14**, e1007367 (2018). doi: 10.1371/journal.pgen.1007367; pmid: 29727464
28. M. Wang *et al*., High-resolution 3D spatiotemporal transcriptomic maps of developing *Drosophila* embryos and larvae. *Dev. Cell* **57**, 1271–1283.e4 (2022). doi: 10.1016/j.devcel.2022.04.006; pmid: 35512700
29. G. Junion *et al*., A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148**, 473–486 (2012). doi: 10.1016/j.cell.2012.01.030; pmid: 22304916
30. S. Secchia, M. Forneris, T. Heinen, O. Stegle, E. E. M. Furlong, Simultaneous cellular and molecular phenotyping of embryonic mutants using single-cell regulatory trajectories. *Dev. Cell* **57**, 496–511.e8 (2022). doi: 10.1016/j.devcel.2022.01.016; pmid: 35176234
31. G. Frommer, G. Vorbrüggen, G. Pasca, H. Jäckle, T. Volk, Epidermal egr-like zinc finger protein of Drosophila participates in myotube guidance. *EMBO J.* **15**, 1642–1649 (1996). doi: 10.1002/j.1460-2075.1996.tb00509.x; pmid: 8612588
32. R. M. Fox, A. Vaishnavi, R. Maruyama, D. J. Andrew, Organ-specific gene expression: The bHLH protein Sage provides tissue specificity to *Drosophila* FoxA. *Development* **140**, 2160–2171 (2013). doi: 10.1242/dev.092924; pmid: 23578928
33. R. Reuter, The gene serpent has homeotic properties and specifies endoderm versus ectoderm within the Drosophila gut. *Development* **120**, 1123–1135 (1994). doi: 10.1242/dev.120.5.1123; pmid: 7913013
34. J. Curtiss, J. S. Heilig, *Arrowhead* encodes a LIM homeodomain protein that distinguishes subsets of *Drosophila* imaginal cells. *Dev. Biol.* **190**, 129–141 (1997). doi: 10.1006/dbio.1997.8659; pmid: 9331336
35. E. Preger-Ben Noon, F. P. Davis, D. L. Stern, Evolved Repression Overcomes Enhancer Robustness. *Dev. Cell* **39**, 572–584 (2016). doi: 10.1016/j.devcel.2016.10.010; pmid: 27840106
36. M. M. Myat, D. J. Andrew, Fork head prevents apoptosis and promotes cell shape change during formation of the Drosophila salivary glands. *Development* **127**, 4217–4226 (2000). doi: 10.1242/dev.127.19.4217; pmid: 10976053
37. V. K. Kartha *et al*., Functional Inference of Gene Regulation using Single-Cell Multi-Omics. bioRxiv 2021.07.28.453784 [Preprint] (2021). doi: 10.1101/2021.07.28.453784
38. S. A. Blythe, E. F. Wieschaus, Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. *Cell* **160**, 1169–1181 (2015). doi: 10.1016/j.cell.2015.01.050; pmid: 25748651
39. M. M. Harrison, X.-Y. Li, T. Kaplan, M. R. Botchan, M. B. Eisen, Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLOS Genet.* **7**, e1002266 (2011). doi: 10.1371/journal.pgen.1002266; pmid: 22028662
40. J. Duan *et al*., CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. *eLife* **10**, e69937 (2021). doi: 10.7554/eLife.69937; pmid: 34342574
41. M. N. Özel *et al*., Neuronal diversity and convergence in a visual system developmental atlas. *Nature* **589**, 88–95 (2021). doi: 10.1038/s41586-020-2879-3; pmid: 33149298
42. J. Bageritz *et al*., Gene expression atlas of a developing tissue by single cell expression correlation analysis. *Nat. Methods* **16**, 750–756 (2019). doi: 10.1038/s41592-019-0492-x; pmid: 31363221
43. C. N. McLaughlin *et al*., Single-cell transcriptomes of developing and adult olfactory receptor neurons in *Drosophila*. *eLife* **10**, e63856 (2021). doi: 10.7554/eLife.63856; pmid: 33555999
44. Z. Shi *et al*., Single-cyst transcriptome analysis of *Drosophila* male germline stem cell lineage. *Development* **147**, dev.184259 (2020). doi: 10.1242/dev.184259; pmid: 32122991
45. N. J. Everetts, M. I. Worley, R. Yasutomi, N. Yosef, I. K. Hariharan, Single-cell transcriptomics of the *Drosophila* wing disc reveals instructive epithelium-to-myoblast interactions. *eLife* **10**, e61276 (2021). doi: 10.7554/eLife.61276; pmid: 33749594
46. R.-J. Hung *et al*., A cell atlas of the adult *Drosophila* midgut. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1514–1523 (2020). doi: 10.1073/pnas.1916820117; pmid: 31915294
47. H. Li *et al*., Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022). doi: 10.1126/science.abk2432; pmid: 35239393
48. J. S. Packer *et al*., A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution. *Science* **365**, eaax1971 (2019). doi: 10.1126/science.aax1971; pmid: 31488706
49. K. L. Howe *et al*., Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021). doi: 10.1093/nar/gkaa942; pmid: 33137190

ACKNOWLEDGMENTS

We thank B. Kim, C. Qiu, S. Domcke, J. Cao, and S. Srivatsan for helpful discussions on data processing, analysis, and visualization as well as all members of the Shendure and Furlong laboratories for helpful input and discussions. **Funding:** D.C. was supported by award no. T32HL007828 from the National Heart, Lung, and Blood Institute along with support by award no. F32HG011817 from the National Human Genome Research Institute. J.K. was supported by award no. T32GM136566 from the National Institute of General Medical Sciences. K.M.O.-G. and E.L. and work in their laboratories were funded by NIH National Institute of Neurological Disorders and Stroke award no. NS125864-01. E.E.M.F. and work in the Furlong laboratory were supported by grants from the Deutsche Forschungsgemeinschaft (DFG-SPP 2202 and CRC 1550) agreement FU 750, Baden-Württemberg Stiftung (BWST-ISF2019-032), and European Research Council (ERC advanced grant) agreement no. 787611 (DeCRyPT). J.S. and work in the Shendure laboratory were supported by the Paul G. Allen Frontiers Group (Allen Discovery Center for Cell Lineage Tracing). J.S. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** S.S. and A.D. performed embryo collections and carried out embryo fixation and nuclei isolation for sci-ATAC-seq. B.M. and R.B.-G. performed the sci-RNA-seq experiments. R.M.D. and R.B.-G. performed the sci-ATAC-seq experiments. D.C., X.H., S.S., and R.B.-G. carried out the formal analysis. D.C., R.B.-G., and X.H. wrote the first draft of the manuscript, which was then finalized with guidance from E.E.M.F. and J.S. and input from all authors. S.S. performed data analysis and results interpretation and participated in figure generation and manuscript writing. C.S. performed in situ hybridization experiments. J.K. aided with additional analyses and fine-scale neuronal cell type annotation supervised by E.L. and K.M.O.-G. C.T. contributed to conceptualizing the time modeling and comparisons to pseudotime and revised the manuscript. E.E.M.F. and J.S. conceptualized, supervised, and funded the project. **Competing interests:** J.S. is a SAB member and a consultant and/or cofounder of Cajal Neuroscience, Guardant Health, Maze Therapeutics, Camp4 Therapeutics, Phase Genomics, Adaptive Biotechnologies, and Scale Biosciences. C.T. is a SAB member and a consultant and/or cofounder of Algen Biotechnologies, Altius Therapeutics, and Scale Biosciences. The authors declare no other competing interests. **Data and materials availability:** All other data are in the main paper or the supplementary materials. All raw data are available through the GEO series GSE190149. Additional scripts and intermediate files, including bigwigs for all time windows and clusters, and a custom web app to visualize UMAPs are available at https://shendure-web.gs.washington.edu/content/members/DEAP_website/public/. We downloaded the Kah ChIP-seq data from the ENCODE portal with identifier ENCSR161YRO. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn5800
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S13
References (*50*–*90*)
MDAR Reproducibility Checklist

Submitted 6 December 2021; accepted 28 June 2022
10.1126/science.abn5800

RESEARCH ARTICLES

MEDICINE

Mitochondrial remodeling and ischemic protection by G protein–coupled receptor 35 agonists

Gregory A. Wyant[1,2], Wenyu Yu[1], Ilias P. Doulamis[3], Rio S. Nomoto[3], Mossab Y. Saeed[3], Thomas Duignan[3], James D McCully[3], William G. Kaelin, Jr.[1,2,4]*

Kynurenic acid (KynA) is tissue protective in cardiac, cerebral, renal, and retinal ischemia models, but the mechanism is unknown. KynA can bind to multiple receptors, including the aryl hydrocarbon receptor, the a7 nicotinic acetylcholine receptor (a7nAChR), multiple ionotropic glutamate receptors, and the orphan G protein–coupled receptor GPR35. Here, we show that GPR35 activation was necessary and sufficient for ischemic protection by KynA. When bound by KynA, GPR35 activated G_i- and $G_{12/13}$-coupled signaling and trafficked to the outer mitochondria membrane, where it bound, apparently indirectly, to ATP synthase inhibitory factor subunit 1 (ATPIF1). Activated GPR35, in an ATPIF1-dependent and pertussis toxin–sensitive manner, induced ATP synthase dimerization, which prevented ATP loss upon ischemia. These findings provide a rationale for the development of specific GPR35 agonists for the treatment of ischemic diseases.

Kynurenic acid (KynA) has been shown to protect tissues in preclinical organ ischemia models and to mediate cardioprotection in a mouse model of remote ischemic preconditioning, a phenomenon whereby an ischemic tissue confers ischemic protection to other tissues at a distance (*1*–*5*). Mice, however, are a suboptimal model for cardiac physiology studies because of their small size, extremely rapid heart rates, and short life spans (*6*). Therefore, we tested KynA in rabbits. Pretreatment of rabbits with KynA decreased infarct size when their excised, beating, and perfused hearts (Langendorff prep) were subjected to a brief period of ischemia followed by reperfusion to model ischemia/reperfusion (I/R) injury (fig. S1, A to D). KynA also increased cardiac function after I/R, as measured by increased fractional shortening, increased left ventricular contractility (dP/dt max), and left ventricular peak developed pressure (LVPDP) during reperfusion (fig. S1, E to H).

We tested which of the putative KynA receptors (*7*–*13*), if any, mediated KynA's tissue-protective effects. Because KynA protects isolated hearts against ischemia, we focused on the aryl hydrocarbon receptor (AhR) and G protein–coupled receptor 35 (GPR35) rather than on ionotropic glutamate receptors, which are linked to neurotransmission. KynA promoted the binding of the AhR to its partner, aryl hydrocarbon receptor nuclear translocator (ARNT), and activated AhR-responsive transcription (fig. S2, A and B). However, cardiac AhR loss in mice did not block KynA-induced cardioprotection in the setting of I/R injury in vivo (fig. S2, C and D). Moreover, stable expression of a constitutively active *AHR* (AHR-CA) cardiomyocytes derived from human induced pluripotent stem (hIPS) cells did not promote their survival when they were transiently deprived of oxygen and nutrients and then re-exposed to oxygen and nutrients to simulate I/R injury ex vivo (fig. S2, E to G). Therefore, AhR activation is neither necessary nor sufficient for KynA-induced cardioprotection.

GPR35 is broadly expressed, and its abundance is increased after hypoxia under the control of the HIFα transcription factor and also in human heart failure (fig. S3, A to C) (*14*, *15*). Moreover, GPR35 is linked to tissue protection, although studies disagree as to whether GPR35 agonism or antagonism is protective (*16*–*18*). Whether KynA is an authentic endogenous GPR35 ligand is also controversial, partly because its binding affinity for GPR35 varies across species and, in some species, is relatively low (in the micromolar range) (*19*).

We confirmed that radiolabeled KynA bound to purified human and mouse GPR35 (Fig. 1A). This was specific, because radiolabeled KynA did not bind to a control G protein–coupled receptor (GPCR), GPR160, and its binding to GPR35 was prevented by excess unlabeled KynA and the known GPR35 activators pamoic acid, zaprinast, and lodoxamide, but not by the KynA-derived metabolite quinolinic acid or tryptophan (the parent molecule from which KynA is derived) (Fig. 1B) (*20*–*22*). KynA activated G_i and $G_{12/13}$ in a GPR35-dependent manner, as determined by G_i and $G_{12/13}$ recruitment, activation of the protein kinase Rho, and suppression of forskolin-induced cAMP abundance and downstream signaling by protein kinase A (PKA) (fig. S4, A to E).

We sought to create a GPR35 mutant that could not bind to KynA so that it could be used as a specificity control. We made a series of GPR35 missense mutants in which specific residues within GPR35 transmembrane domains III and IV were converted to alanine, motivated by previous homology modeling and ligand-docking studies that implicated this region (particularly specific arginine residues within this region) in ligand binding (*23*, *24*). One such variant, GPR35 R151A, failed to bind to radiolabeled KynA in assays using GPR35-expressing cell membrane fractions or highly purified GPR35 (Fig. 1, C to F). Similarly, KynA bound to wild-type GPR35, but not GPR35 R151A, in thermal shift assays based on thermostabilization upon ligand binding (Fig. 1G). Because these studies used GPR35 that was purified from mammalian cells (293T cells), we cannot yet formally exclude that additional cellular proteins are required for the binding of KynA to GPR35. Many GPCRs, including GPR35, are internalized upon activation (*25*). We confirmed that wild-type GPR35, but not GPR35 R151A, was internalized in early endosome antigen 1 (EEA1)–containing endosomes in cells treated with KynA or the GPR35 agonists pamoic acid or zaprinast, but not in cells treated with quinolinic acid (Fig. 1H and fig. S5A) (*11*, *15*). GPR35-containing endosomes also contained transferrin receptor (TfR), which is normally present on plasma membranes, consistent with the GPR35 endosomes participating in endosomal trafficking (fig. S5B). β-arrestin localization was not regulated by KynA or by the GPR35 agonist pamoic acid in the cardiomyocytes that we studied, although it is regulated by many other GPCRs that are internalized upon activation (fig. S5C) (*25*).

Cardioprotection by KynA administered either 2 or 24 hours before injury in vivo or 10 or 30 min before I/R in ex vivo hearts was completely abrogated in $GPR35^{-/-}$ mice that we made using CRISPR/Cas9–based gene editing of mouse embryos (Fig. 2, A and B, and fig. S6A) (*26*). GPR35 loss did not alter, either positively or negatively, the induction of hypoxia-inducible transcription factor 1α (HIF1α) after cardiac I/R injury (fig. S6B). In cell culture experiments performed with murine neonatal cardiomyocytes, KynA, in a GPR35-dependent manner, decreased resting oxygen consumption and, after simulated I/R injury, decreased mitochondrial reactive oxygen species production and preserved mitochondrial membrane potential (fig. S6, C to E), changes that would be predicted to promote survival after I/R.

The drug FG-4592 (roxadustat), which stabilizes HIF1α by inactivating the egg laying

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA. [2]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA. [3]Department of Cardiac Surgery, Boston Children's Hospital, Department of Surgery, Harvard Medical School, Boston, MA 02215, USA. [4]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, USA.
*Corresponding author. Email: william_kaelin@dfci.harvard.edu

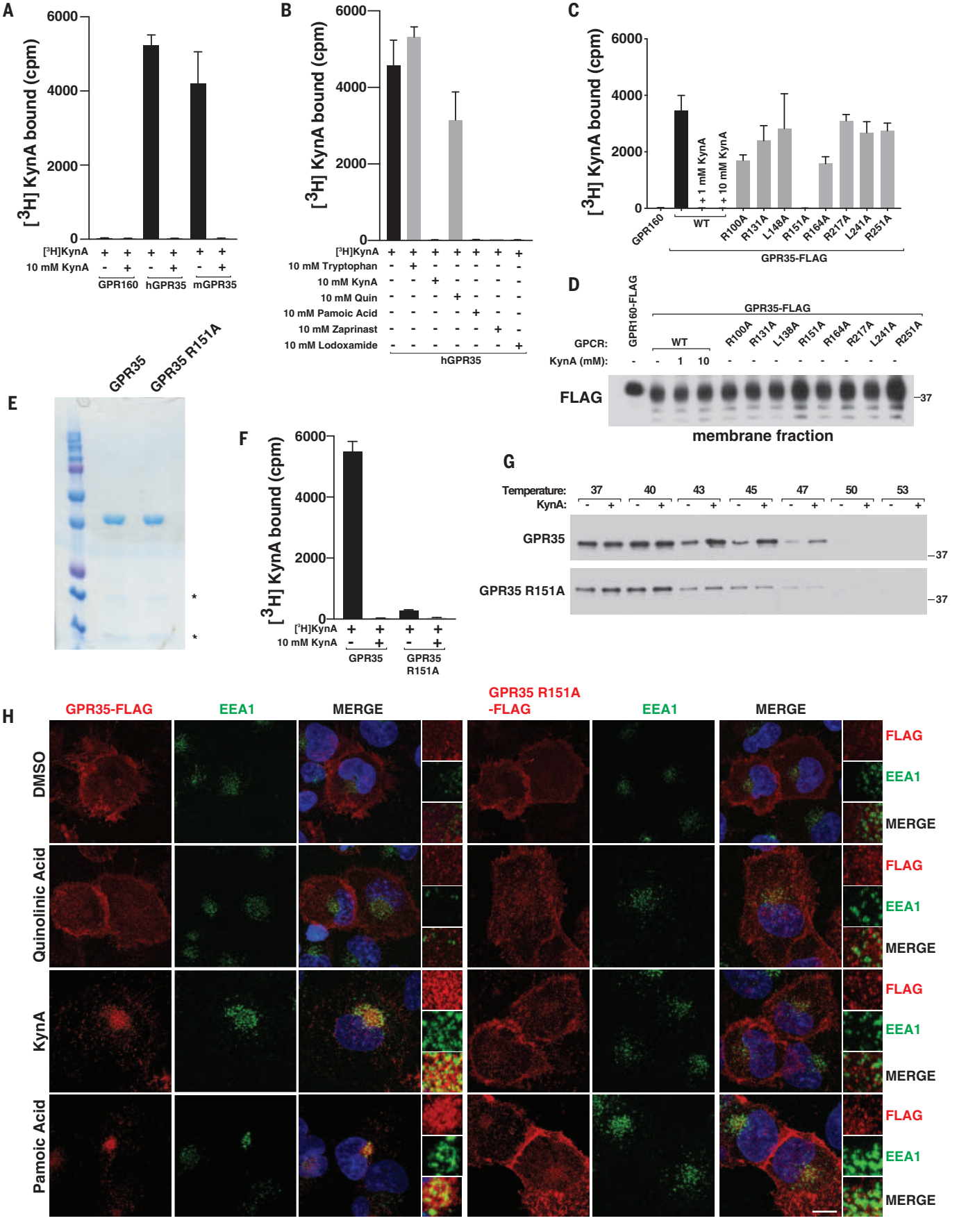

A
[3H] KynA bound (cpm)
6000
4000
2000
0
[3H]KynA + + + + + +
10 mM KynA - + - + - +
GPR160 hGPR35 mGPR35
B
[3H] KynA bound (cpm)
[3H]KynA + + + + + + +
10 mM Tryptophan - + - - - - -
10 mM KynA - - + - - - -
10 mM Quin - - - + - - -
10 mM Pamoic Acid - - - - + - -
10 mM Zaprinast - - - - - + -
10 mM Lodoxamide - - - - - - +
hGPR35
C
[3H] KynA bound (cpm)
+ 1 mM KynA
+ 10 mM KynA
GPR160
WT
R100A R131A L148A R151A R164A R217A L241A R251A
GPR35-FLAG
D
GPR160-FLAG
GPR35-FLAG
GPCR:
WT
R100A R131A L138A R151A R164A R217A L241A R251A
KynA (mM): - - 1 10 - - - - - - - -
FLAG
37
membrane fraction
E
GPR35
GPR35 R151A
*
*
F
[3H] KynA bound (cpm)
[3H]KynA + + + +
10 mM KynA - + - +
GPR35
GPR35 R151A
G
Temperature: 37 40 43 45 47 50 53
KynA: - + - + - + - + - + - + - +
GPR35
GPR35 R151A
37
37
H
GPR35-FLAG
EEA1
MERGE
GPR35 R151A -FLAG
EEA1
MERGE
DMSO
Quinolinic Acid
KynA
Pamoic Acid
FLAG
EEA1
MERGE

Fig. 1. KynA binds GPR35 and promotes GPR35 internalization. (**A, B**, and **F**) [^{3}H]-KynA–binding assay using tandem affinity–purified versions of the indicated high-density lipoprotein–reconstituted hemagglutinin (HA)-FLAG–tagged GPCRs in the presence or absence of the indicated unlabeled competitor compounds (10 mM). (**C** and **D**) [^{3}H]-KynA binding (C) and immunoblot (D) assays of isolated membrane fractions from 293T cells expressing FLAG-GPR160 or FLAG-GPR35 (wild-type or mutant). Where indicated, 1 and 10 mM unlabeled KynA was used as a competitor. (**E** and **G**) Commassie blue staining (E) and immunoblot (G) assay of the purified GPCRs used in (F). In (G), the proteins were preincubated with 1 mM KynA or dimethyl sulfoxide (DMSO) for 1 hour before exposure to the indicated temperatures. (**H**) Confocal microscopy of mouse neonatal cardiomyocytes stably expressing wild-type or R151A GPR35-FLAG stimulated with 20 μM quinolinic acid, 20 μM KynA, or 1 μM pamoic acid for 20 min. Scale bar, 20 μm. In (A), (B), (C), and (F), values are shown as means ± SDs for three technical replicates from one representative experiment.

defective 9 (EGLN) prolyl hydroxylases, was shown to be cardioprotective in preclinical models (*27*–*29*). Cardioprotection is likely an on-target effect of FG-4592 because similar effects have been observed with other structurally unrelated EglN inhibitors and after systemic or cardiac-specific genetic inactivation of EglN1, which is the primary regulator of HIF1α stability (*1*, *28*, *30*–*32*). Although EglN inactivation is thought to cardioprotect by stabilizing HIF1α, systemic inactivation of EglN also increases the plasma concentrations of 2-oxoglutarate (2-OG), which is converted to KynA by the liver and released into the circulation (*1*). Cardiac protection by FG-4592 in vivo was partially abrogated by loss of GPR35 (Fig. 2A). By contrast, FG-4592 was cardioprotective in wild-type and GPR35$^{-/-}$ hearts when administered ex vivo, a setting in which KynA would not be produced, before I/R (Fig. 2, B and C). Protection by FG-4592 was observed when given 30 min, but not 10 min, before I/R, which correlated with successful induction of HIF1α 30 min after FG-4592 administration (Fig. 2C). By contrast, KynA protected more rapidly (Fig. 2B). These results suggest that cardiac protection in vivo by roxadustat is caused by the direct effects of HIFα and the indirect effects on GPR35 mediated by endogenous KynA.

Pertussis toxin, a classic G_i and G_o protein family inhibitor, abolished cardiac ischemic preconditioning in rats (*33*) and likewise inhibited KynA-induced cardiac ischemic protection in vivo, suggesting that intact G_i or G_o family signaling by GPR35 is required for KynA-induced ischemic protection (Fig. 2D). Multiple structurally unrelated GPR35 agonists, including lodoxamide, zaprinasat, and pamoic acid, were cardioprotective in vivo and ex vivo, whereas quinolinic acid, which fails to activate GPR35, was not (Fig. 2, E and F). Where tested, cardioprotection by these other GPR35 agonists was GPR35 dependent (Fig. 2A).

To formally test whether KynA binding to GPR35 is necessary for KynA-induced ischemic protection, we used CRISPR-Cas9 to generate hIPS cells lacking GPR35 and then infected them with a lentivirus encoding wild-type GPR35, GPR35 R151A, or the empty vector (Fig. 2G). These cells were then induced to become cardiomyocytes and subjected to simulated I/R ex vivo. KynA pretreatment protected against I/R-induced injury, and this protection was lost in the absence of GPR35. Reexpression of the wild-type GPR35, but not GPR35 R151A, in the GPR35$^{-/-}$ cardiomyocytes restored protection by KynA (Fig. 2H). Thus, GPR35 mediates KynA-induced ischemic protection in both mouse and human cells, and this protection requires the binding of KynA to GPR35.

GPR35 is annotated to bind to the mitochondrial protein ATP synthase inhibitory factor subunit 1 (ATPIF1) and ATP synthase in the BioPlex protein-protein interaction database (*34*). To test whether GPR35 can associate with mitochondria, we reintroduced wild-type or R151A GPR35 into GPR35$^{-/-}$ murine neonatal cardiomyocytes by lentiviral infection. Treatment of the wild-type cells with GPR35 agonists, including KynA, led to colocalization of GPR35 with mitochondrial proteins such as citrate synthase and the ATP synthase component ATP5H, but not the Golgi apparatus protein GM130 or the endoplasmic reticulum protein calreticulin, as shown by confocal microscopy and quantitative image analysis (Fig. 3A and figs. S7, A to C, and S8, A to D). This was specific because colocalization with mitochondria was not observed with GPR35 R151A or with quinolinic acid. KynA and the GPR35 agonist pamoic acid, but not quinolinic acid, also increased GPR35 and TOM20 association in situ based on proximity ligation assays (fig. S7, D to E). To visualize GPR35 trafficking to mitochondria, we transiently overexpressed GPR35-GFP together with mCherry-TOMM20 in GPR35$^{-/-}$ murine cardiomyocytes and performed live-cell imaging. KynA treatment caused GPR35-GFP to become internalized within minutes on cytoplasmic punctae that associated with mCherry-TOMM20 (fig. S8E). To corroborate these findings, we rapidly isolated mitochondria from cells expressing 3XHA-EGFP-OMP25 (HA-MITO) after KynA treatment and then performed immunoblot assays (*35*). We reproducibly detected the association of GPR35 with mitochondria, and this was increased in cells treated with KynA or other GPR35 agonists, but not in cells treated with quinolinic acid (Fig. 3B and fig. S9A). We introduced a biotin ligase (APEX) fused to wild-type or R151A GPR35 into the GPR35$^{-/-}$ murine cardiomyocytes and captured GPR35-associated proteins using streptavidin agarose after treatment with biotin tyramide and hydrogen peroxide (to activate the APEX enzyme) in the presence or absence of GPR35 agonists (fig. S9B) (*36*). Multiple mitochondrial outer membrane proteins, such as VDAC, TOM20, and TOM70, were associated with wild-type, but not R151A, GPR35 in cells treated with KynA (Fig. 3C). Similarly, KynA promoted biotin labeling of mitochondria, as determined by confocal microscopy using streptavidin-568 as a probe, in cells expressing wild-type GPR35-APEX, but not in those expressing GPR35 R151A-APEX (fig. S9C) (*37*).

We confirmed, using two different epitope tags, that epitope-tagged human and mouse GPR35 coimmunoprecipitated with endogenous ATPIF1 and ATP synthase, as determined by immunoblot analysis (fig. S10, A to D). Consistent with our localization studies, the binding of GPR35 to ATPIF1 was enhanced by KynA and the GPR35 agonists zaprinast and pamoic acid, but not by quinolinic acid (Fig. 4A and fig. S10A). As additional specificity controls, we looked for binding of ATPIF1 to cytosolic METAP2, the transmembrane mitochondrial protein TMEM141, GPR160, or a GPR35 C-terminal truncation mutant (Fig. 4B and fig. S10B), but found none. The electrophoretic mobility of wild-type GPR35 was decreased under these gel conditions compared with the C-terminal truncation mutant because of a slightly higher molecular size and because of glycosylation events that require the GPR35 C terminus (*38*).

ATPIF1 binds to, and regulates, the multimeric ATP synthase complex (*39*). The ATP synthase subunits ATP5B, ATP5H, ATP5O, and ATP5F1 also coimmunoprecipitated with GPR35, and their abundance mirrored that of ATPIF1 (Fig. 4, A and B, and fig. S10, A to C). Given that ATPIF1 and ATP synthase are localized inside mitochondria, we tested whether GPR35 is present inside or outside mitochondria after treatment with KynA. We treated murine cardiomyocytes expressing wild-type GPR35 with KynA and used differential centrifugation to purify mitochondria, which were then exposed to increasing amounts of proteinase K. Mitochondrially associated GPR35 was proteinase K sensitive, arguing that it is largely associated with the outer mitochondria membrane (Fig. 4C). Consistent with this, KynA promoted GPR35-APEX biotinylation of the outer mitochondrial membrane proteins VDAC and TOM70, but did not promote biotinylation of ATPIF1 or ATP synthase

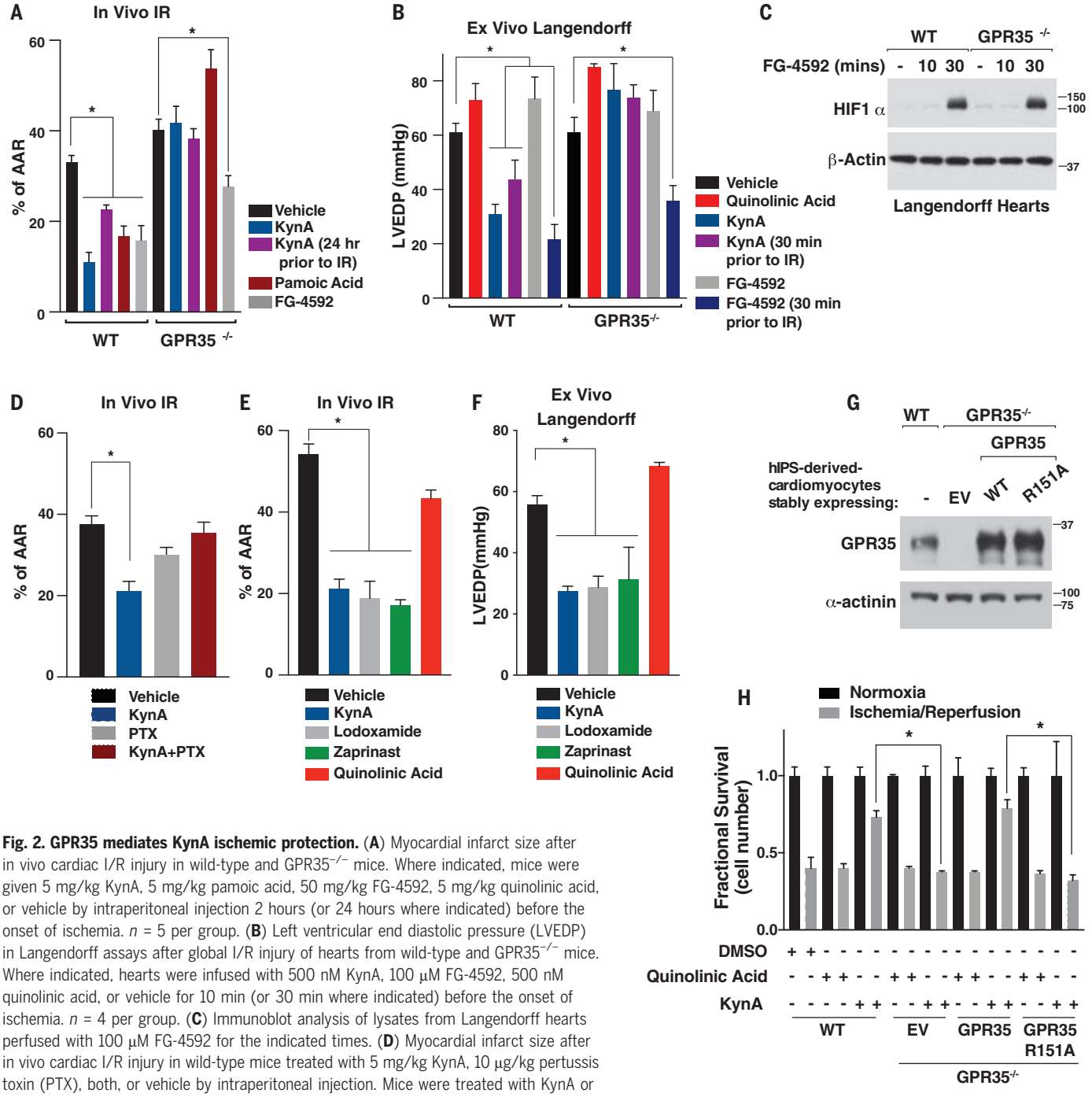


Fig. 2. GPR35 mediates KynA ischemic protection. (**A**) Myocardial infarct size after in vivo cardiac I/R injury in wild-type and GPR35$^{-/-}$ mice. Where indicated, mice were given 5 mg/kg KynA, 5 mg/kg pamoic acid, 50 mg/kg FG-4592, 5 mg/kg quinolinic acid, or vehicle by intraperitoneal injection 2 hours (or 24 hours where indicated) before the onset of ischemia. n = 5 per group. (**B**) Left ventricular end diastolic pressure (LVEDP) in Langendorff assays after global I/R injury of hearts from wild-type and GPR35$^{-/-}$ mice. Where indicated, hearts were infused with 500 nM KynA, 100 μM FG-4592, 500 nM quinolinic acid, or vehicle for 10 min (or 30 min where indicated) before the onset of ischemia. n = 4 per group. (**C**) Immunoblot analysis of lysates from Langendorff hearts perfused with 100 μM FG-4592 for the indicated times. (**D**) Myocardial infarct size after in vivo cardiac I/R injury in wild-type mice treated with 5 mg/kg KynA, 10 μg/kg pertussis toxin (PTX), both, or vehicle by intraperitoneal injection. Mice were treated with KynA or vehicle 2 hours before or with PTX 24 hours before the onset of ischemia. n = 4 per group. (**E**) Myocardial infarct size after in vivo cardiac I/R injury in wild-type mice given GPR35 agonists. Where indicated, mice were given 5 mg/kg KynA, 5 mg/kg lodoxamide, 5 mg/kg zaprinast, 5 mg/kg quinolinic acid, or vehicle by intraperitoneal injection 2 hours before the onset of ischemia. n = 4 per group. (**F**) LVEDP in Langendorff assays after global I/R injury. Where indicated, hearts were infused with 500 nM KynA, 1 μM lodoxamide, 1 μM zaprinast, 500 nM quinolinic acid, or vehicle for 10 min before the onset of ischemia. n = 4 per group. (**G**) Immunoblot of wild-type or GPR35$^{-/-}$ hIPS-derived cardiomyocytes stably expressing wild-type or R151A GPR35. (**H**) Fractional survival after simulated I/R injury ex vivo of hIPS cell–derived cardiomyocytes modified as in (G). Cell were pretreated with 20 μM KynA, 20 μM quinolinic acid, or DMSO for 1 hour before I/R. Data are shown as mean ± SD. *P < 0.05.

(Fig. 4D). Therefore, KynA promotes the association of GPR35 with ATPIF1 and ATP synthase, but this interaction is likely indirect.

During ischemia, ATP synthase shifts from ATP synthesis to ATP hydrolysis, with loss of ATP eventually contributing to cell death (*40*). Binding of ATPIF1 to ATP synthase promotes the formation of ATP synthase dimers and thereby inhibits ATP hydrolysis, oxidative phosphorylation, and oxygen consumption (*41*–*44*). ATPIF1 mimetic compounds and transgenic overexpression of ATPIF1 both protect tissues in I/R models (*44*, *45*). Here, KynA increased the recovery of ATP synthase after immunoprecipitation of ATPIF1 and decreased the recovery of ATPIF1 after the immunoprecipitation of ATP synthase in a pertussis toxin–sensitive manner, consistent

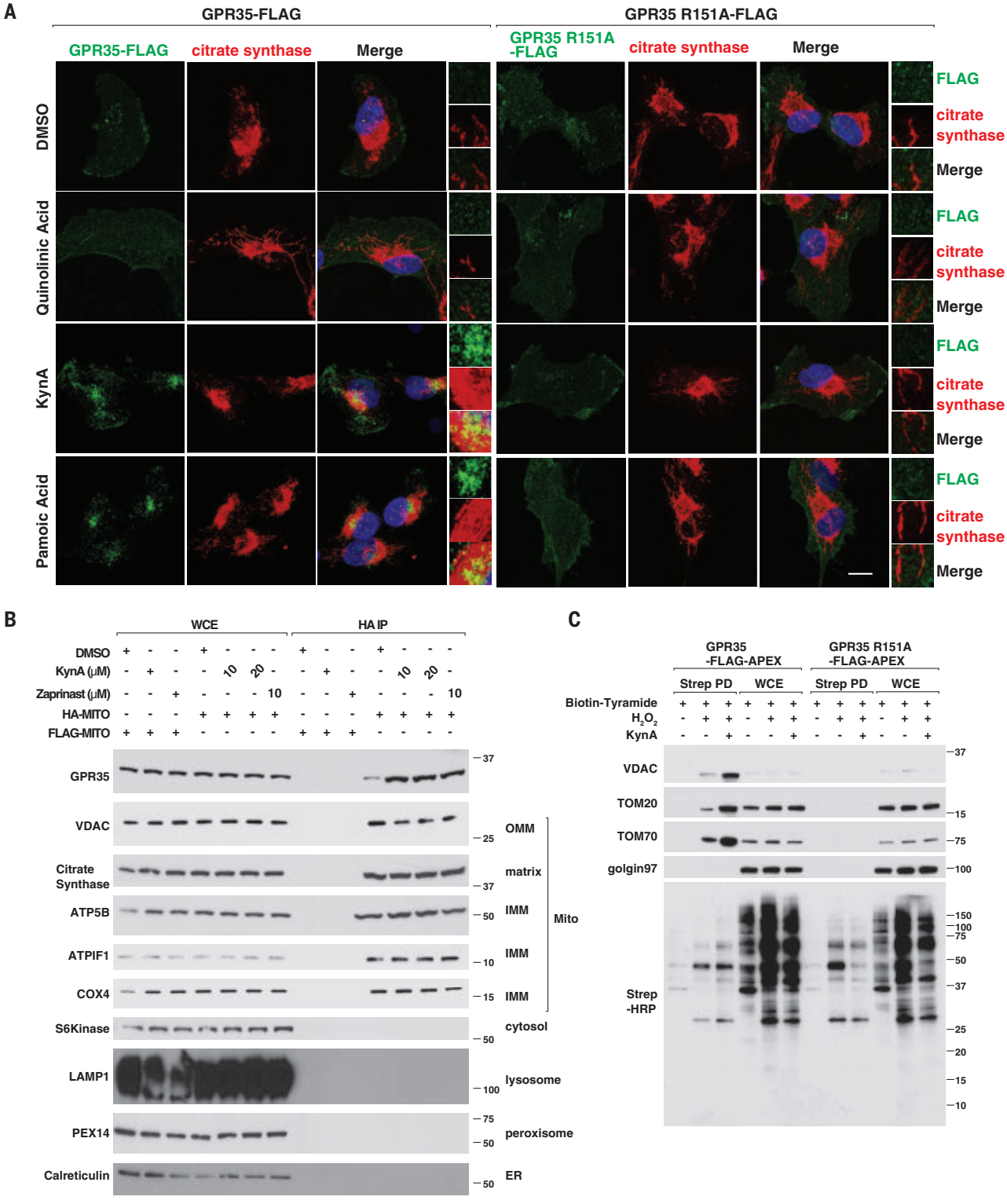


Fig. 3. KynA promotes GPR35 association with mitochondria. (**A**) Confocal microscopy of GPR35$^{-/-}$ mouse neonatal cardiomyocytes stably expressing FLAG-tagged wild-type or R151A GPR35 that were treated with 20 μM quinolinic acid, 20 μM KynA, 1 μM pamoic acid, or DMSO for 20 min. Scale bar, 20 μm. (**B**) Immunoblot analysis of whole-cell extract (WCE) or mitochondria that were rapidly immunopurified by anti-HA immunoprecipitation (HA-IP) from 293T cells engineered to contain HA-tagged mitochondria (HA-MITO) or, as a control, FLAG-tagged mitochondria (FLAG-MITO). Where indicated, the cells were pretreated with 20 μM KynA or 10 μM zaprinast for 20 min before analysis. (**C**) GPR35$^{-/-}$ mouse cardiomyocytes stably expressing FLAG-APEX-tagged wild-type or R151A GPR35 were treated with biotin-tryamide for 30 min and, where indicated, with 20 μM KynA for 20 min and H_2O_2 for 1 min (to enable biotinylation). Cell lysates (WCE) and biotinylated proteins captured on streptavidin agarose (Strep-PD) were resolved by SDS-polyacrylamide gel electrophoresis (PAGE), transferred to nitrocellulose, and immunoblotted with antibodies against the indicated proteins or probed with horseradish peroxidase–conjugated streptavidin (Strep-HRP).

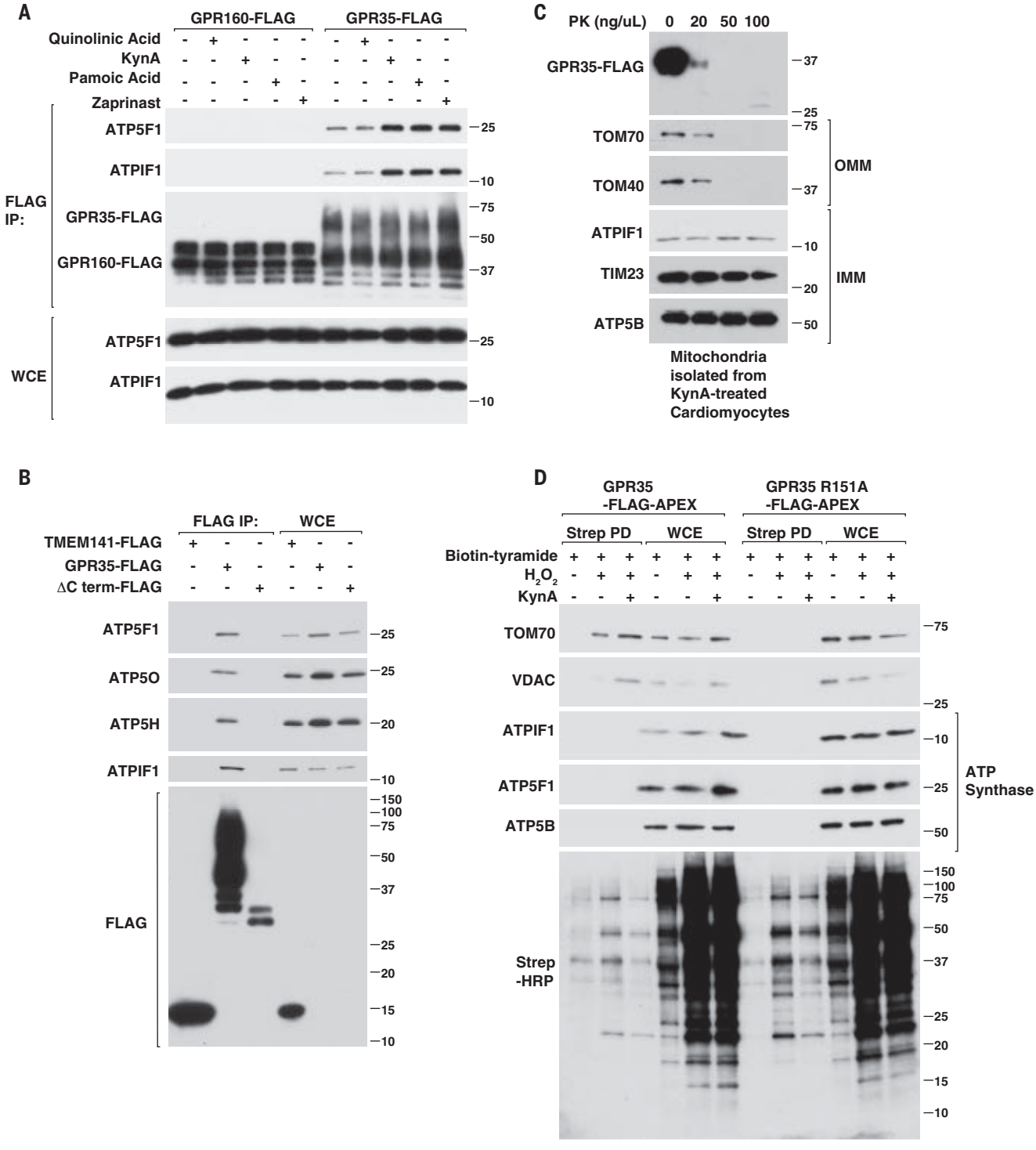


Fig. 4. KynA promotes the GPR35-ATPIF1-ATP synthase interaction. (**A**) Immunoblot analysis of anti-FLAG immunoprecipitates (FLAG-IP) and WCE from mouse neonatal GPR35$^{-/-}$ cardiomyocytes stably expressing GPR160-FLAG or GPR35-FLAG that were treated with 20 μM quinolinic acid, 20 μM KynA, 1 μM pamoic acid, 10 μM zaprinast, or DMSO for 20 min. (**B**) Immunoblot analysis of anti-FLAG immunoprecipitates from human AC16 cardiomyocytes stably expressing TMEM141-FLAG, GPR35 (wild-type)-FLAG, or GPR35 (ΔC terminus)-FLAG. (**C**) Immunoblot analysis of mitochondria isolated from mouse neonatal cardiomyocytes stably expressing GPR35-FLAG that were treated with 20 μM KynA for 20 min. The isolated mitochrondria were incubated with increasing concentrations of proteinase K (PK) for 30 min. (**D**) GPR35$^{-/-}$ mouse neonatal cardiomyocytes stably expressing FLAG-APEX–tagged wild-type or R151A GPR35 were treated with biotin-tryamide for 30 min and, where indicated, with 20 μM KynA for 20 min and H_2O_2 for 1 min (to enable biotinylation). Cell lysates (WCE) and biotinylated proteins captured on streptavidin agarose (STREP-PD) were resolved by SDS-PAGE, transferred to nitrocellulose, and immunoblotted with antibodies against the indicated proteins or probed with Strep-HRP.

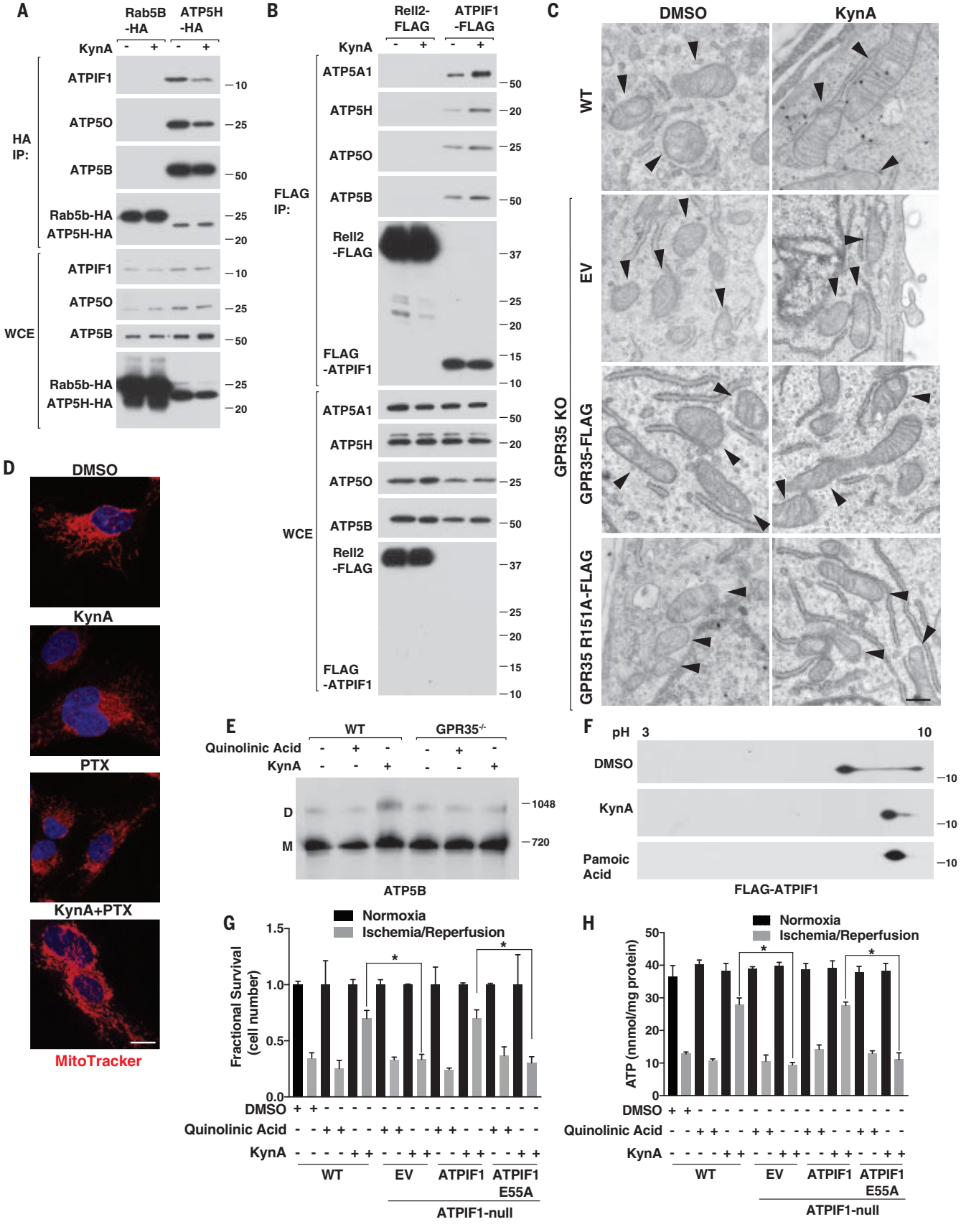

A
Rab5B -HA
ATP5H -HA
KynA
HA IP:
ATPIF1
ATP5O
ATP5B
Rab5b-HA
ATP5H-HA
WCE
B
Rell2-FLAG
ATPIF1 -FLAG
FLAG IP:
ATP5A1
ATP5H
ATP5O
ATP5B
Rell2 -FLAG
FLAG -ATPIF1
C
DMSO
KynA
WT
EV
GPR35 KO
GPR35-FLAG
GPR35 R151A-FLAG
D
DMSO
KynA
PTX
KynA+PTX
MitoTracker
E
WT
GPR35-/-
Quinolinic Acid
D
M
ATP5B
F
pH 3
10
DMSO
KynA
Pamoic Acid
FLAG-ATPIF1
G
Normoxia
Ischemia/Reperfusion
Fractional Survival (cell number)
DMSO
Quinolinic Acid
KynA
WT
EV
ATPIF1
ATPIF1 E55A
ATPIF1-null
H
Normoxia
Ischemia/Reperfusion
ATP (nmol/mg protein)
DMSO
Quinolinic Acid
KynA
WT
EV
ATPIF1
ATPIF1 E55A
ATPIF1-null

Fig. 5. KynA promotes ATP synthase dimerization and ATPIF1 is required for KynA-induced ischemic protection. (**A**) Immunoblot analysis of HA-IP and WCE from mouse neonatal cardiomyocytes stably expressing Rab5B-HA or ATP5H-HA that were or were not treated with 20 μM KynA for 20 min. (**B**) Immunoblot analysis of FLAG-IP and WCE from mouse neonatal cardiomyocytes stably expressing Rell2-FLAG or FLAG-ATPIF1 that were or were not treated with 20 μM KynA for 20 min. (**C**) Electron micrographs of wild-type and GPR35 $^{-/-}$ cardiomyocytes stably expressing GPR35-FLAG, R151A GPR35-FLAG, or empty vector (EV) that were treated with 20 μM KynA or DMSO for 20 min. Arrowheads indicate mitochondria. Scale bar, 500 nm. (**D**) Live-cell fluorescence imaging of mouse neonatal cardiomyocytes treated with 20 μM KynA, 200 ng/ml PTX, or both before staining with 100 nM Mitotracker FM. Scale bar, 20 μm. (**E**) Blue-native page analysis of mitochondria isolated from mouse neonatal cardiomyocytes treated with 20 μM quinolinic acid, 20 μM KynA, or DMSO for 20 min. (**F**) Anti-Flag immunoblot analysis after two-dimensional gel electrophoresis of cell extracts made from mouse neonatal cardiomyocytes stably expressing FLAG-ATPIF1 that were treated with 20 μM KynA, 1 μM pamoic acid, or DMSO for 20 min. (**G** and **H**) Fractional survival (G) and ATP levels (H) of wild-type and ATPIF1$^{-/-}$ hIPS-derived cardiomyocytes stably expressing wild-type or E55A ATPIF1 or EV that were pretreated with 20 μM KynA, 20 μM quinolinic acid, or DMSO for 1 hour before simulated I/R ex vivo. Data are shown as mean ± SD. *P < 0.05.

with the ATPIF1:ATP synthase stoichiometry shifting from 1:1 to 1:2 (Fig. 5, A and B, and fig. S11, A to D). We independently confirmed that KynA promotes ATP synthase dimerization using native gel electrophoresis and immunofluorescence and electron microscopy, which revealed hallmark changes in mitochondrial morphology and mitochondrial cristae, respectively, usually associated with ATP synthase dimerization (Fig. 5, C to E, and figs. S11E and S12, A to C) (*46*).

Because phosphorylation of ATPIF1 by PKA prevents it from inhibiting ATP synthase (*47*), we tested whether GPR35 activation caused dephosphorylation of ATPIF1. Both KynA and the GPR35 agonist pamoic acid blocked ATPIF1 phosphorylation in a GPR35-dependent manner, as shown by characteristic electrophoretic mobility changes after two-dimensional gel electrophoresis (Fig. 5F and fig. S12D).

We used CRISPR/Cas9 to generate ATPIF1$^{-/-}$ cardiomyocytes from hIPS cells and then infected them with wild-type ATPIF1, an ATPIF1 point mutant that cannot bind to ATP synthase (ATPIF1 E55A), or an empty vector (fig. S13, A and B) (*48*). Wild-type ATPIF1, but not ATPIF1 E55A, allowed KynA to prevent ATP consumption and preserve cell viability during simulated ischemia ex vivo (Fig. 5, G and H). Collectively, these results show that KynA ischemic protection depends on ATPIF1.

Although GPCRs are typically thought to function at the plasma membrane, there are multiple examples of mitochondrial GPCRs and G proteins (*49*, *50*). For example, the G proteins guanine nucleotide binding protein-β subunit 2 (Gβ2) and guanine nucleotide binding protein subunit alpha-12 (Gα12) have roles in mitochondria fission versus fusion (*51*, *52*). Moreover, many GPCRs, including GPR35, are internalized once activated, presumably so that they can signal in a spatially distinct manner (*53*). The endoplasmic reticulum has important roles in endosomal trafficking and forms multiple contacts with mitochondria, and a physical interaction between endosomes and mitochondria serves critical functions in cell homeostasis (*54*–*56*). The endosomal trafficking of GPR35 to mitochondria is reminiscent of Rab5, a small GTPase that translocates from early endosomes to mitochondria in response to oxidative stress to confer cytoprotection (*57*).

We found that GPR35 could bind to ATPIF1 and that the latter is necessary for cardioprotection by former. Nonetheless, GPR35 associates with the mitochondrial outer membrane in response to KynA, whereas ATPIF1 is a mitochondrial matrix protein (*41*). Moreover, we did not detect biotinylation of ATPIF1 or ATP synthase by GPR35-APEX, and neither ATPIF1 nor ATP synthase was shown to be proteinase K sensitive. Thus, the interaction of GPR35 with ATPIF1 is not direct, but instead it is likely bridged by one or more proteins, perhaps at contact sites that have been demonstrated between outer and inner mitochondrial membranes and endosome-mitochondrial membrane contact sites (*55*, *58*). Pertussis toxin did not block the internalization of GPR35 by KynA but blocked the KynA-mediated regulated interaction between ATPIF1 and ATP synthase, ATPIF1 dephosphorylation, and cardioprotection; the latter two therefore appear to be linked and to require $G_{i/o}$ signaling (figs. S14A and S15A and Fig. 2D). We propose that GPR35, once activated and translocated to mitochondria, inhibits mitochondrial adenylate cyclase and thereby PKA. This, in turn, would allow ATPIF1 to promote ATP synthase dimerization and prevent ATP hydrolysis. Although KynA might have direct effects on mitochondria and may regulate metabolism, the effects that we observed clearly depend on GPR35 (*59*–*65*).

It is clear that ischemic preconditioning involves both an early and a late phase of ischemic protection, and we have shown that GPR35 is necessary for KynA-mediated cardioprotection whether administered acutely or 24 hours before I/R injury (Fig. 2A) (*66*). Although our data show that GPR35 is necessary in both settings, it remains unknown whether ATPIF1 is required at longer time scales.

Ischemic diseases such as myocardial infarction and stroke are major causes of death in the developed world. Remote ischemic preconditioning has been demonstrated in animal models, but attempts to harness it for therapeutic purposes in humans, such as through repeated hyperinflation of a limb blood pressure cuff before elective heart surgery, have not proven to be successful (*67*). Without an understanding of the underlying mechanism, however, it was impossible to know whether the responsible tissue-protective factor was properly induced by these interventions and if its effects would be mitigated by other factors. Our findings support further exploration of KynA, and more broadly GPR35 agonists, for the prevention and treatment of ischemic injuries. A number of chemically synthetic GPR35 agonists have now been described, including compounds with far greater potency than KynA (*20*, *21*). In addition, our findings provide a unifying explanation for the tissue-protective effects of kynurenine hydroxylase and kyurenine monooxygenase inhibitors. Although these inhibitors were originally developed to prevent the production of neuroexcitatory and neurotoxic tryptophan metabolites, they also promote the accumulation of KynA (*68*, *69*).

REFERENCES AND NOTES

1. B. A. Olenchock *et al*., *Cell* **164**, 884–895 (2016).
2. C. E. Murry, R. B. Jennings, K. A. Reimer, *Circulation* **74**, 1124–1136 (1986).
3. X. Zheng *et al*., *Exp. Mol. Med.* **51**, 1–14 (2019).
4. L. Gellért *et al*., *Eur. J. Pharmacol.* **667**, 182–187 (2011).
5. R. B. Nahomi *et al*., *Int. J. Mol. Sci.* **21**, 1795 (2020).
6. N. Milani-Nejad, P. M. Janssen, *Pharmacol. Ther.* **141**, 235–249 (2014).
7. C. Hilmas *et al*., *J. Neurosci.* **21**, 7463–7473 (2001).
8. T. W. Stone, *Eur. J. Neurosci.* **25**, 2656–2665 (2007).
9. M. H. Mok, A. C. Fricker, A. Weil, J. N. Kew, *Neuropharmacology* **57**, 242–249 (2009).
10. F. Resta *et al*., *Neuropharmacology* **108**, 136–143 (2016).
11. J. Wang *et al*., *J. Biol. Chem.* **281**, 22021–22028 (2006).
12. B. C. DiNatale *et al*., *Toxicol. Sci.* **115**, 89–97 (2010).
13. D. Wang *et al*., *Acta Pharm. Sin. B* **11**, 763–780 (2021).
14. K. D. Min *et al*., *Biochem. Biophys. Res. Commun.* **393**, 55–60 (2010).
15. V. P. Ronkainen *et al*., *Cardiovasc. Res.* **101**, 69–77 (2014).
16. K. Chen *et al*., *J. Cardiovasc. Pharmacol.* **75**, 556–563 (2020).
17. O. Sharmin *et al*., *Sci. Rep.* **10**, 9400 (2020).
18. L. Olivares-González *et al*., *PLOS ONE* **11**, e0166717 (2016).
19. G. Milligan, *Trends Pharmacol. Sci.* **32**, 317–325 (2011).
20. Y. Taniguchi, H. Tonai-Kachi, K. Shinjo, *FEBS Lett.* **580**, 5003–5008 (2006).
21. P. Zhao *et al*., *Mol. Pharmacol.* **78**, 560–568 (2010).
22. A. E. MacKenzie *et al*., *Mol. Pharmacol.* **85**, 91–104 (2014).
23. L. Jenkins *et al*., *Br. J. Pharmacol.* **162**, 733–748 (2011).
24. P. Zhao *et al*., *J. Biol. Chem.* **289**, 3625–3638 (2014).
25. E. V. Moo, J. R. van Senten, H. Bräuner-Osborne, T. C. Møller, *Mol. Pharmacol.* **99**, 242–255 (2021).
26. H. Yang, H. Wang, R. Jaenisch, *Nat. Protoc.* **9**, 1956–1968 (2014).
27. H. Deguchi *et al*., *Circ. J.* **84**, 1028–1033 (2020).
28. W. Bao *et al*., *J. Cardiovasc. Pharmacol.* **56**, 147–155 (2010).
29. T. L. Yeh *et al*., *Chem. Sci.* **8**, 7651–7668 (2017).
30. J. I. Nwogu *et al*., *Circulation* **104**, 2216–2221 (2001).
31. M. Vogler *et al*., *Pflugers Arch.* **467**, 2141–2149 (2015).
32. M. Hölscher *et al*., *J. Biol. Chem.* **286**, 11185–11194 (2011).

33. J. D. Thornton, G. S. Liu, J. M. Downey, *J. Mol. Cell. Cardiol.* **25**, 311–320 (1993).
34. E. L. Huttlin *et al.*, *Cell* **184**, 3022–3040.e28 (2021).
35. W. W. Chen, E. Freinkman, T. Wang, K. Birsoy, D. M. Sabatini, *Cell* **166**, 1324–1337.e11 (2016).
36. S. S. Lam *et al.*, *Nat. Methods* **12**, 51–54 (2015).
37. S. Y. Lee *et al.*, *Cell Rep.* **15**, 1837–1847 (2016).
38. C. Dong, C. M. Filipeanu, M. T. Duvernay, G. Wu, *Biochim. Biophys. Acta* **1768**, 853–870 (2007).
39. M. Fujikawa, H. Imamura, J. Nakamura, M. Yoshida, *J. Biol. Chem.* **287**, 18781–18787 (2012).
40. G. J. Grover *et al.*, *Am. J. Physiol. Heart Circ. Physiol.* **287**, H1747–H1755 (2004).
41. J. J. García, E. Morales-Ríos, P. Cortés-Hernandez, J. S. Rodríguez-Zavala, *Biochemistry* **45**, 12695–12703 (2006).
42. L. Formentini, M. Sánchez-Aragó, L. Sánchez-Cenizo, J. M. Cuezva, *Mol. Cell* **45**, 731–742 (2012).
43. L. Sánchez-Cenizo *et al.*, *J. Biol. Chem.* **285**, 25308–25313 (2010).
44. L. Formentini *et al.*, *EMBO J.* **33**, 762–778 (2014).
45. F. Ivanes *et al.*, *Br. J. Pharmacol.* **171**, 4193–4206 (2014).
46. A. García-Aguilar, J. M. Cuezva, *Front. Physiol.* **9**, 1322 (2018).
47. J. García-Bermúdez *et al.*, *Cell Rep.* **12**, 2143–2155 (2015).
48. N. Ichikawa *et al.*, *J. Biochem.* **130**, 687–693 (2001).
49. R. Valenzuela *et al.*, *Cell Death Dis.* **7**, e2427 (2016).
50. A. Belous *et al.*, *J. Cell. Biochem.* **92**, 1062–1073 (2004).
51. J. Zhang *et al.*, *Nat. Commun.* **1**, 101 (2010).
52. A. V. Andreeva, M. A. Kutuzov, T. A. Voyno-Yasenetskaya, *FASEB J.* **22**, 2821–2831 (2008).
53. B. T. Lobingier, M. von Zastrow, *Traffic* **20**, 130–136 (2019).
54. A. D. Sheftel, A. S. Zhang, C. Brown, O. S. Shirihai, P. Ponka, *Blood* **110**, 125–132 (2007).
55. K. Todkar, L. Chikhi, M. Germain, *Mitochondrion* **49**, 284–288 (2019).
56. Y. Elbaz-Alon *et al.*, *Nat. Commun.* **11**, 3645 (2020).
57. F. Hsu *et al.*, *eLife* **7**, e32282 (2018).
58. A. S. Reichert, W. Neupert, *Biochim. Biophys. Acta* **1592**, 41–49 (2002).
59. H. Baran *et al.*, *Int. J. Tryptophan Res.* **9**, 17–29 (2016).
60. H. Baran *et al.*, *Pharmacology* **62**, 119–123 (2001).
61. H. Baran *et al.*, *Life Sci.* **72**, 1103–1115 (2003).
62. D. Y. Lee *et al.*, *Eur. J. Cell Biol.* **87**, 389–397 (2008).
63. L. Juhász *et al.*, *Front. Med. (Lausanne)* **7**, 566582 (2020).
64. G. Schneditz *et al.*, *Sci. Signal.* **12**, eaau9048 (2019).
65. L. Z. Agudelo *et al.*, *Cell Metab.* **27**, 378–392.e5 (2018).
66. R. Bolli, *Circ. Res.* **87**, 972–983 (2000).
67. D. J. Hausenloy *et al.*, *N. Engl. J. Med.* **373**, 1408–1417 (2015).
68. D. Zwilling *et al.*, *Cell* **145**, 863–874 (2011).
69. A. Cozzi, R. Carpenedo, F. Moroni, *J. Cereb. Blood Flow Metab.* **19**, 771–777 (1999).

ACKNOWLEDGMENTS

We thank the members of the Kaelin laboratory and J. Moslehi (UCSF) for helpful discussions and critical reading of the manuscript. Special thanks go to the Cardiovascular Physiology Core (Longwood) at BWH/HMS for assistance with cardiac physiology experiments, as well as B. Olenchock and R. Liao for experimental advice. We also thank T. Sakmar for critical reading of the manuscript and V. Mootha for helpful discussions and experimental advice. **Funding:** W.G.K. is supported by the National Institutes of Health (grant 5R35CA210068) and is an Howard Hughes Medical Institute (HHMI) investigator. G.A.W. is an HHMI Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. I.P.D. is a Fellow of the George and Marie Vergottis Foundation. J.D.M. is supported by Richard A. and Susan F. Smith President's Innovation Award, the Sidman Family Foundation, the Michael B. Rukin Charitable Foundation, the Kenneth C. Griffin Charitable Research Fund, The George and Marie Vergottis Foundation, and the Boston Investment Council. **Author contributions:** G.A.W. performed experiments and, together with W.G.K., designed experiments, analyzed data, and assembled and wrote the manuscript. W.Y. performed two-dimensional gel electrophoresis experiments and assisted in biochemical isolation of GPCRs. I.P.D., R.S.N., M.Y.S., and T.D. designed and performed the rabbit I/R model experiments under the direction and guidance of J.D.M. **Competing interests:** W.G.K. has financial interests in Lilly Pharmaceuticals, Fibrogen, Cedilla Therapeutics, IconOvir Bio, Nextech Invest, Tango Therapeutics, LifeMine Therapeutics, Circle Pharma, and Casdin Capital. J.D.M. is a consultant for Cellvie Scientific, Inc. The remaining authors declare no competing interests. **Data and materials availability:** All data needed to evaluate and reproduce the conclusions in the paper are present in the paper and/or the supplementary materials. All plasmids are available from the authors at request.

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm1638

Materials and Methods
Figs. S1 to S15
References (*70*–*74*)
MDAR Reproducibility Checklist

Submitted 30 August 2021; accepted 28 June 2022
10.1126/science.abm1638

PLANT SCIENCE

Meiotic exit in *Arabidopsis* is driven by P-body–mediated inhibition of translation

Albert Cairo[1], Anna Vargova[1], Neha Shukla[1], Claudio Capitao[2], Pavlina Mikulkova[1], Sona Valuchova[1], Jana Pecinkova[1], Petra Bulankova[2]†, Karel Riha[1]*

Meiosis, at the transition between diploid and haploid life cycle phases, is accompanied by reprograming of cell division machinery and followed by a transition back to mitosis. We show that, in *Arabidopsis*, this transition is driven by inhibition of translation, achieved by a mechanism that involves processing bodies (P-bodies). During the second meiotic division, the meiosis-specific protein THREE-DIVISION MUTANT 1 (TDM1) is incorporated into P-bodies through interaction with SUPPRESSOR WITH MORPHOGENETIC EFFECTS ON GENITALIA 7 (SMG7). TDM1 attracts eIF4F, the main translation initiation complex, temporarily sequestering it in P-bodies and inhibiting translation. The failure of *tdm1* mutants to terminate meiosis can be overcome by chemical inhibition of translation. We propose that TDM1-containing P-bodies down-regulate expression of meiotic transcripts to facilitate transition of cell fates to postmeiotic gametophyte differentiation.

Differentiation of plant gametes is a multistep process that involves meiosis, a reductional cell division that produces haploid cells from diploid progenitors. Meiotic chromosome segregation requires modification of the cell division program that is implemented through ordered expression of specialized proteins, reflected in transcriptional changes (*1*, *2*). The duration of meiosis can range from half a day to decades, as in the case of human oogenesis, and the prolonged prophase I, during which chromosomes are condensed, limits instructive transcription. Therefore, the activities of many meiotic genes are regulated by posttranscriptional and translational mechanisms (*3*, *4*). Posttranscriptional gene regulation has been linked to biomolecular condensates. Studies in yeast indicate such condensates in meiosis, in which late meiotic genes are translationally repressed in meiosis I by inclusion in amyloid-like aggregates. These aggregates dissolve during meiosis II, releasing translational inhibition and allowing synthesis of proteins required for meiotic exit (*5*, *6*).

In animals, haploid products of meiosis directly differentiate into gametes. In plants, fungi, and some protists, however, haploid spores undergo mitotic divisions to form haploid free-living organisms or reproductive structures. In such haploid cell lineages, completion of meiosis must be followed by reprograming of cell division machinery and a transition back to mitosis.

In angiosperm plants, differentiation of gametes includes the formation of pollen and the embryo sac, rudimentary multicellular gametophytes generated by postmeiotic (i.e., mitotic) divisions of haploid cells. In *Arabidopsis*, meiotic exit and the diploid-to-haploid transition require the evolutionarily conserved nonsense-mediated mRNA decay (NMD) factor SUPPRESSOR WITH MORPHOGENETIC EFFECTS ON GENITALIA 7 (SMG7), together with the plant-specific protein THREE-DIVISION MUTANT 1 (TDM1) (*7*–*9*). Mutants deficient in these proteins fail to terminate male meiosis and do not form microspores. Instead, haploid nuclei produced in meiosis undergo multiple cycles of spindle reassembly and chromatin recondensation without intervening DNA synthesis, resulting in aborted polyads (*8*, *10*, *11*). Meiotic exit is driven by TDM1. Its unscheduled activation in meiosis I results in premature termination of meiosis and formation of diploid microspores (*9*), but the molecular mechanism by which TDM1 terminates meiosis is unknown. Here we demonstrate that, in *Arabidopsis*, meiotic exit and the diploid-to-haploid transition are mediated by inhibition of translation through temporally limited

[1]Central European Institute of Technology (CEITEC), Masaryk University, 625 00 Brno, Czech Republic. [2]Gregor Mendel Institute (GMI), Austrian Academy of Sciences (OAW), Vienna BioCenter (VBC), 1030 Vienna, Austria.
***Corresponding author. Email: karel.riha@ceitec.muni.cz**
†Present address: VIB Center for Plant Systems Biology, Technologiepark 71, 9052, Ghent, Belgium.

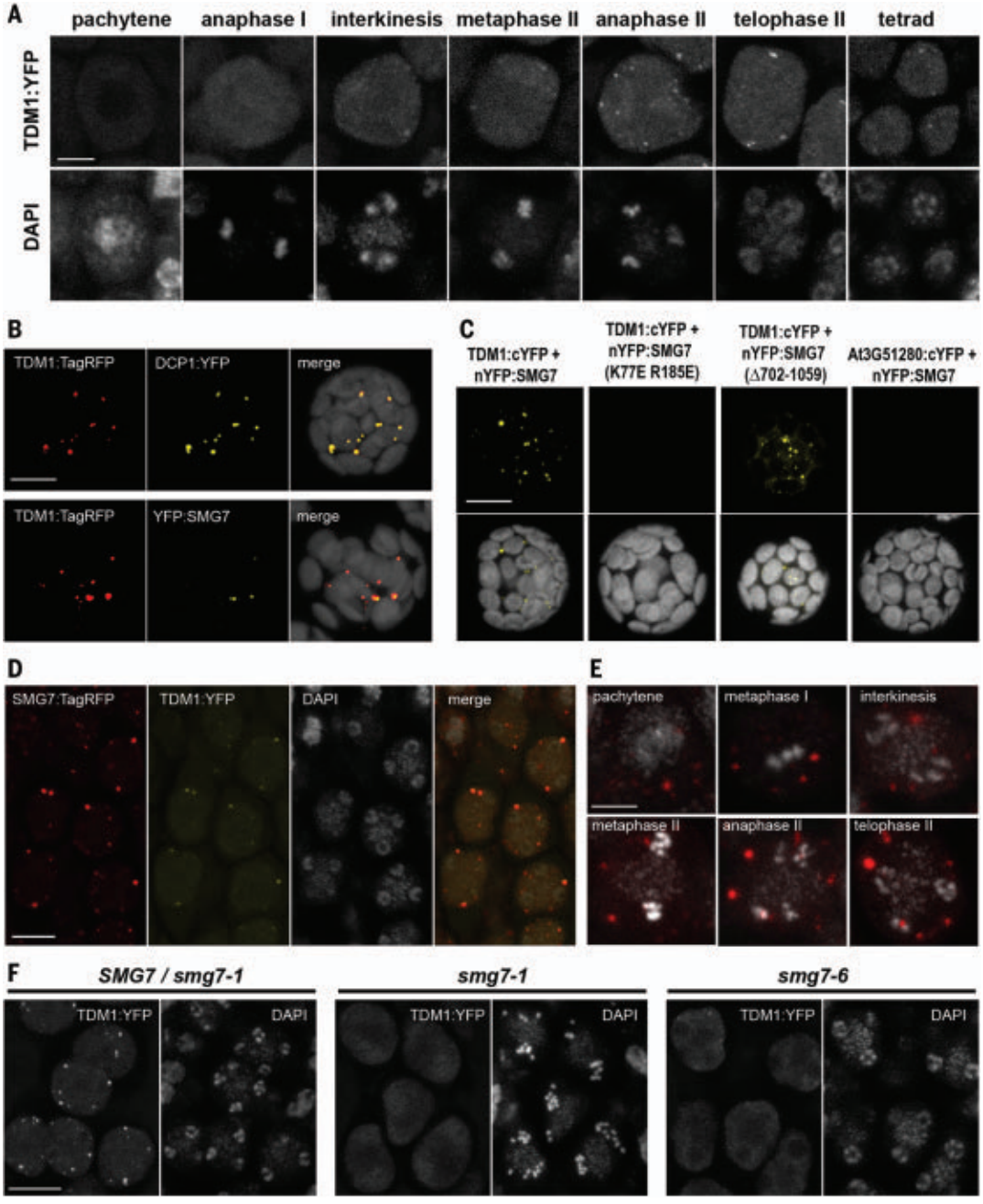


Fig. 1. SMG7-dependent localization of TDM1 in P-bodies. (**A**) Localization pattern of TDM1 in PMCs from *tdm1-4* plants complemented with *pTDM1::TDM1:YFP*. DNA was counterstained with 4′,6-diamidino-2-phenylindole (DAPI). (**B**) Colocalization of DCP1 (*35S::DCP1:YFP*) and SMG7 (*35S::YFP:SMG7*) with TDM1 (*35S::TagRFP:TDM1*) in mesophyll protoplasts 16 hours after cotransfection. (**C**) TDM1-SMG7 interaction assessed by BiFC assay in protoplasts. The TDM1 paralog At3G51280 was used as a specificity control. Representative images from at least three independent transfections are shown. In (B) and (C), autofluorescence from chloroplasts is indicated in gray. (**D**) Colocalization of SMG7 and TDM1 in telophase II PMCs expressing *pSMG7::SMG7:TagRFP* and *pTDM1::TDM1:YFP*. DNA was counterstained with DAPI. (**E**) Meiotic localization of SMG7 in PMCs from *smg7-1* mutants complemented with *pSMG7::SMG7:TagRFP*. DNA was counterstained with DAPI. (**F**) Localization of TDM1 at the stages corresponding to telophase II in the indicated *smg7* mutants expressing *pTDM1::TDM1:YFP*. Scale bars: 5 μm in (A) and (D); 10 μm in (B), (C), (E), and (F).

sequestration of translation initiation complexes by TDM1 in processing bodies (P-bodies).

TDM1 is recruited to P-bodies by SMG7

TDM1 is a meiosis-specific protein expressed from late prophase I, as evidenced by reporter constructs and protein immunodetection (*9*, *12*). To determine the localization of TDM1 during meiosis, we generated *pTDM1::TDM1:YFP* and *pTDM1::TagRFP:TDM1* reporter lines (fig. S1). Whereas TDM1 shows mostly uniform cytoplasmic distribution in pollen mother cells (PMCs) during meiosis I, it forms prominent cytoplasmic foci in meiosis II that decline after tetrad formation (Fig. 1A, fig. S2, and movie S1). These foci are reminiscent of P-bodies, cytoplasmic condensates generated by liquid-liquid phase separation (LLPS) that predominantly consist of RNA decay proteins (*13*). We confirmed that these foci were P-bodies by colocalizing ectopically expressed TDM1:TagRFP in *Arabidopsis* mesophyll protoplasts with DECAPPING PROTEIN 1 (DCP1), an mRNA decapping factor, and SMG7 (Fig. 1B and figs. S3 and S4), both of which are core components of P-bodies (*14*). Formation of LLPS condensates depends on intracellular protein concentration. To test that the observed signal distribution was not caused by ectopic protein overexpression, we confirmed the colocalization of SMG7 and TDM1 at meiotic telophase II in PMCs stably expressing the respective reporter constructs from native promoters (Fig. 1D and fig. S1). We found that P-bodies marked with SMG7:TagRFP are present throughout all stages of meiosis, not just in meiosis II (Fig. 1E). This indicates that TDM1 does not elicit P-body formation specifically in meiosis II but that it is recruited to preexisting P-bodies.

tdm1 is genetically epistatic to *smg7* with respect to the meiotic exit phenotype (*8*, *11*), which, together with the colocalization in P-bodies, suggests a mechanistic link between these proteins. SMG7 is a phosphoserine binding

protein that has so far been implicated in only NMD (*15*). In NMD, SMG7 binds phosphorylated UP-FRAMESHIFT 1 (UPF1) that marks mRNAs containing premature termination codon, and mediates their recruitment to P-bodies (*16*, *17*). However, the meiotic function of SMG7 is independent of NMD (*18*, *19*). We therefore hypothesized that, analogous to its function in NMD, SMG7 binds and recruits TDM1 to P-bodies.

We observed SMG7 interaction with TDM1 in P-bodies by bimolecular fluorescence complementation (BiFC) in mesophyll protoplasts (Fig. 1C). SMG7 has a tripartite structure that consists of an N-terminal 14-3-3 like domain, a central helical domain, and an evolutionarily diverged C terminus (*15*) (fig. S4A). A combination of two mutations in conserved residues of the 14-3-3 phosphoserine binding pocket, Lys^{77}→Glu (K77E) and Arg^{185}→Glu (R185E) (*15*, *17*), abolished the SMG7-TDM1 interaction (Fig. 1C). P-bodies are archetypes of cytoplasmic LLPS biocondensates, whose formation is driven by multivalent interactions that often include proteins with intrinsically disordered regions (*20*). We found that the C terminus of SMG7 is highly disordered and contains a prion-like domain enriched in polar amino acids that contribute to LLPS (*21*) (figs. S4 and S5). Deletion of the C terminus or the prion-like domain decreased localization of SMG7 to P-bodies but did not abolish the SMG7-TDM1 interaction (Fig. 1C and fig. S4B). Together, these data suggest that SMG7 acts as an adaptor protein that binds TDM1 at its 14-3-3 domain and recruits it to P-bodies through the intrinsically disordered region in its C terminus. Indeed, TDM1 fails to form cytoplasmic foci in PMCs of *Arabidopsis smg7-1* null mutants, and its association with P-bodies is reduced in *smg7-6* plants, which have a truncation in the SMG7 prion-like domain (Fig. 1F and fig. S4A).

TDM1 mediates retention of eIF4F in P-bodies

We used two approaches to identify additional genes involved in meiotic exit. First, we performed a genetic suppressor screen for mutations that improve pollen formation and fertility of semifertile *smg7-6* plants (*11*). Two suppressor lines, *EMS43* and *EMS98*, harbored mutations in the *At2G24050* gene, which encodes the eukaryotic translation initiation factor eIFiso4G2 (Fig. 2A, fig. S6, and data S1 and S2). The causality of these mutations was validated by allelic test and genetic complementation (fig. S7). A portion of PMCs in *EMS43* and *EMS98* lines formed tetrads (fig. S8), indicating that down-regulation of eIFiso4G2 facilitates meiotic exit in SMG7-deficient plants.

In the second approach, we used a yeast two-hybrid assay to screen *Arabidopsis* cDNA libraries for TDM1-interacting proteins. The *Arabidopsis* translation initiation factors eIFiso4G2 and its homolog eIFiso4G1 were among the top hits (data S3). eIF4G proteins form a scaffold of the main eukaryotic translation initiation complex eIF4F, and *Arabidopsis* contains three eIF4G paralogs—canonical eIF4G and its shortened isoforms eIFiso4G1 and eIFiso4G2 (*22*), which we collectively refer to as eIF4Gs. The interaction between eIFiso4G2 and TDM1 was validated by in vitro coimmunoprecipitation (fig. S9).

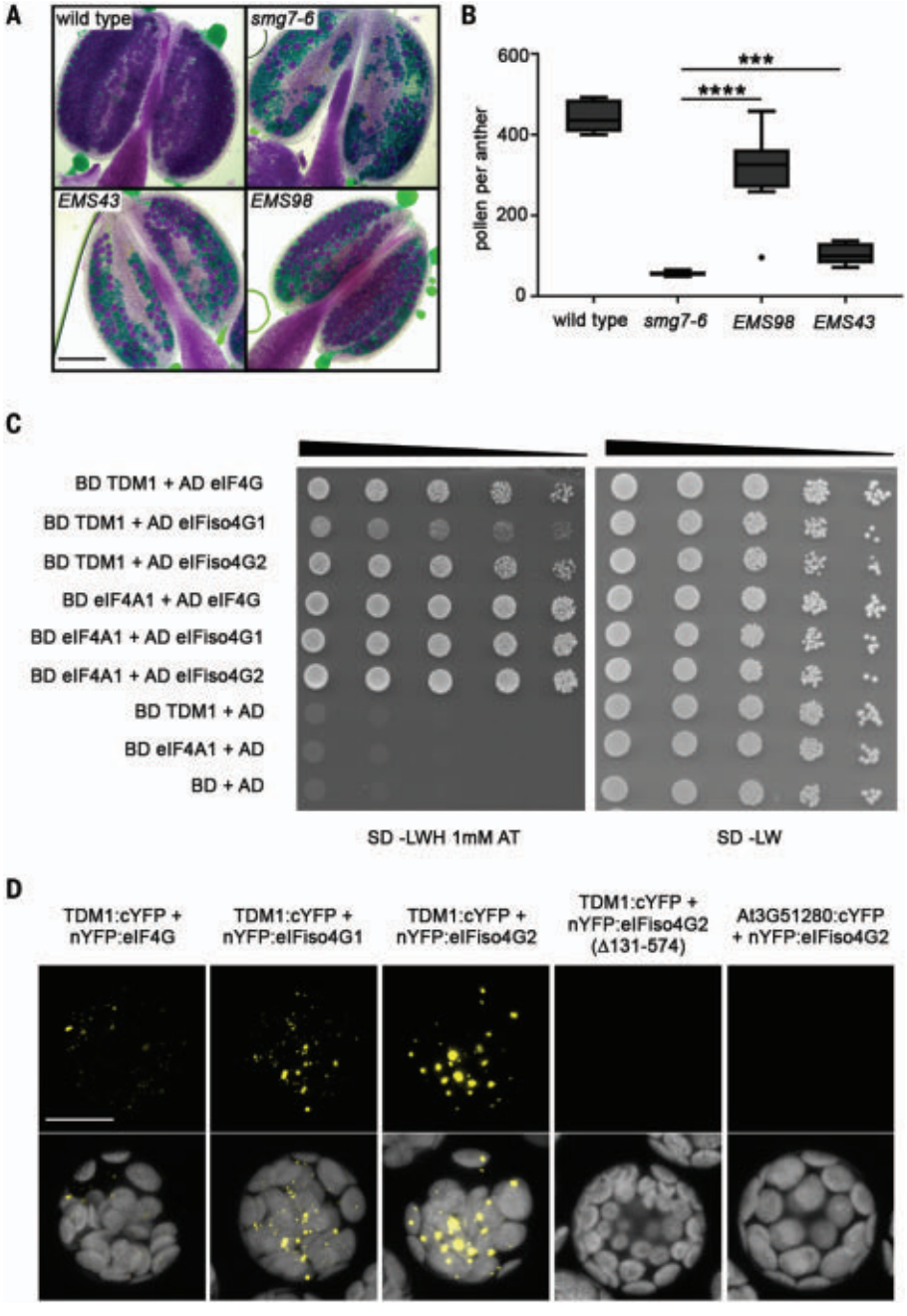


Fig. 2. Genetic and biochemical interactions between eIF4Gs and SMG7/TDM1. (**A**) Alexander staining of anthers and (**B**) quantification of pollen count per anther in *EMS98* and *EMS43* lines (n = 8 anthers). Significance of the difference is indicated (two-tailed t test; ***$P < 0.005$; ****$P < 0.0001$). Scale bar in (A): 100 μm. (**C**) TDM1- eIF4Gs interactions assessed by yeast two-hybrid assay. A truncated eIF4G protein composed of 819 to 1529 amino acids and the full-length eIFiso4G1 and eIFiso4G2 were used in the assay. Dilution series of yeast cells containing pGBKT7 (bait) and pGADT7 (pray) plasmids were spotted on synthetic media lacking either Leu and Trp (SD-LW) or Leu, Trp, and His and were supplemented with 1 mM 3-amino-1,2,3-triazole (SD-LWH 1mM AT). (**D**) Interaction between TDM1 and eIF4Gs assessed by BiFC in protoplasts. TDM1 paralog At3G51280 and eIFis4G2 lacking the central HEAT1 domain were used as a negative control. Chloroplast autofluorescence is indicated in gray. Scale bar: 10 μm.

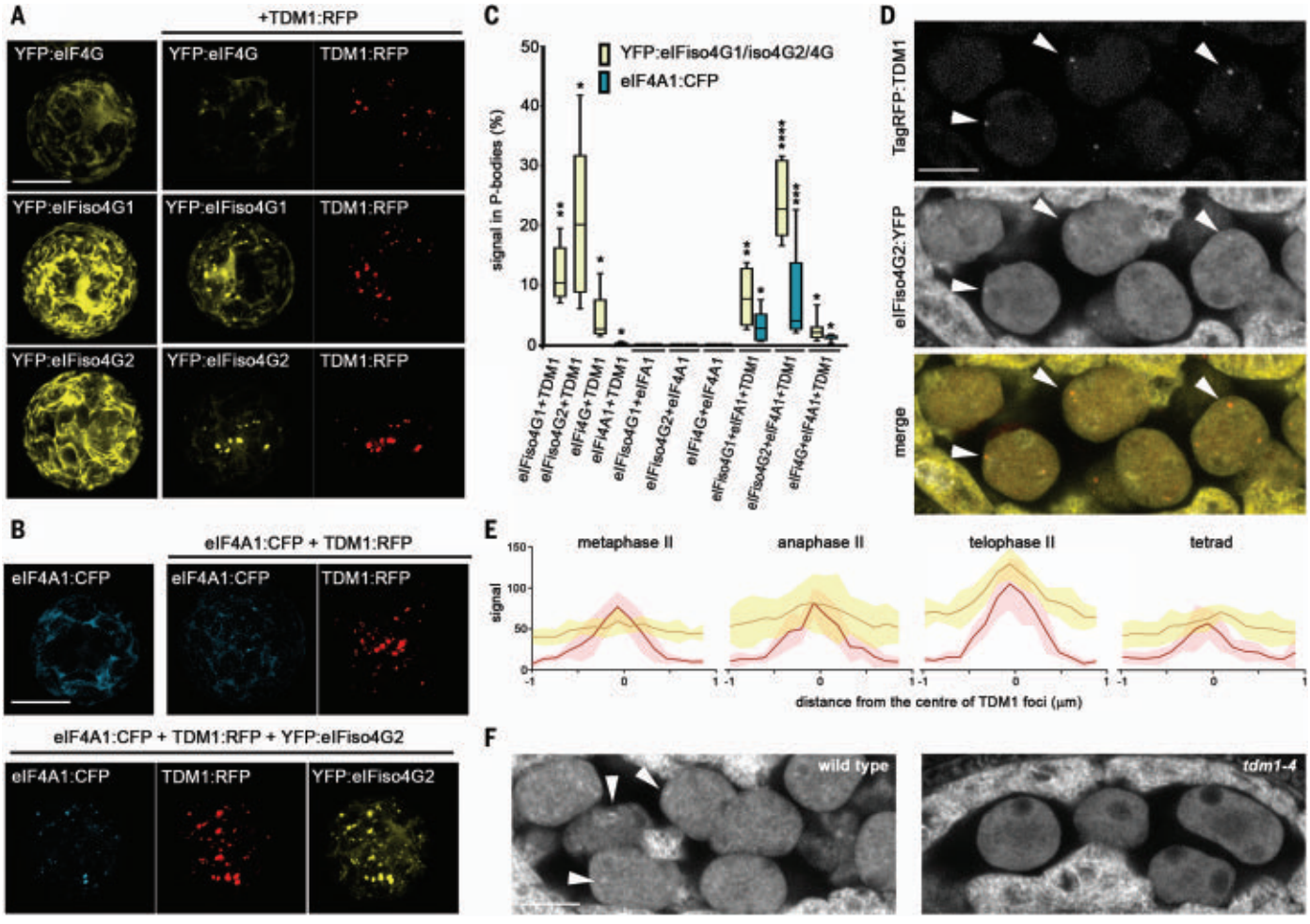


Fig. 3. TDM1 recruits eIF4Gs to P-bodies. Colocalization of (**A**) eIF4Gs (*35S::YFP:eIF4Gs*) and TDM1 (*35S::TDM1:TagRFP*) and (**B**) eIF4A1 (*35S::eIF4A1:CFP*), eIFiso4G2 (*35S::YFP:eIFiso4G2*), and TDM1 (*35S::TDM1:TagRFP*) in mesophyll protoplasts. (**C**) Quantification of the YFP:eIF4Gs and eIF4A1:CFP signals in P-bodies from (A) and (B) from total signal in a cell. Error bars indicate SD (n = 5 to 7). Significance of the difference between sample and respective control without TDM1 is indicated (two tailed t test; *P < 0.05; **P < 0.01; ***P < 0.005; ****P < 0.0001). (**D**) Colocalization of TDM1 and eIFiso4G2 in telophase II PMCs expressing *peIFiso4G2::eIFiso4G2:YFP* and *pTDM1::TagRFP:TDM1*. Arrowheads indicate signal in cytoplasmic foci. The signal in granules represents 1.2 ± 0.2% of the total signal (n = 8). (**E**) Signal intensity profiles of TDM1 (red) and eIFiso4G2 (yellow) centered at P-bodies during meiosis II. Shaded areas around lines indicate SDs (n = 8 to 9). (**F**) Representative pictures of eIFiso4G2 localization in telophase II PMCs from wild-type and *tdm1-4* mutants (five anthers were scored for each genotype). Arrowheads indicate eiFiso4G2 foci. Scale bars: 10 µm.

Pairwise yeast two-hybrid assays showed that TDM1 binds to all three paralogs (Fig. 2C), though the interactions appear to be weaker than between the eIF4Gs and another subunit of the eIF4F complex, the eIF4A1 helicase (fig. S9B). The TDM1-eIFiso4G2 interaction is mediated by the C-terminal portion of TDM1 and the central region of eIFiso4G2 encompassing the HEAT1 domain (fig. S9, B to D).

BiFC assays showed that the interaction between TDM1 and eIF4G factors takes place in P-bodies (Fig. 2D). However, the eIF4F complex is not a constituent of P-bodies (*14*, *23*), and *Arabidopsis* eIF4G paralogs are uniformly localized in the cytoplasm of mesophyll cells (Fig. 3A). Because TDM1 is normally expressed only in meiocytes and absent in mesophyll cells, we tested whether the association of eIF4Gs with P-bodies is caused by TDM1. Indeed, co-expression of TDM1 in mesophyll protoplasts led to relocalization of eIF4Gs into P-bodies, whereas they remained in cytoplasm when coexpressed with SMG7 (Fig. 3, A and C, and fig. S10). We next assessed whether localization of the eIF4A1 helicase is also affected by TDM1. Although we observed only a small portion of eIF4A1 in P-bodies when coexpressed with TDM1, this portion is increased as much as ninefold when TDM1 was coexpressed with individual eIF4Gs (Fig. 3, B and C, and fig. S11). These data suggest that TDM1 recruits eIF4A1, and likely the entire eIF4F complex, into P-bodies through interaction with its eIF4G scaffold subunits.

The *Arabidopsis* eIF4Gs act redundantly, although they exhibit distinct functional specificities in the translation of different transcripts (*22*, *24*). C-terminally–tagged gene-YFP reporters expressed from native promoters revealed that all eIF4Gs are expressed in PMCs throughout meiosis, with eIFiso4G2 being the most abundant (fig. S12). Whereas eIF4Gs are mainly distributed throughout the cytoplasm, eIF4G and eIFiso4G2 exhibit a more dynamic pattern, associating with the spindle in anaphase I and granules during meiosis II (Fig. 3D and fig. S12). The eIFiso4G2 granules colocalize with TDM1, indicating that they represent P-bodies, and a quantitative analysis of the signal revealed a peak of TDM1-eIFiso4G2 association in telophase II (Fig. 3, D and E, and fig. S13). The granules are TDM1 dependent, as they do not form in *tdm1*-null mutants (Fig. 3F). These data validate our observations in mesophyll protoplasts and demonstrate the TDM1-dependent recruitment of eIFiso4G2 (and likely of eIF4G) to P-bodies during telophase II.

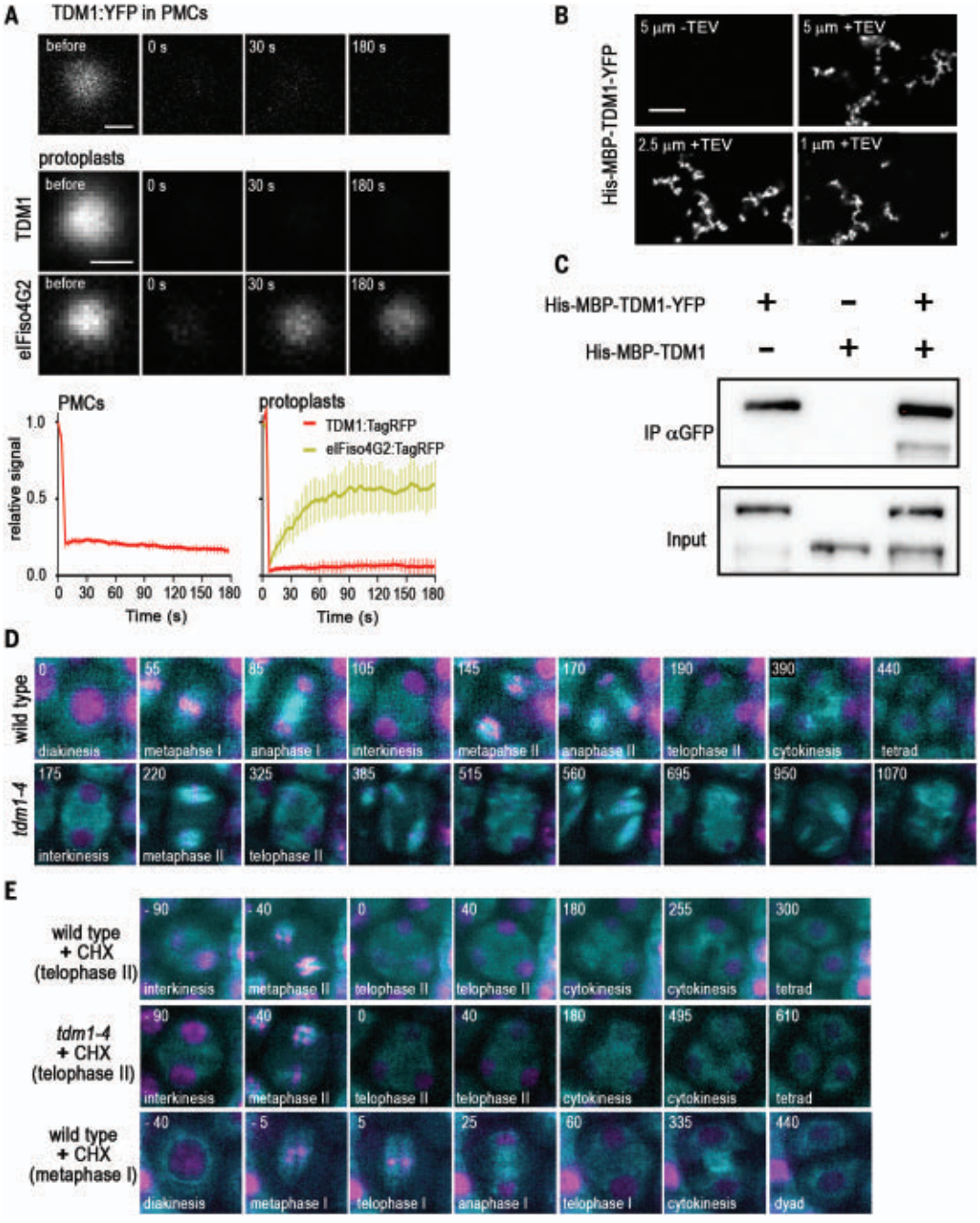


Fig. 4. Inhibition of translation drives meiotic exit. (**A**) FRAP of TDM1:YFP, TDM1:TagRFP, and eIFiso4G2:TagRFP signals in P-bodies. A line harboring *pTDM1::TDM1:YFP* was used to measure FRAP in PMCs, whereas transient coexpression of either *35S::TDM1:TagRFP* with *35S::YFP:eIFiso4G2* or *35S::eIFiso4G2:TagRFP* with *35S::TDM1:YFP* was performed to measure FRAP in protoplasts. Charts depict quantification of signal recovery in PMCs and protoplasts over 180 s after photobleaching. Error bars indicate SDs from three measurements. Scale bar: 1 µm. (**B**) Micrographs of TDM1:YFP aggregates formed in vitro 30 min after the addition of tobacco etch virus (TEV) protease to remove the His-MBP solubility tag by TEV protease. Scale bar: 5 µm. (**C**) TDM1 multimerization detected by in vitro protein immunoprecipitation (IP). IP was performed using α-GFP nanobody, and proteins were detected by Western blotting with an α-His antibody. (**D**) Time-lapse series of wild-type and *tdm1-4* PMCs expressing *pRPS5A::G3GFP:TUA5* and *pHTA10::HTA10:TagRFP* for microtubule (magenta) and chromatin (violet) labeling, respectively. Time in minutes is indicated, with diakinesis set as time point zero; meiosis I is depicted for only the wild type. (**E**) Time-lapse series of PMCs treated with CHX at time point 0. Time in minutes relative to the CHX treatment is indicated.

In contrast to mesophyll protoplasts, where TDM1 was ectopically expressed from a strong promoter, only a fraction (1.2%) of the eIFiso4G2 in PMCs is localized to cytoplasmic granules (Fig. 3D). This is consistent with the observation that the expression level of TDM1 in PMCs is lower than for any of the eIF4Gs (fig. S12B).

Meiotic exit is driven by inhibition of translation

Biocondensates formed by LLPS, such as P-bodies, typically display a dynamic exchange of their protein constituents with the surrounding environment (*20*). Notably, fluorescence recovery after photobleaching (FRAP) showed no recovery of TDM1:TagRFP signal in P-bodies in either PMCs or mesophyll protoplasts (Fig. 4A), and this behavior was also recapitulated in TDM1:YFP condensates formed in vitro (fig. S14). TDM1 has a tendency to aggregate (Fig. 4B), and it multimerizes when kept soluble by fusion with the maltose binding protein (Fig. 4C). These data indicate that TDM1 likely forms multimeric structures that are rigidly embedded within P-bodies. By contrast, eIFiso4G2 exhibits a more dynamic association with TDM1-containing P-bodies, as its FRAP signal rapidly recovers (Fig. 4A). Thus, despite its specific interaction with TDM1, eIFiso4G2 can rapidly shuttle between P-bodies and the cytoplasm.

These observations suggest that, during meiosis II, TDM1 bestows P-bodies with the capability to temporarily recruit eIF4F complexes—and, by this mechanism, to terminate meiosis and enable the transition to gametophytic development. Because P-bodies are associated with translational repression and because *eifiso4G2* mutations reverse the inability of *smg7-6* plants to exit meiosis (Fig. 2, A and B), we hypothesized that termination of meiosis is implemented through inhibition of

translation. In support of this, TDM1 impaired in vitro translation in wheat germ extract (fig. S15A) and suppressed expression of *TCS::GFP* cytokinin inducible reporter in protoplasts (fig. S15, B and C). To test whether TDM1 facilitates meiotic exit through translational repression, we assessed the impact of cycloheximide (CHX), an inhibitor of eukaryotic translation, on meiotic progression in *tdm1* mutants. Live-cell imaging of meiosis, using a GFP:TUA5 reporter to visualize microtubules and HTA10:TagRFP to visualize chromatin (*25*), showed that cytokinesis and tetrad formation normally occur ~4 hours after telophase II in controls as well as in PMCs treated with CHX at telophase II (movies S2 and S4 and Fig. 4, D and E). TDM1-null plants failed to undergo cytokinesis and instead entered multiple cycles of chromatin condensation-decondensation and spindle assembly-disassembly (movie S3 and Fig. 4D). By contrast, treatment of *tdm1* PMCs with CHX at telophase II prevented these postmeiotic cycles of aberrant chromosome segregation and induced orderly cytokinesis and tetrad formation in all scored meiocytes (movie S5 and Fig. 4E; 87 PMCs scored). TDM1 is inhibited during meiosis I by phosphorylation with CDKA;1-TAM. Its unscheduled activation, either by a phospho-site mutation or inactivation of the cyclin TAM, leads to cytokinesis after telophase I and formation of dyads and diploid microspores (*9*). Application of CHX in wild-type PMCs during metaphase or anaphase I led to premature cytokinesis and dyad formation, recapitulating the ectopic activation of TDM1 in all 81 scored meiocytes (movie S6 and Fig. 4E). These observations show that TDM1 mediates meiotic exit by inhibiting translation.

Discussion

In this study, we dissected the molecular function of SMG7 and TDM1 in meiotic exit in *Arabidopsis*. We demonstrate that meiotic exit is driven by inhibition of translation through TDM1-mediated recruitment of translation initiation factors to P-bodies. Despite their tight association, TDM1 does not have an intrinsic affinity for P-bodies and requires SMG7 for its incorporation. TDM1 is therefore another client of SMG7, in addition to the known interacting partner UPF1, implying a broader role for SMG7 in P-body metabolism and remodeling beyond its role in NMD. TDM1 appears to act as a scaffold protein that, through its specific interaction with eIF4Gs, transiently recruits the main eukaryotic translation initiation factor eIF4F to P-bodies. Meiotic exit in angiosperm plants marks the cell fate transition, which coincides with the remodeling of cell division machinery as well as the alternation between genetically distinct sporophyte and gametophyte life forms. This is expected to be accompanied by the clearance of transcripts specifying the preceding state. Thus, it is likely that recruitment of eIF4F to P-bodies leads not only to translational repression of associated mRNAs but perhaps also to their degradation.

Although P-bodies were identified more than two decades ago, their biological function is still subject to debate. P-bodies associate with RNA decay proteins, but the prevalent view is that they mainly act as sites of translational repression and RNA storage (*26*, *27*). Nevertheless, some studies suggest that they function in selective RNA decay (*28*, *29*). We propose that the interaction between TDM1, embedded within P-bodies, and eIF4F leads to disassembly of the translation initiation complex and exposure of the mRNA 5′ end to decapping proteins. The TDM1 binding site of eIFiso4G2 spans the HEAT1 domain (fig. S8), which is required for the interaction between poly-A binding protein PABP and eIF4G. This interaction tethers the 3′ and 5′ ends of mRNA and facilitates reinitiation of translation (*30*). Disruption of the eIFiso4G2-PABP interaction by TDM1 is expected to impair reinitiation of translation and weaken eIF4F binding to the 5′-mRNA cap (*31*). Together with the high local concentration of decapping factors in P-bodies, this may create a permissive environment for replacement of eIF4F with the decapping complex, priming mRNAs for degradation. This possibility is consistent with the observation that whereas TDM1 is tightly bound to P-bodies, eIFiso4G shows a more dynamic association. In this model, P-bodies with embedded TDM1 may be perceived as catalysts that mediate the disassembly and turnover of RNA-bound eIF4F (fig. S16). It is currently unknown how many transcripts are affected by this mechanism. On one hand, comparative proteomics analysis reported down-regulation of ~25% of detectable proteins between tetrads and microspores, suggesting a pronounced alteration of proteome at the diploid-to-haploid transition (*32*). On the other hand, single-cell transcriptomics in maize showed persistence of sporophytic mRNAs until pollen mitosis I (*32*, *33*), suggesting that P-body–mediated inhibition of translation does not have a global effect on transcriptomes but may exhibit specificity toward a subset of transcripts. This is supported by the observation that only 1.2% of cellular eIFiso4G2 is retained in P-bodies at a given time point. Nevertheless, considering the dynamic nature of this association with 50% signal recovery within ~90 s (Fig. 4A), a considerable portion of eIFiso4G2-associated mRNA can be shuttled through P-bodies during telophase II, leading to remodeling of the translatome and hence facilitating a cell fate transition.

REFERENCES AND NOTES

1. S. Chu *et al*., *Science* **282**, 699–705 (1998).
2. B. Nelms, V. Walbot, *Science* **364**, 52–56 (2019).
3. J. M. D. Legrand, R. M. Hobbs, *Semin. Cell Dev. Biol.* **79**, 80–91 (2018).
4. G. A. Brar *et al*., *Science* **335**, 552–557 (2012).
5. F. Wang *et al*., *Dev. Cell* **52**, 141–151.e5 (2020).
6. L. E. Berchowitz *et al*., *Cell* **163**, 406–418 (2015).
7. N. Riehs *et al*., *J. Cell Sci.* **121**, 2208–2216 (2008).
8. P. Bulankova, N. Riehs-Kearnan, M. K. Nowack, A. Schnittger, K. Riha, *Plant Cell* **22**, 3791–3803 (2010).
9. M. Cifuentes *et al*., *PLOS Genet.* **12**, e1005856 (2016).
10. K. J. Ross *et al*., *Chromosome Res.* **5**, 551–559 (1997).
11. C. Capitao *et al*., *PLOS Genet.* **17**, e1009779 (2021).
12. S. Andreuzza, B. Nishal, A. Singh, I. Siddiqi, *PLOS Genet.* **11**, e1005396 (2015).
13. W. Xing, D. Muhlrad, R. Parker, M. K. Rosen, *eLife* **9**, e56525 (2020).
14. J. Y. Youn *et al*., *Mol. Cell* **76**, 286–294 (2019).
15. N. Fukuhara *et al*., *Mol. Cell* **17**, 537–547 (2005).
16. L. Unterholzner, E. Izaurralde, *Mol. Cell* **16**, 587–596 (2004).
17. Z. Mérai *et al*., *Plant J.* **73**, 50–62 (2013).
18. N. Riehs-Kearnan, J. Gloggnitzer, B. Dekrout, C. Jonak, K. Riha, *Nucleic Acids Res.* **40**, 5615–5624 (2012).
19. C. Capitao, N. Shukla, A. Wandrolova, O. Mittelsten Scheid, K. Riha, *Front. Plant Sci.* **9**, 1602 (2018).
20. G. A. Corbet, R. Parker, *Cold Spring Harb. Symp. Quant. Biol.* **84**, 203–215 (2019).
21. X. Fang *et al*., *Nature* **569**, 265–269 (2019).
22. D. R. Gallie, *J. Biol. Chem.* **291**, 1501–1513 (2016).
23. T. Chantarachot, J. Bailey-Serres, *Plant Physiol.* **176**, 254–269 (2018).
24. H. Y. Cho, M. J. Lu, M. C. Shih, *New Phytol.* **222**, 366–381 (2019).
25. S. Valuchova *et al*., *eLife* **9**, e52546 (2020).
26. A. Hubstenberger *et al*., *Mol. Cell* **68**, 144–157.e5 (2017).
27. N. Standart, D. Weil, *Trends Genet.* **34**, 612–626 (2018).
28. A. Aizer *et al*., *J. Cell Sci.* **127**, 4443–4456 (2014).
29. C. Wang *et al*., *eLife* **7**, e41300 (2018).
30. D. R. Gallie, *Translation* **2**, e959378 (2014).
31. C. C. Wei, M. L. Balasta, J. Ren, D. J. Goss, *Biochemistry* **37**, 1910–1916 (1998).
32. T. Ischebeck *et al*., *Mol. Cell. Proteomics* **13**, 295–310(2014).
33. B. Nelms, V. Walbot, *Science* **375**, 424–429 (2022).

ACKNOWLEDGMENTS

The genome sequencing was performed by the Next Generation Sequencing Facility at Vienna BioCenter Core Facilities (VBCF), a member of the Vienna BioCenter (VBC), Austria. We also acknowledge the Plant Sciences Core Facility of CEITEC MU for support with plant cultivation, as well as the core facility CELLIM of CEITEC MU for assistance with microscopy. **Funding:** This work was supported from the European Regional Development Fund-Project "REMAP" (no. CZ.02.1.01/0.0/0.0/15_003/0000479). CELLIM of CEITEC MU is supported by MEYS CR (LM2018129 Czech-BioImaging). **Author contributions:** A.C. designed and performed experiments and data analysis and wrote the paper. A.V. designed and performed experiments and data analysis. N.S. designed and performed experiments. C.C. performed experiments and data analysis. P.M. performed experiments. S.V. analyzed and formalized data. J.P. performed experiments. P.B. performed experiments. K.R. designed experiments, acquired funding, and wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main paper or the supplementary materials. Reagents, mutants, and reporter lines are available from K.R. upon request. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo0904
Materials and Methods
Figs. S1 to S16
Table S1
References (*34*–*47*)
MDAR Reproducibility Checklist
Movies S1 to S6
Data S1 to S3

Submitted 12 January 2022; accepted 30 June 2022
10.1126/science.abo0904

HYBRIDIZATION

Selection against admixture and gene regulatory divergence in a long-term primate field study

Tauras P. Vilgalys[1,2]†, Arielle S. Fogel[1,3]†, Jordan A. Anderson[1], Raphael S. Mututua[4], J. Kinyua Warutere[4], I. Long'ida Siodi[4], Sang Yoon Kim[1], Tawni N. Voyles[1], Jacqueline A. Robinson[5], Jeffrey D. Wall[5], Elizabeth A. Archie[6], Susan C. Alberts[1,7,8], Jenny Tung[1,7,8,9,10]*

Genetic admixture is central to primate evolution. We combined 50 years of field observations of immigration and group demography with genomic data from ~9 generations of hybrid baboons to investigate the consequences of admixture in the wild. Despite no obvious fitness costs to hybrids, we found signatures of selection against admixture similar to those described for archaic hominins. These patterns were concentrated near genes where ancestry is strongly associated with gene expression. Our analyses also show that introgression is partially predictable across the genome. This study demonstrates the value of integrating genomic and field data for revealing how "genomic signatures of selection" (e.g., reduced introgression in low-recombination regions) manifest in nature; moreover, it underscores the importance of other primates as living models for human evolution.

The ancestors of modern humans intermixed with Neanderthals and other close, now-extinct lineages, leaving a genetic legacy that continues to shape human trait variation today (*1*–*3*). Even as these findings reshape our conception of human origins, they also bring us more closely in line with our primate relatives, where hybridization is observed in many species (*4*, *5*). Studies of other living primates therefore provide context for understanding admixture dynamics in our own lineage. Field studies in hybrid zones, for instance, offer the opportunity to integrate demographic (e.g., reproductive success, immigration/emigration), phenotypic, and genomic data on early-generation hybrids, which studies in humans suggest experienced the greatest fitness costs (*6*, *7*).

Thus far, studies suggest that ancestry frequently predicts trait variation in primate hybrid zones, but admixture often does not result in overt fitness costs (*8*–*11*). However, field observations have not been combined with population and functional genomic analyses to investigate both the organismal and molecular consequences of admixture in primates. Here, we took such an approach to investigate whether selection against introgression (i.e., alleles introduced by gene flow from one distinct lineage to another) is compatible with apparently healthy hybrids, investigated the functional consequences of introgressed alleles, and followed the course of hybridization and natural selection across generations.

We focused on admixture between yellow baboons (*Papio cynocephalus*) and anubis baboons (*P. anubis*): large-bodied, terrestrial primates long used as a model for human biology and evolution (*12*). Although baboon taxonomy has undergone many revisions over time, six extant baboon species are currently recognized on the basis of distinct phenotypic differences and a pattern of phylogenetic divergence supported by recent whole-genome sequencing data (*12*–*14*). This phylogeny establishes two major baboon lineages (the "northern" and "southern" clades) that separated ~1.4 million years ago, although the complex evolutionary history of baboons means that they may have experienced episodes of gene flow since that time (*14*–*16*). Anubis and yellow baboons belong to the northern and southern clades, respectively, and both phylogenetic and population genomic analyses confirm their divergence into distinct taxa (*13*, *14*). Nonetheless, they interbreed to produce viable and fertile offspring where their ranges meet (Fig. 1A).

We concentrated on the region in and around the Amboseli basin of Kenya, where data from 50 years of continuous observation of a population near the center of the hybrid zone are available (*17*). Members of this majority-yellow baboon population include descendants of historical admixture prior to the start of monitoring in 1971, as well as descendants of a directly observed, recent wave of admixture beginning in 1982 (*15*, *18*). In Amboseli, hybrids do not experience obvious fitness costs, and anubis ancestry may in fact confer benefits, including accelerated maturation, increased mating success, and higher rates of male-female affiliation (*19*–*21*). However, field and microsatellite data indicate that the hybrid zone is narrow (*22*), which suggests that natural selection may act to limit gene flow.

Structure of the baboon hybrid zone

To assess selection against introgression in hybrid anubis-yellow baboons, we used whole-genome resequencing data to evaluate ancestry patterns for animals sampled in and near the Amboseli hybrid zone (Fig. 1 and table S1). We generated resequencing data from 430 wild baboons from Amboseli and Mikumi National Park in Tanzania [17 high-coverage (mean = 22.51×); 413 low-coverage (mean = 1.04×)], which we combined with published baboon genomes from Amboseli (n = 22), Mikumi (n = 5), the Maasai Mara National Reserve (n = 7), and the Aberdares region of central Kenya (n = 2) (*14*, *15*, *23*). In Amboseli, our sample included 442 baboons born between 1969 and 2016. Finally, we also included the genomes of 39 captive baboons from the Southwest National Primate Research Center (SNPRC; n = 31 colony founders, 33 total) and the Washington National Primate Research Center (WNPRC; n = 6) (*14*, *15*, *24*).

We estimated global and local ancestry for each individual using a composite likelihood method suitable for low-coverage data, LCLAE, which uses genotype likelihoods across genomic windows rather than requiring genotypes at specific variants (*13*, *15*). These results confirmed that admixture is minimal or absent in the anubis baboon founders of the SNPRC colony, anubis baboons from Maasai Mara, and yellow baboons from Mikumi (Fig. 1 and fig. S5), although we cannot exclude ancient bouts of admixture that affect all living baboons. In contrast, all baboons from Amboseli are admixed [Fig. 1A; mean = 30 to 37% genome-wide anubis ancestry ± 10% SD), including many whose ancestry can be traced to anubis-like immigrants within the most recent seven generations. These results closely match F_4-ratio estimates (*25*) (<2% difference for nine high-coverage Amboseli genomes), indicating that putative anubis ancestry in Amboseli reflects admixture, not incomplete lineage sorting (*13*).

We also detected a signal of ~17% mean anubis ancestry in the putatively yellow baboon founders of the SNPRC colony, which were previously thought to be unadmixed (Fig. 1) (*24*). Identity-by-descent (IBD) analysis using IBDMix (*26*) confirmed this pattern (Fig. 1C). Because IBD between Mikumi yellow baboons and anubis baboons is <5%, these findings also implicate admixture rather than incomplete lineage sorting (Fig. 1C) (*13*). Combined with evidence for yellow ancestry in a central Kenyan anubis baboon (*13*, *14*), our results indicate that gene flow has been a common feature of baboon evolution in east Africa.

[1]Department of Evolutionary Anthropology, Duke University, Durham, NC, USA. [2]Section of Genetic Medicine, University of Chicago, Chicago, IL, USA. [3]University Program in Genetics and Genomics, Duke University, Durham, NC, USA. [4]Amboseli Baboon Research Project, Amboseli National Park, Kenya. [5]Institute for Human Genetics, University of California, San Francisco, CA, USA. [6]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA. [7]Department of Biology, Duke University, Durham, NC, USA. [8]Duke University Population Research Institute, Durham, NC, USA. [9]Canadian Institute for Advanced Research, Toronto, Canada. [10]Department of Primate Behavior and Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
*Corresponding author. Email: jenny.tung@duke.edu
†These authors contributed equally to this work.

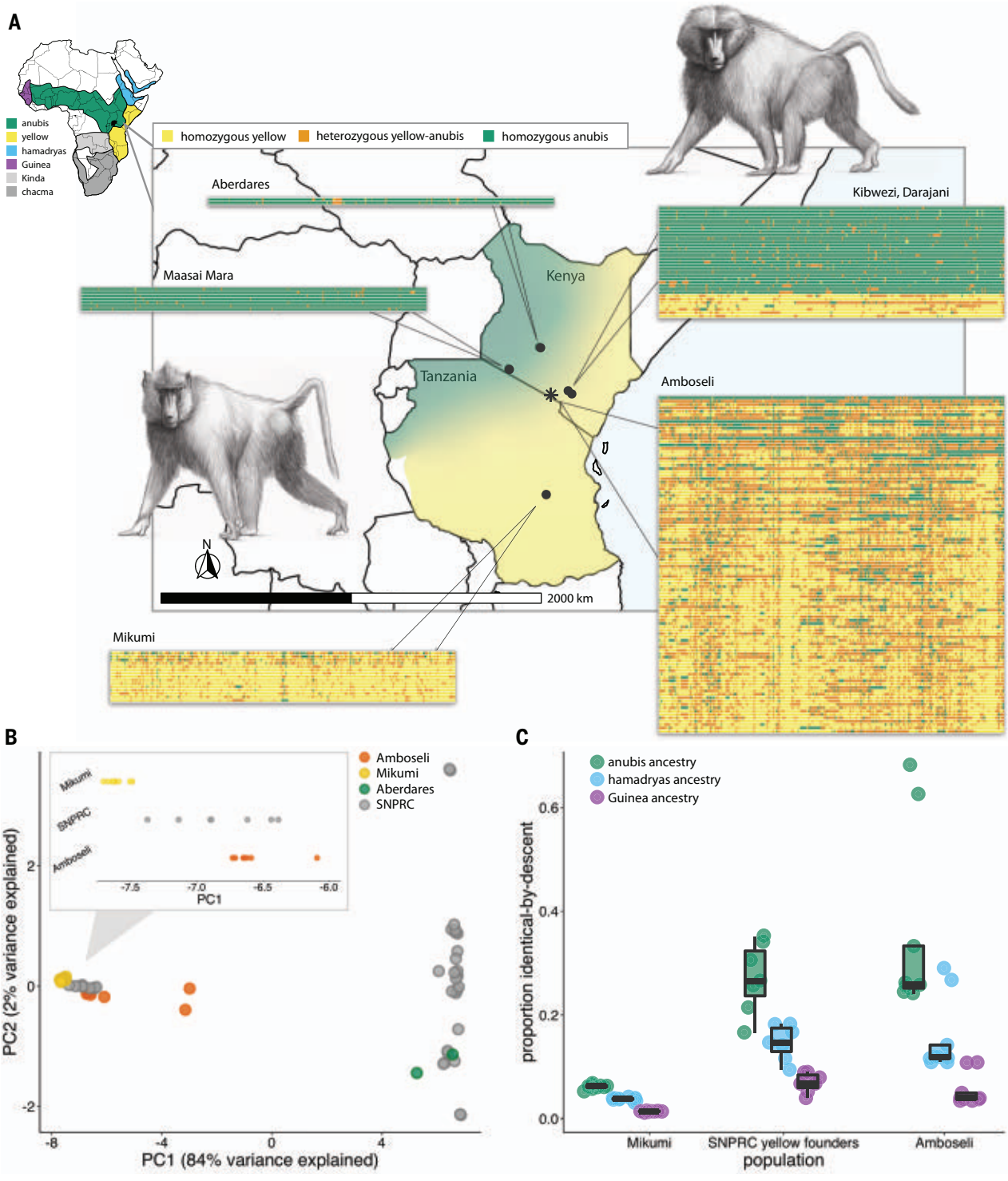


Fig. 1. The structure of the baboon hybrid zone in Amboseli and the surrounding region. (**A**) Geographic locations and local ancestry estimates for baboons in this study (black asterisk = Amboseli). For each population, each row corresponds to the first 20 Mb of chromosome 1 for one individual, organized vertically by global ancestry. For Amboseli, a subsample of 100 individuals is shown. Central map: Ranges of yellow baboons and anubis baboons in Kenya and Tanzania. Small map: Ranges of all six African baboons (*47*), modified from a map by Kenneth Chiou (CC BY 3.0 license). [Baboon drawings by Christopher Smith] (**B**) Principal components analysis (PCA) of genotype data for high-coverage genomes. Inset: Distribution of "yellow-like" individuals along PC1. SNPRC yellow baboon founders resemble admixed Amboseli baboons. (**C**) IBDMix (*26*) results for three sets of yellow or majority-yellow baboons. SNPRC yellow baboon founders and Amboseli baboons exhibit substantial identity by descent (IBD) with anubis baboons, whereas IBD estimates for Mikumi yellow and anubis baboons are low. The excess IBD in the SNPRC and Amboseli samples points to the contribution of gene flow beyond residual incomplete lineage sorting.

Selection against introgression in Amboseli

To investigate whether selection restricts gene flow between anubis and yellow baboons, we focused on the multigenerational dataset from Amboseli. We replicated three analyses used to infer selection against Neanderthal or Denisovan introgression in humans (*27*–*29*). First, we tested for a relationship between anubis ancestry in Amboseli and yellow-anubis genetic divergence [based on unadmixed populations (*13*)]. Because the Amboseli population is largely of yellow baboon origin, if hybridization is deleterious, selection is expected to be less permissive of anubis alleles that are more diverged from their yellow counterparts. Indeed, we found that anubis ancestry is systematically lower in regions of the genome with more fixed differences (Fig. 2A; Spearman's rho = −0.119, $P = 8.05 \times 10^{-34}$). In Amboseli, anubis alleles were 6.7% less common in the most diverged percentile relative to the least diverged percentile of the baboon genome. These results are similar to the negative correlation between the density of fixed human-Neanderthal differences and introgressed Neanderthal ancestry in modern humans (*27*) (Fig. 2B and table S2).

Second, we tested whether introgressed anubis ancestry is depleted in genomic regions that are likely to be affected by linked selection, as summarized by B statistic values calculated for the baboon genome (*13*, *30*) (i.e., due to high gene density per recombination distance). Again paralleling the case of Neanderthal ancestry in modern humans (*28*), anubis ancestry was most common in regions that were predicted to be least affected by linked selection (Fig. 2, C and D; Spearman's rho = 0.168, $P = 1.73 \times 10^{-66}$). Consequently, anubis ancestry per individual was reduced, on average, by 7.03% in protein-coding regions relative to random, size-matched regions of the genome (±4.20% SD; $n = 442$ Amboseli baboons). Reductions in promoters and putative peripheral blood mononuclear cell enhancers were 5.56% (±4.10% SD) and 6.22% (±4.20%), respectively.

Third, we tested whether introgressed anubis ancestry is positively correlated with local recombination rate. This relationship is predicted if recombination influences the rate at which natural selection eliminates deleterious introgressed ancestry and uncouples deleterious from neutral introgressed variants. This prediction, documented across diverse taxa (*29*, *31*, *32*), was also observed in baboons (Fig. 2E; Spearman's rho = 0.127, $P = 1.48 \times 10^{-38}$), with a magnitude similar to that reported for Neanderthal and Denisovan gene flow into modern humans [Fig. 2F; Spearman's rho = 0.17 and 0.14 for Neanderthal and Denisovan ancestry, respectively (*29*)].

To investigate these patterns further, we took advantage of the dynamic history of admixture within the Amboseli population. At the beginning of monitoring in 1971, all Amboseli animals were considered to be yellow baboons (*33*). Phenotypically anubis and admixed animals immigrated into the population starting in 1982, and the proportion of hybrid animals increased over the following decades (*18*, *34*). Whole-genome data recapitulate these patterns, documenting an increase of 11.8% anubis ancestry from 1971 (23.1 to 29.6%) to 2020 (34.9 to 41.4%) (Fig. 3A). However, animals with no known anubis ancestors during the 50-year field study were also clearly admixed (Fig. 3B). Additionally, although immigrant males were more anubis-like than the study population as a whole (Fig. 3A), one immigrant male was among the most yellow-like in our sample (78.8% yellow ancestry), indicating ongoing gene flow involving both parental taxa.

The Amboseli population today therefore contains individuals that descend from ancient, unobserved admixture events as well as those affected by recent hybridization, generating a bimodal distribution of genome-wide ancestry (Fig. 3C) (*15*). By integrating local ancestry calls, pedigree information, and field observations, we identified 188 "recently" admixed individuals whose ancestors include at least one anubis-like immigrant within the last seven generations (mean = 1.7 generations, although these animals are not classical F_1 or F_2 hybrids because historical gene flow is involved). We also classified 214 baboons as "historically" admixed, as their genomes only contain anubis ancestry from before 1971. Forty baboons could not be assigned to either hybrid class (*13*). Based on a single-pulse model of admixture using DATES (*35*), historical admixture is dated to a mean of 283 (±242 SD) generations ago ($n = 7$ high-coverage genomes), in contrast to 5 and 21 generations ago for two recent hybrids sequenced to high coverage.

Stratifying individuals in the dataset by admixture history reveals that signatures of selection against introgression are driven by historical admixture (i.e., genomes sampled dozens to hundreds of generations post-contact). Historically admixed individuals are more depleted of anubis ancestry in highly diverged and low B-value regions of the genome than recently admixed animals (Fig. 3, D and E). Further, the relationship between anubis ancestry and recombination rate is exclusive to the historically admixed dataset, even when recombination rates are measured on chromosome-level scales (Fig. 3F and table S3) (*13*). The weaker signature of selection in recent hybrids likely reflects intermittent gene flow in the most recent generations and stochastic inheritance processes. In contrast, sufficient generations have passed since historical admixture to break apart large introgressed haplotypes, allowing us to observe nonrandom patterns of ancestry across the genome. This result emphasizes the importance of complementing field observations with genomic data, which provide insight into selective processes that operate over time scales longer than even the longest-running field studies.

Selection against regulatory divergence

Analyses of human-Neanderthal admixture suggest a consistent pattern of selection against regulatory variants (*36*). If so, the introgressed regions that persist in modern humans have likely been purged of many alleles with large regulatory effects (*37*, *38*). However, direct comparisons between the effect sizes of retained versus lost archaic alleles are difficult, as only a fraction of archaic hominin alleles (e.g., Neanderthal, Denisovan, or other ghost lineages) segregate in modern human genomes today (*28*, *39*). Extant primate populations, where hybridization and selection are ongoing, provide an opportunity to test this hypothesis.

To test for selection against gene regulatory divergence in baboons, we paired genetic ancestry data with blood-derived RNA-sequencing data from 145 individuals ($n = 157$ samples) (*40*–*42*) (table S1). This dataset includes whole blood and white blood cells, which were analyzed separately while controlling for age, sex, and kinship (*13*). Among 10,192 analyzed genes, we identified no significant associations between genome-wide ancestry and gene expression levels (10% false discovery rate). In contrast, local ancestry predicted gene expression levels for 20.1% (2046) of tested genes in one or both datasets (Fig. 4A), with concordant additive effects between datasets (Pearson's $R = 0.43$, $P < 10^{-200}$) and little evidence for non-additivity (*13*).

If introgressed alleles that perturb gene regulation are a primary target of selection, we reasoned that selection should purge anubis ancestry near genes where ancestry strongly affects gene expression. In support of this prediction, the top 15% of genes with the largest local ancestry effects on gene expression harbored 1.5% less anubis ancestry, on average, than the bottom 15% of genes with the smallest local ancestry effects (Fig. 4B; paired t test, $P = 1.10 \times 10^{-36}$, $n = 442$). This difference was exaggerated within historically admixed individuals (1.9% reduction, $P = 1.26 \times 10^{-27}$, $n = 214$; table S4). Further, the correlation between anubis ancestry and local recombination rate is larger for genes with the largest local ancestry effects than for those with the smallest (Fig. 4C; $\text{rho}_{\text{diff}} = 0.07$ for the top and bottom 15% of genes, bootstrapped $P = 0.027$; table S4). Combined with the depletion of introgressed sequence in regulatory elements, these results support the hypothesis that introgressed alleles that affect gene regulation are nonrandomly purged after hybridization. They are therefore consistent with the idea that natural selection removed archaic variants with large

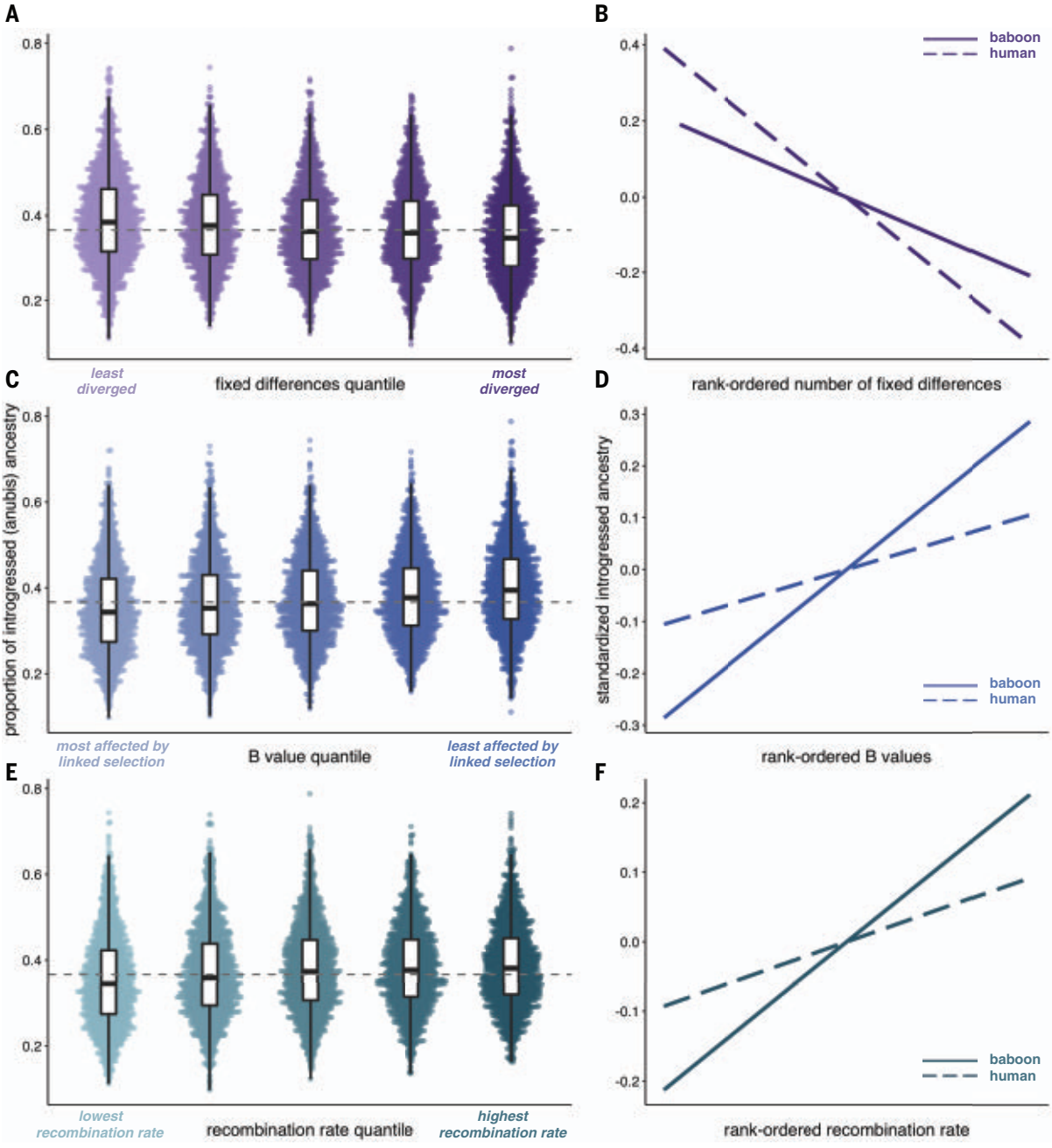


Fig. 2. Selection against introgression in the Amboseli baboons mirrors patterns described for archaic hominin admixture. (**A**, **C**, and **E**) Proportion of introgressed (anubis) ancestry in Amboseli in 250-kb windows (n = 10,324 total windows) as a function of (A) fixed differences between yellow and anubis baboons (Spearman's rho = −0.119, $P = 8.05 \times 10^{-34}$), (C) mean B value (rho = 0.168, $P = 1.73 \times 10^{-66}$), and (E) mean recombination rate (rho = 0.127, $P = 2.49 \times 10^{-38}$), divided into quintiles for visualization purposes only. Dashed gray lines show median anubis ancestry across all windows. (**B**, **D**, and **F**) Predicted relationships between introgressed ancestry and all three measures are observed for both anubis ancestry in the Amboseli baboons (solid lines) and Neanderthal ancestry in modern human genomes (dashed lines) (*27*–*29*), consistent with selection against introgression. Panels show the relationship between introgressed ancestry and the rank-ordered (B) number of fixed differences, (D) mean B statistic, and (F) mean local recombination rate. Mean introgressed ancestry is centered on 0 and divided by the standard deviation for each species to facilitate comparison.

regulatory effects from the genomes of modern humans (*37*).

Predicting the genomic landscape of introgression

Finally, we investigated our ability to predict the genomic locations most and least affected by introgression. We modeled mean anubis ancestry as a function of local recombination rate, single-nucleotide polymorphism density in the reference yellow and anubis populations, yellow-anubis genetic divergence, gene and enhancer content, linked selection, and local ancestry–associated gene expression in blood. We iteratively trained an elastic net regression model on nonoverlapping 250-kb windows of the genome, representing 75% of the genome, and applied the model to a test set of windows in the remaining 25% (*13*). We found that our

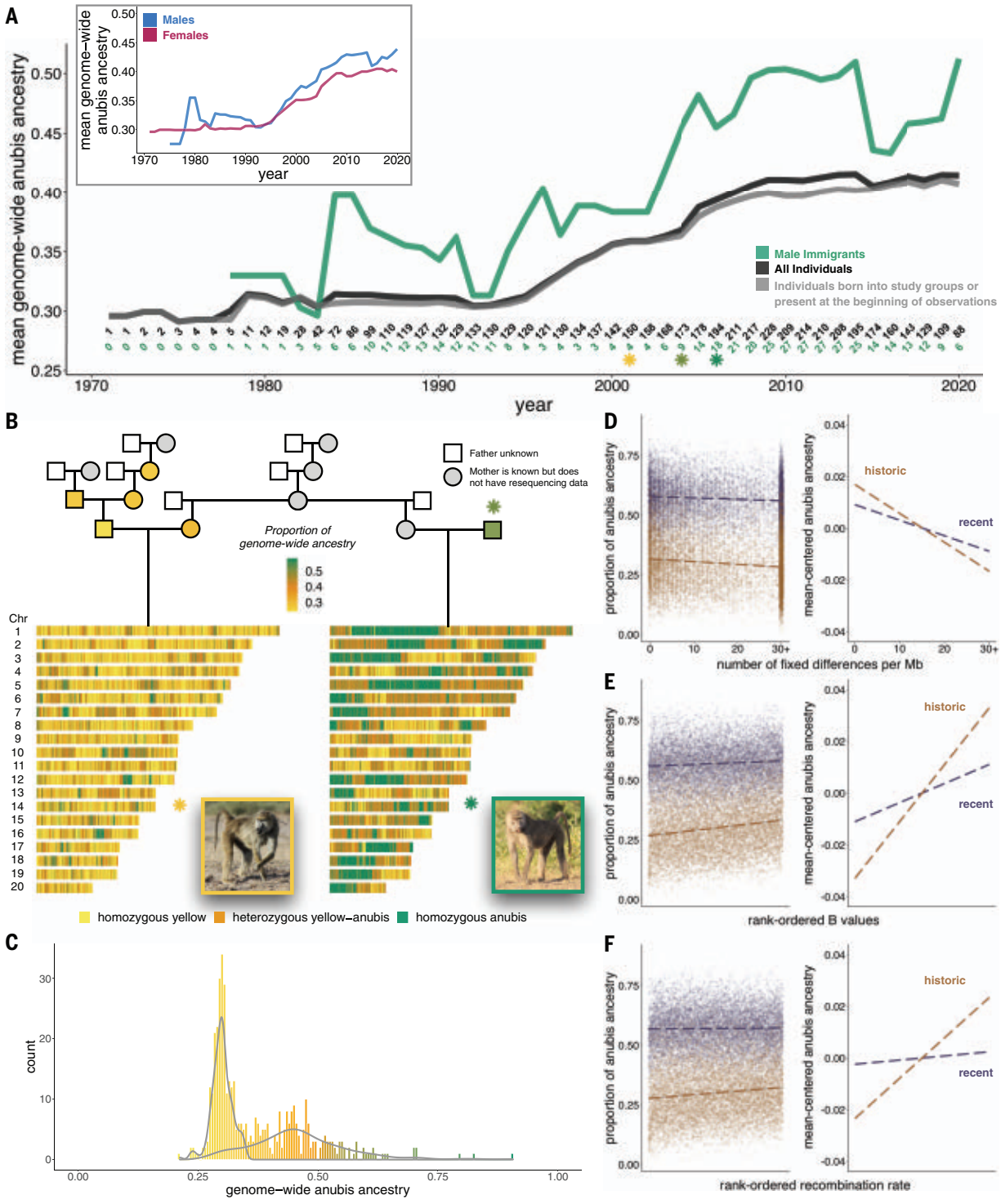


Fig. 3. Recent and historical hybrid ancestry in Amboseli. (**A**) Mean genome-wide anubis ancestry in the Amboseli population has increased since the 1970s. Numbers above the *x* axis indicate the number of individuals used to calculate annual ancestry (black, all individuals; green, male immigrants). (**B**) Pedigree and ancestry estimates for example historical (left) and recent (right) hybrids. Pedigree individuals with resequencing data are colored by ancestry. The two examples share a maternal grandmother and were born a few years apart [yellow and bright green asterisks in (A)]. The father of the recent hybrid immigrated in 2004 [olive green asterisk in (A)]. (**C**) Genome-wide anubis ancestry in Amboseli, with density plots overlaid for historical and recently admixed individuals. (**D** to **F**) The relationships between introgressed anubis ancestry and the rank-ordered (D) number of fixed anubis-yellow differences, (E) mean B value, and (F) mean local recombination rate. All relationships are stronger for historical hybrids than for recent hybrids. Right panels show anubis ancestry within each dataset mean-centered to 0.

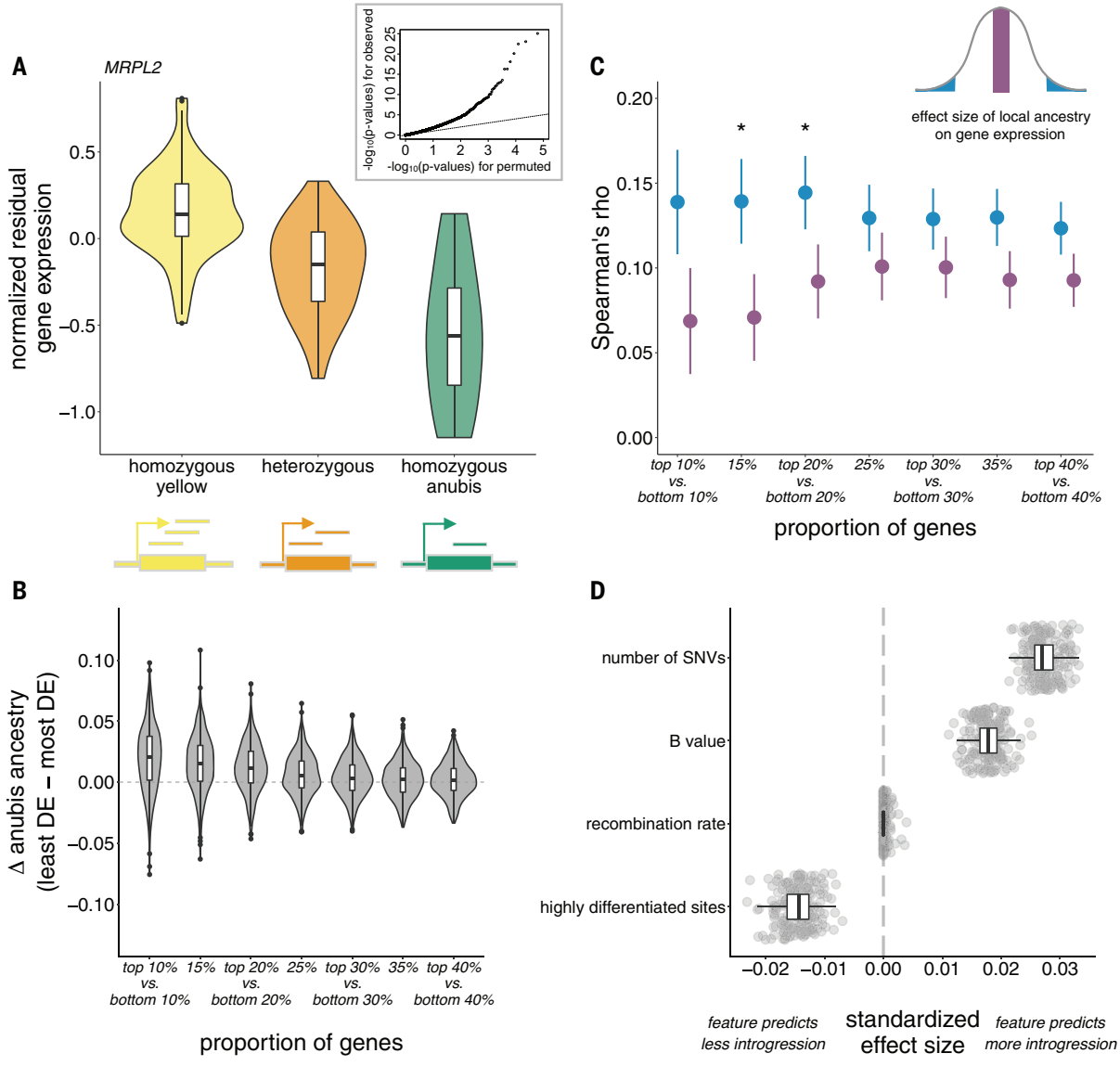


Fig. 4. Selection against gene regulatory divergence and prediction of local introgression levels. (**A**) Local ancestry predicts gene expression in the Amboseli population, as depicted for an example gene (*MRPL2*). Inset: quantile-quantile plot comparing *P* distributions for observed local ancestry effects (*y* axis) to a permutation-based null (*x* axis). (**B**) Difference in introgressed anubis ancestry between genes with the smallest versus largest local ancestry effects on gene expression. Violin plots show the distribution of differences across individuals; boxplots show the median and interquartile range (all $P < 0.05$; table S4). (**C**) Correlation between anubis introgression and recombination rate calculated for sets of genes with the largest (blue) versus smallest (magenta) local ancestry absolute effect sizes. Asterisks denote bootstrapped $P < 0.05$ (table S4); error bars show SD. (**D**) Distribution of effect sizes for features that consistently predict the extent of anubis introgression in Amboseli baboon genomes (table S5). The number of single-nucleotide variants (SNVs) is derived from unadmixed yellow and anubis baboon genomes.

predicted values were consistently positively correlated with observed levels of anubis ancestry in the test sets (mean Pearson's R = 0.254 ± 0.016 SD versus 0.014 ± 0.011 SD for models fit to permuted data), with frequent contributions from features capturing local recombination rate, linked selection, genetic variation, and sequence divergence (Fig. 4D and table S5). We consistently predicted anubis ancestry more accurately in historical hybrids than in recent hybrids (mean Pearson's R = 0.265 ± 0.017 SD versus 0.177 ± 0.018 SD, bootstrapped $P < 10^{-3}$).

Our longitudinal data also indicate that increases in anubis ancestry across the 50-year field study are nonrandomly distributed throughout the genome. Controlling for the starting level of anubis ancestry in 1979, 100-kb windows characterized by lower F_{ST} and higher recombination rates experienced larger increases in anubis ancestry between 1979 and 2020, although both effect sizes were small (F_{ST} and recombination rate P values < 10^{-3}, β = −2.965 × 10^{-4} and 1.020 × 10^{-4}, respectively; n = 25,797 windows; table S6). B statistic values did not predict temporal change in anubis ancestry independently of recombination rate.

Divergence and hybridization in primates

Our genomic analysis reveals evidence for selection against admixture that is remarkably consistent with results obtained for archaic introgression in humans. Our results also support

a hypothesis that can only be indirectly tested in our own lineage: that natural selection has acted to eliminate introgressed alleles that strongly perturb gene regulation (*37*). These results contrast with the behavioral and life history evidence to date in Amboseli—one of the largest and longest-running primate field sites in the world—which indicates that hybrid baboons suffer no obvious fitness costs (*19*–*21*). Our results identify subtle selection against hybridization that may help to explain the maintenance of primate taxonomic integrity in the face of frequent interspecific gene flow (*4*, *5*). Ultimately, the outcome of this process will depend on the relative balance among this selection pressure, possible advantages to introgressed ancestry, migration rates, and demographic stochasticity—potentially explaining cases of nuclear swamping in baboons despite costs to hybridization (*16*).

The mode of selection against hybrids is unclear. Unlike in humans, hybridization load is unlikely to explain our results: Yellow and anubis baboons harbor similar levels of genetic diversity compared to humans and Neanderthals, which differ more [<50% difference in baboons versus a >3-fold difference between humans and Neanderthals (*6*, *14*, *15*, *27*)]. Both hybrid incompatibilities and ecological selection, however, could play a role. For example, some reports suggest that anubis and yellow baboons occupy distinct climatic niches (*43*). Previously described assortative mating by ancestry in the Amboseli baboons (*20*) may also limit introgression. Understanding the genetic and phenotypic mechanisms that influence interspecific gene flow, including the role of the X chromosome and adaptive introgression, remains an important goal for future work.

Combined, our findings illustrate the importance of contextualizing genomic data with phenotypic and demographic information to understand the evolutionary dynamics of admixture. Genomes harbor information about historical processes that stretch back many generations, and can capture subtle signatures of selection that may not be obvious in natural populations where demographic stochasticity is high, sample sizes are modest, and the specific phenotypes under selection may be unknown. Conversely, field data reveal the range of phenotypic and fitness outcomes that are compatible with genomic signatures of selection. Indeed, genomic evidence alone has led some researchers to posit that the costs of modern human-archaic hominin interbreeding must have been high, reflecting species at the brink of reproductive incompatibility (*44*, *45*). Our results point to the limits of these inferences by indicating that qualitatively similar evidence for selection against introgression can be compatible with primate hybrids that thrive (*19*–*21*). This work therefore highlights the crucial role of other primates for understanding human evolution, especially for phenomena that are impossible to study in our lineage alone.

REFERENCES AND NOTES

1. E. Huerta-Sánchez *et al*., *Nature* **512**, 194–197 (2014).
2. C. N. Simonti *et al*., *Science* **351**, 737–741 (2016).
3. H. Zeberg, S. Pääbo, *Nature* **587**, 610–612 (2020).
4. D. Zinner, M. L. Arnold, C. Roos, *Evol. Anthropol.* **20**, 96–103 (2011).
5. J. Tung, L. B. Barreiro, *Curr. Opin. Genet. Dev.* **47**, 61–68 (2017).
6. K. Harris, R. Nielsen, *Genetics* **203**, 881–891 (2016).
7. M. Petr, S. Pääbo, J. Kelso, B. Vernot, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1639–1644 (2019).
8. C. J. Jolly, A. S. Burrell, J. E. Phillips-Conroy, C. Bergey, J. Rogers, *Am. J. Primatol.* **73**, 291–303 (2011).
9. T. J. Bergman, J. E. Phillips-Conroy, C. J. Jolly, *Am. J. Primatol.* **70**, 136–147 (2008).
10. K. M. Detwiler, *Int. J. Primatol.* **40**, 28–52 (2018).
11. D. M. Kitchen *et al*., *Int. J. Primatol.* **40**, 132–152 (2017).
12. J. Fischer *et al*., *eLife* **8**, e50989 (2019).
13. See supplementary materials.
14. J. Rogers *et al*., *Sci. Adv.* **5**, eaau6947 (2019).
15. J. D. Wall *et al*., *Mol. Ecol.* **25**, 3469–3483 (2016).
16. D. Zinner, L. F. Groeneveld, C. Keller, C. Roos, *BMC Evol. Biol.* **9**, 83 (2009).
17. S. C. Alberts, J. Altmann, in *Long-Term Field Studies of Primates*, P. M. Kappeler, D. P. Watts, Eds. (Springer, 2012), pp. 261–287.
18. A. Samuels, J. Altmann, *Int. J. Primatol.* **7**, 131–138 (1986).
19. M. J. Charpentier, J. Tung, J. Altmann, S. C. Alberts, *Mol. Ecol.* **17**, 2026–2040 (2008).
20. J. Tung, M. J. Charpentier, S. Mukherjee, J. Altmann, S. C. Alberts, *Am. Nat.* **180**, 113–129 (2012).
21. A. S. Fogel *et al*., *Anim. Behav.* **180**, 249–268 (2021).
22. M. J. Charpentier *et al*., *Mol. Ecol.* **21**, 715–731 (2012).
23. N. Snyder-Mackler *et al*., *Genetics* **203**, 699–714 (2016).
24. J. A. Robinson *et al*., *Genome Res.* **29**, 848–856 (2019).
25. N. Patterson *et al*., *Genetics* **192**, 1065–1093 (2012).
26. L. Chen, A. B. Wolf, W. Fu, L. Li, J. M. Akey, *Cell* **180**, 677–687. e16 (2020).
27. B. Vernot, J. M. Akey, *Science* **343**, 1017–1021 (2014).
28. S. Sankararaman *et al*., *Nature* **507**, 354–357 (2014).
29. M. Schumer *et al*., *Science* **360**, 656–660 (2018).
30. G. McVicker, D. Gordon, C. Davis, P. Green, *PLOS Genet.* **5**, e1000471 (2009).
31. N. B. Edelman *et al*., *Science* **366**, 594–599 (2019).
32. E. Calfee *et al*., *PLOS Genet.* **17**, e1009810 (2021).
33. S. A. Altmann, J. Altmann, *Baboon Ecology: African Field Research* (Univ. of Chicago Press, 1970).
34. J. Tung, M. J. Charpentier, D. A. Garfield, J. Altmann, S. C. Alberts, *Mol. Ecol.* **17**, 1998–2011 (2008).
35. V. M. Narasimhan *et al*., *Science* **365**, eaat7487 (2019).
36. N. Telis, R. Aguilar, K. Harris, *Nat. Ecol. Evol.* **4**, 1558–1566 (2020).
37. R. C. McCoy, J. Wakefield, J. M. Akey, *Cell* **168**, 916–927.e12 (2017).
38. L. L. Colbran *et al*., *Nat. Ecol. Evol.* **3**, 1598–1606 (2019).
39. M. Steinrücken, J. P. Spence, J. A. Kamm, E. Wieczorek, Y. S. Song, *Mol. Ecol.* **27**, 3873–3888 (2018).
40. J. Tung, X. Zhou, S. C. Alberts, M. Stephens, Y. Gilad, *eLife* **4**, e04729 (2015).
41. A. J. Lea *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E12163–E12171 (2018).
42. J. A. Anderson *et al*., *Philos. Trans. R. Soc. London Ser. B* **377**, 20200441 (2022).
43. D. Chala, C. Roos, J. C. Svenning, D. Zinner, *J. Hum. Evol.* **132**, 215–226 (2019).
44. M. Currat, L. Excoffier, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15129–15134 (2011).
45. D. Reich, *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past* (Oxford Univ. Press, 2018).
46. T. P. Vilgalys *et al*., Data and code for: Selection against admixture and gene regulatory divergence in a long-term primate field study. Zenodo (2021): https://doi.org/10.5281/zenodo.5912279.
47. D. Zinner, J. Wertheimer, R. Liedigk, L. F. Groeneveld, C. Roos, *Am. J. Phys. Anthropol.* **150**, 133–140 (2013).

ACKNOWLEDGMENTS

We thank J. Altmann for her fundamental contributions to research on the Amboseli baboons. We thank the University of Nairobi, Institute of Primate Research, the National Museums of Kenya, the Amboseli-Longido pastoralist communities, the Enduimet Wildlife Management Area, Ker & Downey Safaris, Air Kenya, and Safarilink for assistance in Kenya; S. Sayialel, G. Marinka, B. Oyath, T. Wango, and V. Oudu as long-term members of the Amboseli Baboon Research Project (ABRP) team; and K. Pinc, N. H. Learn, and J. B. Gordon for contributions to the ABRP database. The Kenya-based research reported here was conducted with permission from the Kenya Wildlife Service (KWS), the National Environmental Management Authority (NEMA), and the National Commission for Science, Technology and Innovation (NACOSTI), which we currently renew on an annual basis. Samples were collected and exported under NEMA permits and CITES permits from KWS. We also hold a Memorandum of Agreement (MoA) with KWS, the University of Nairobi, the Institute of Primate Research, and the National Museums of Kenya that details the ABRP research mission and commitment to benefits sharing under the Nagoya Protocol. Finally, we thank G. McVicker, P. Moorjani, K. Samuk, S. He, and members of the Tung lab for discussion and assistance with code, and L. Barreiro, A. Goldberg, P. Moorjani, and G. Coop and members of the Coop lab for feedback on a previous version of this manuscript. The research in this study was approved by the Institutional Animal Care and Use Committees at Duke University, Princeton University, and the University of Notre Dame and adhered to the laws and guidelines of the Kenyan government. **Funding:** Supported by NSF IOS 1456832 (S.C.A.), NIH R01AG053308 (S.C.A.), NIH R01AG053330 (E.A.A.), R01HD088558 (J.T.), P01AG031719 (S.C.A.), NSF BCS-1751783 (J.T., T.P.V.), NSF BCS-2018897 (J.T., A.S.F.), a Leakey Foundation Research Grant (T.P.V.), and the North Carolina Biotechnology Center (2016-IDG-1013). A.S.F. was supported by NSF GRFP (DGE 1644868) and NIH T32GM007754. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of our funding bodies. **Author contributions:** T.P.V., A.S.F., and J.T. conceived and designed the study. A.S.F., J.A.A., R.S.M., J.K.W., I.L.S., S.Y.K., T.N.V., E.A.A., S.C.A., and J.T. collected data. J.D.W. contributed data analysis tools. T.P.V., A.S.F., J.A.A., and J.A.R. processed data. T.P.V. and A.S.F. performed analyses. T.P.V., A.S.F., and J.T. interpreted analyses. T.P.V., A.S.F., and J.T. wrote the manuscript with edits and revisions from all other coauthors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All sequencing data have been deposited in the NCBI Sequence Read Archive under BioProject accession number PRJNA755322. Previously generated resequencing data are available under PRJNA308870, PRJNA433868, PRJNA54005, PRJNA20425, PRJNA162517, PRJNA54003, PRJNA54009, PRJNA54007, PRJNA251424, and PRJNA295782, and previously generated RNA-seq data are deposited under PRJNA269070, PRJNA480672, and PRJNA731520. Processed data and scripts are available on Zenodo (*46*). All data and code required to reproduce or reanalyze these data are public. The biological samples are CITES-regulated and can only be shared with third parties with prior written authorization from KWS. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm4917
Materials and Methods
Figs. S1 to S8
Tables S1 to S9
References (*48*–*136*)
MDAR Reproducibility Checklist

Submitted 24 September 2021; accepted 1 July 2022
10.1126/science.abm4917

SYSTEMS BIOLOGY

Multimodal perception links cellular state to decision-making in single cells

Bernhard A. Kramer[1,2], Jacobo Sarabia del Castillo[1], Lucas Pelkmans[1]*

Individual cells make decisions that are adapted to their internal state and surroundings, but how cells can reliably do this remains unclear. To study the information processing capacity of human cells, we conducted multiplexed quantification of signaling responses and markers of the cellular state. Signaling nodes in a network displayed adaptive information processing, which led to heterogeneous growth factor responses and enabled nodes to capture partially nonredundant information about the cellular state. Collectively, as a multimodal percept this gives individual cells a large information processing capacity to accurately place growth factor concentration within the context of their cellular state and make cellular state–dependent decisions. Heterogeneity and complexity in signaling networks may have coevolved to enable specific and context-aware cellular decision-making in a multicellular setting.

Contextual decision-making by cells in a collective is a hallmark of multicellular systems (*1*–*4*). To achieve context-aware behavior, individual cells must integrate the input they receive from growth factors with complex information on their physicochemical state. Cells perceive this information through activation of intracellular signaling networks; however, individual signaling nodes in such networks are thought to have low capacity for processing information as a result of their highly variable growth factor responses in single cells (*5*–*7*). It thus remains largely unknown if and how individual cells can process a large amount of information in a contextual manner.

We explored the possibility that variable growth factor responses do not reflect a limited information processing capacity but instead represent adaptive information processing. In adaptive information processing the response of a signaling node in an individual cell is adapted to the physicochemical state of the cell and its surroundings (here collectively referred to as the cellular state), through mechanisms by which the cellular state controls the activities of signaling nodes (*8*, *9*). This implies that signaling responses not only capture information about the amount of growth factor a cell is exposed to but also—and perhaps primarily—obtain information about its cellular state. If the activation of different signaling nodes in a network is dependent on different properties of the cellular state then each node would carry partially nonredundant information. As a whole, the network could then generate a multimodal percept that captures a comprehensive picture of a cell's multicellular context and internal state, facilitating accurate and contextual decision-making.

[1]Department of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
[2]Molecular Life Sciences PhD program, Life Science Zurich Graduate School, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.
*Corresponding author. Email: lucas.pelkmans@mls.uzh.ch

Results

Multiplexed quantitative imaging of signaling and cellular state

To obtain multiple readouts of the cellular state in addition to signaling responses we applied 4i—a high-resolution multiplexing technology on the basis of iterative staining and elution of antibodies (*10*)—to human epithelial cells (184A1). After 4 days of growth, cells were deprived of serum and growth factors for 12 hours and subsequently exposed to five different concentrations of epidermal growth factor (EGF) for 5 min (Fig. 1A). After 30-plex 4i, cell segmentation, and quality control, the dataset contained ~8000 individual cells per replicate ($n = 3$) and condition (Fig. 1B and fig. S1, A to C), showing high technical and biological reproducibility (fig. S1, D to F). Images of cells reveal the highly heterogeneous nature of acute signaling responses as well as that of the cellular state (Fig. 1C). Quantifying the abundances of three signaling responses and three cellular state markers in every cell and comparing single-cell distributions of the five conditions revealed that although signaling responses typically display changing levels with increasing amounts of EGF in either a gradual or switch-like bimodal manner, their responses were highly heterogeneous between cells. This results in overlapping single-cell distributions between different doses of EGF (Fig. 1D and fig. S2, A and B). By contrast, cellular state markers did not show any change (Fig. 1D). A systematic assessment of all single-cell features quantified from the images revealed that ten signaling responses downstream of EGF displayed significant changes whereas 650 features of the cellular state did not change during a 5 min exposure to EGF (Fig. 1E). These features quantify or act as proxies of properties across multiple spatial scales such as relative cell positioning within a population, position in the cell cycle, and subcellular textures of organelles (Fig. 1F and fig. S2, C and D). Thus although the cellular state changes in response to EGF stimulation (*11*), this occurs at longer timescales. Inhibitors of signaling nodes abrogated signaling responses to EGF but did not change markers of the cellular state (fig. S2E). We can thus test whether the heterogeneity seen in acute EGF-induced signaling responses is linked to the preexisting heterogeneity in cellular states.

The preexisting cellular state landscape shapes signaling responses in single cells

A projection of the multidimensional cellular state space into a two-dimensional (2D) landscape (*12*) was largely continuous except for two parts that reflect G1 and G2 of the cell cycle (Fig. 1F and fig. S2, F and G). Signaling responses of single cells distribute in different patterns across this landscape [Fig. 2A (middle) and fig. S4A], which was accurately predicted by features of the cellular state (Fig. 2A and fig. S3). This required multiple features in different combinations for different signaling nodes (fig. S4, B to D). For instance, although local cell density was an important feature for most signaling responses, the abundance of Paxillin—a proxy for cell spreading—was particularly important for predicting levels of pS6 in single cells, the abundance of Sec13 was particularly important for predicting pERK, and the amount of nuclear Yap1 was important for predicting nuclear translocation of FoxO3a [Fig. 2A (middle) and fig. S4B]. Thus the cellular state has a specific multivariate effect on signaling responses in individual cells. It also predicts switch-like bimodal response properties (being either a low or high responder) with high accuracy, as shown by some signaling nodes at low doses of EGF (Fig. 2B, Fig. 1D, fig. S2B, and fig. S4E), and achieves high prediction accuracy across the full range of EGF doses tested (Fig. 2C, fig. S4F, and fig. S5). As a result, variation in signaling responses induced by different cellular states is typically larger than variation induced by different concentrations of EGF (Fig. 2, D and E).

Multimodal perception accurately decodes EGF concentration

The above suggests that the sensitivity of signaling nodes—quantified by the effective concentration of EGF at which a node is half-maximally activated (EC_{50})—is adapted to the cellular state. To study this, we defined 18 cellular state classes and compared their dose-response curves (Fig. 3A and fig. S6, A to E). This showed that, for instance, the EC_{50} of pERK differed by a factor of at least three between

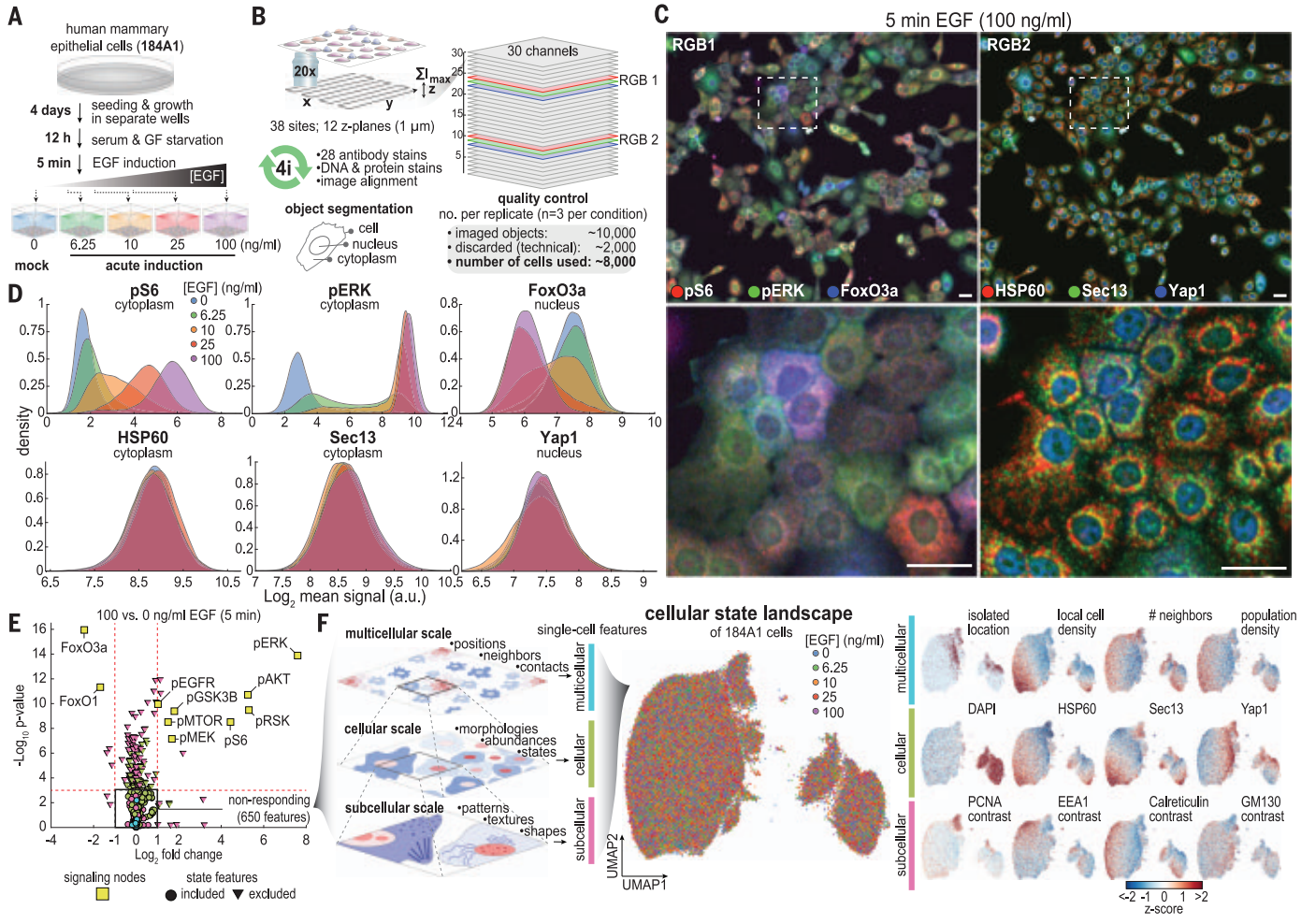


Fig. 1. Acute signaling responses and preexisting cellular states across spatial scales. (**A**) Experimental workflow. (**B**) 4i, imaging setup and quality control pipeline. (**C**) Two combinations of three channels of one imaging site (top) and zoom-in of cells (bottom). Scale bars: 7.5 µm. (**D**) Density distributions of mean intensities in either the cytoplasm or nucleus of six markers. (**E**) Systematic identification of features that change or do not change upon EGF treatment. *P* value was estimated with a two-tailed *t* test comparing replicates treated with 100 ng/ml to 0 ng/ml EGF. (**F**) Single-cell nonresponding features (143 principal components) define the cellular state space, projected using UMAP embedding. (Left) Schematic of features derived from multiple spatial scales. (Middle) Position of cells from each condition. (Right) Values from selected cellular state features.

cellular states (Fig. 3B), with class 2 cells (small cells grown in densely populated areas and having abundant early endosomes) showing a particularly high EC_{50} and class 15 cells (grown in sparsely populated regions and having nuclear Yap1) a low EC_{50}. Inferring single-cell EC_{50}s revealed that the sensitivity of each signaling node was adapted to the cellular state in distinct ways and for different ranges of EGF concentrations (Fig. 3C and fig. S6F).

To test whether signaling node–specific, cellular state–conditioned sensitivity to EGF could provide enough information to individual cells to perceive a wide range of EGF concentrations in a contextual manner, we quantified the amount of mutual information between the signaling responses of individual cells and the five doses of EGF they are exposed to (*13*). Perfect decoding ability would show mutual information of log2(5) or 2.3 bits. When cells were considered to use information from one node (unimodal perception) and it was assumed that their signaling was not conditioned to the cellular state (noncontextual), cells could only distinguish large differences between individual doses (Fig. 3D) and had little decoding capacity (0.7 bits) (Fig. 3E and fig. S7, A to C). When we considered that the cellular state affects the signaling response (contextual), smaller differences between individual doses could be distinguished and the decoding capacity of individual cells increased (1.2 bits). When cells were considered to use information from multiple nodes (multimodal perception) (*14*) they could approach perfect decoding (2.1 bits), but only when conditioning by the cellular state was considered (Fig. 3, D to E). Inhibitor experiments showed that a multimodal response was required to achieve the greatest accuracy. Inhibiting either MEK or AKT reduced the perception accuracy of EGF concentration and inhibiting both led to a stronger reduction (Fig. 3F and fig. S7, D to F). This indicates that multimodal perception could enable individual cells to accurately distinguish a range of EGF concentrations in a manner that is conditioned to the cellular state.

Multimodal perception comprehensively maps cellular state space

Conditioning by the cellular state implies that the multimodal response of an individual

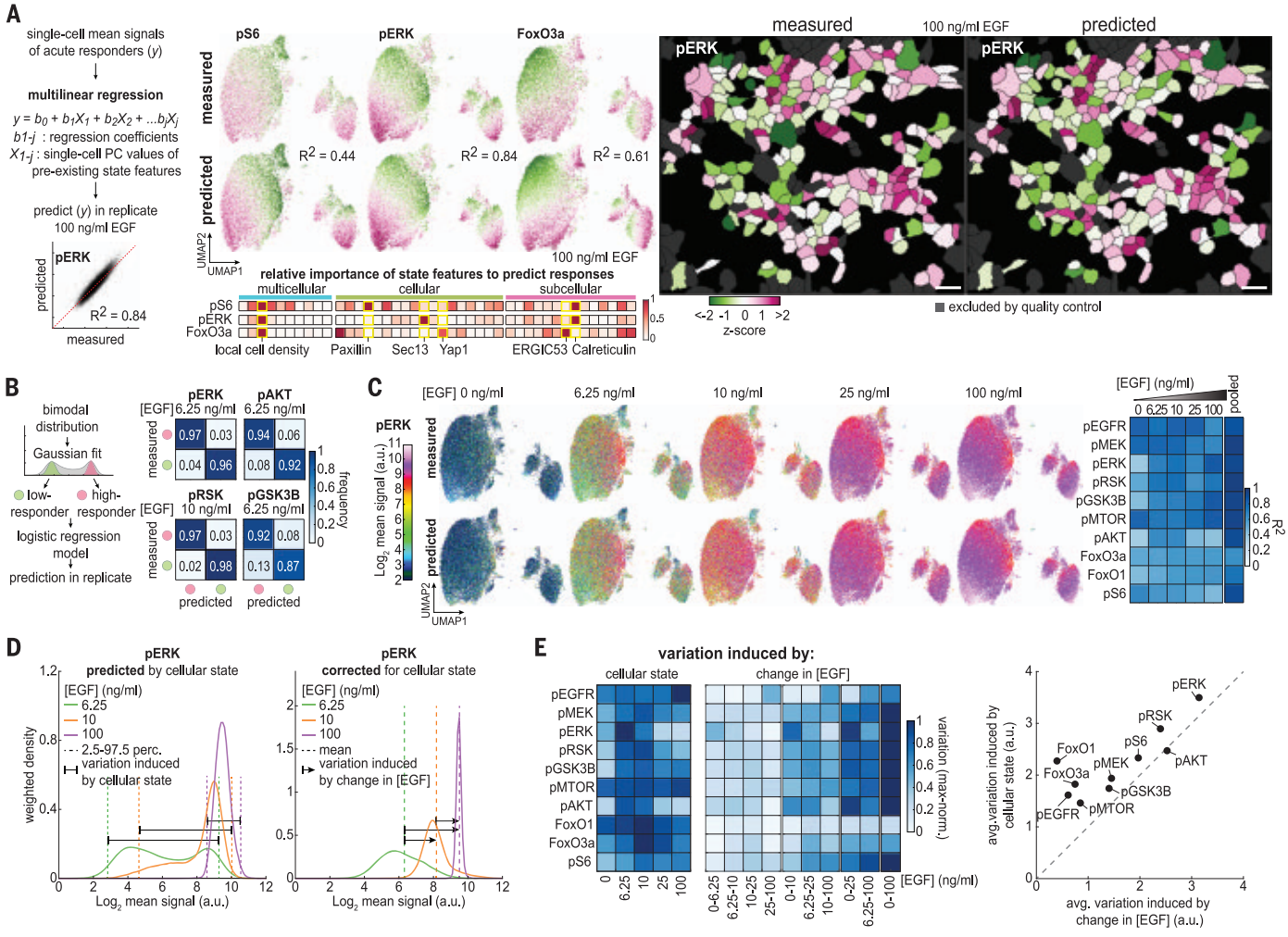


Fig. 2. Cellular state determines the heterogeneous signaling response. (**A**) (Left) Regression approach. (Upper middle) Measured and predicted levels of three signaling response markers. R^2, explained variance. (Lower middle) Dominance analysis of cellular state features in explaining signaling responses. Yellow boxes indicate specific examples. (Right) Side-by-side projection of measured and predicted levels of pERK. Scale bars: 15 μm. (**B**) (Left) Logistic regression. (Right) Confusion matrices for class predictions of bimodal responses. (**C**) (Left) Measured and predicted levels of pERK across all concentrations of EGF. (Right) R^2 for all signaling responses across all EGF concentrations. (**D**) Variation in signaling responses induced by the preexisting cellular state- and variation-induced EGF treatment. (Left) pERK at three EGF concentrations predicted by cellular state. (Right) pERK at three EGF concentrations corrected for variation in cellular state. (**E**) (Left) Quantification of cellular state- and EGF-induced variation as described in (D). (Right) Mean variation induced by all concentration changes in EGF (*x* axis) versus the mean variation induced by cellular state at all concentrations of EGF (*y* axis). Dashed line, $x = y$.

cell to EGF carries a considerable amount of information about the cellular state, especially if nodes in a network not only contain redundant but also unique information and can act in a synergistic manner (Fig. 4A). For instance, the cytoplasmic abundances of pERK and pMTOR in cells show a strong positive correlation, suggesting redundancy, which is observed to various degrees for all nodes in the network (Fig. 4A and fig. S8A). However, the amounts of pERK and pMTOR also scale with cellular state properties that are distinct when compared with the other as illustrated for Sec13 and HSP60 (Fig. 4A). This can be revealed by partial correlation analysis, indicating that the amounts of pERK scale with the abundance of Sec13 whereas the amounts of pMTOR scale with the abundance of HSP60, and can be appreciated in images of individual cells that differ in the abundance of one but not the other cellular state marker (Fig. 4A). To estimate the amount of information pERK and pMTOR carry about these cellular state properties and to decompose it into redundant (captured by either node), unique (captured by one node), and synergistic components (captured only by the combined response of both nodes), we employed partial information decomposition (PID) (*15*). pERK and pMTOR capture redundant but also unique information about Sec13 and HSP60 abundance, respectively (Fig. 4A, right bar graphs). For the latter, pERK and pMTOR also capture synergistic information.

We next applied this analysis to all tested nodes (Fig. 4, B and C, and fig. S8B), showing that each node displays some distinct scaling with various cellular state properties (Fig. 4B). Many of these effects have previously been reported such as (negative) pMEK scaling with local cell density (*16*), pAKT scaling with cell area (*8*), pRSK scaling with the

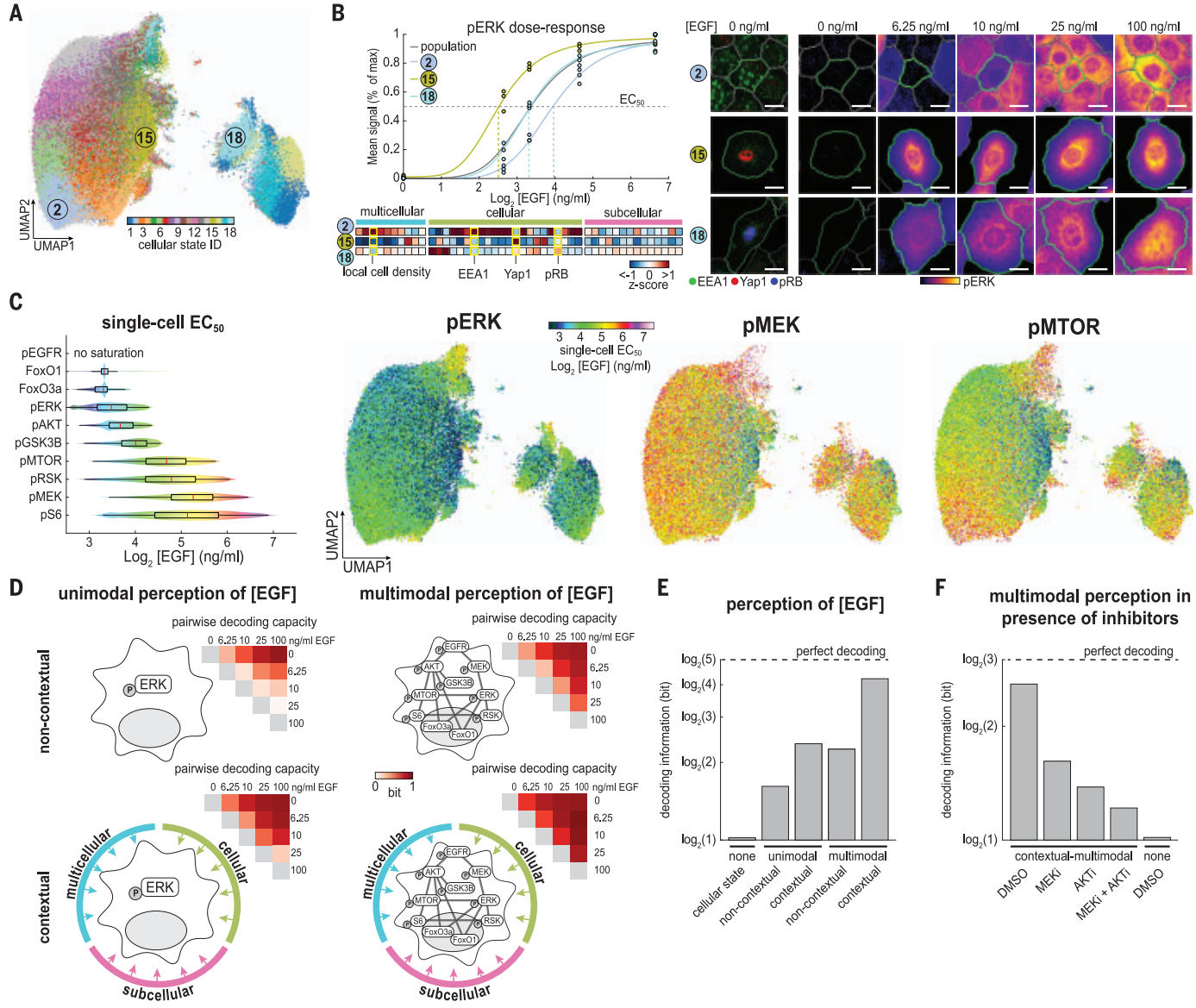


Fig. 3. Contextual multimodal perception of EGF concentration. (**A**) Cellular state classes (highest membership coefficient) from fuzzy clustering on the cellular state landscape. (**B**) (Left) dose-response curves of pERK for the whole population (gray) and for three cellular state classes. Dots, individual replicates. (Bottom) Values of cellular state features. Yellow boxes indicate examples. (Right) Cellular state markers and (left) pERK in representative cells of three state classes outlined in green. Scale bars: 2.5 μm. (**C**) Single-cell EC_{50} was calculated by combining measured responses with predicted (regression). (Left) Single-cell EC_{50} values for each signaling node. (Right) EC_{50} values on cellular state landscape for three signaling nodes. (**D**) Capacity to distinguish two EGF concentrations based on different perception modes. Decoding capacity is calculated from predicted and true EGF exposure labels. (**E**) Decoding capacity of EGF concentrations for different perception modes for all concentrations simultaneously. (**F**) Decoding capacity of contextual multimodal perception in the presence of different inhibitors.

transcriptional state of cells (pPolII) (*17*), pGSK3b scaling with the abundance of late endosomes (VPS35) (*18*), and pEGFR scaling with the abundance of Paxillin (*19*) (Fig. 4B). Only by analyzing these multiple connections between signaling responses and cellular state properties collectively in the same cells and across many cells can we observe that nodes capture not only redundant but also unique and synergistic information about the cellular state (Fig. 4C and fig. S9). To test whether this information propagates through the network, we used inhibitors. The redundant information about cellular state features between upstream (pAKT and pERK) and downstream nodes (respectively nuclear FoxO3a and pS6) was reduced upon inhibition of the upstream node (Fig. 4D), whereas the unique information about the cellular state captured by the downstream node was not reduced and sometimes increased (fig. S8C). For instance, although nuclear depletion of FoxO3a was reduced when AKT was inhibited it was still heterogeneous and correlated more strongly with position in the cell cycle (fig. S8D). Thus, cellular state information can propagate through the network and nodes can integrate information from multiple sources depending on the activity of other nodes.

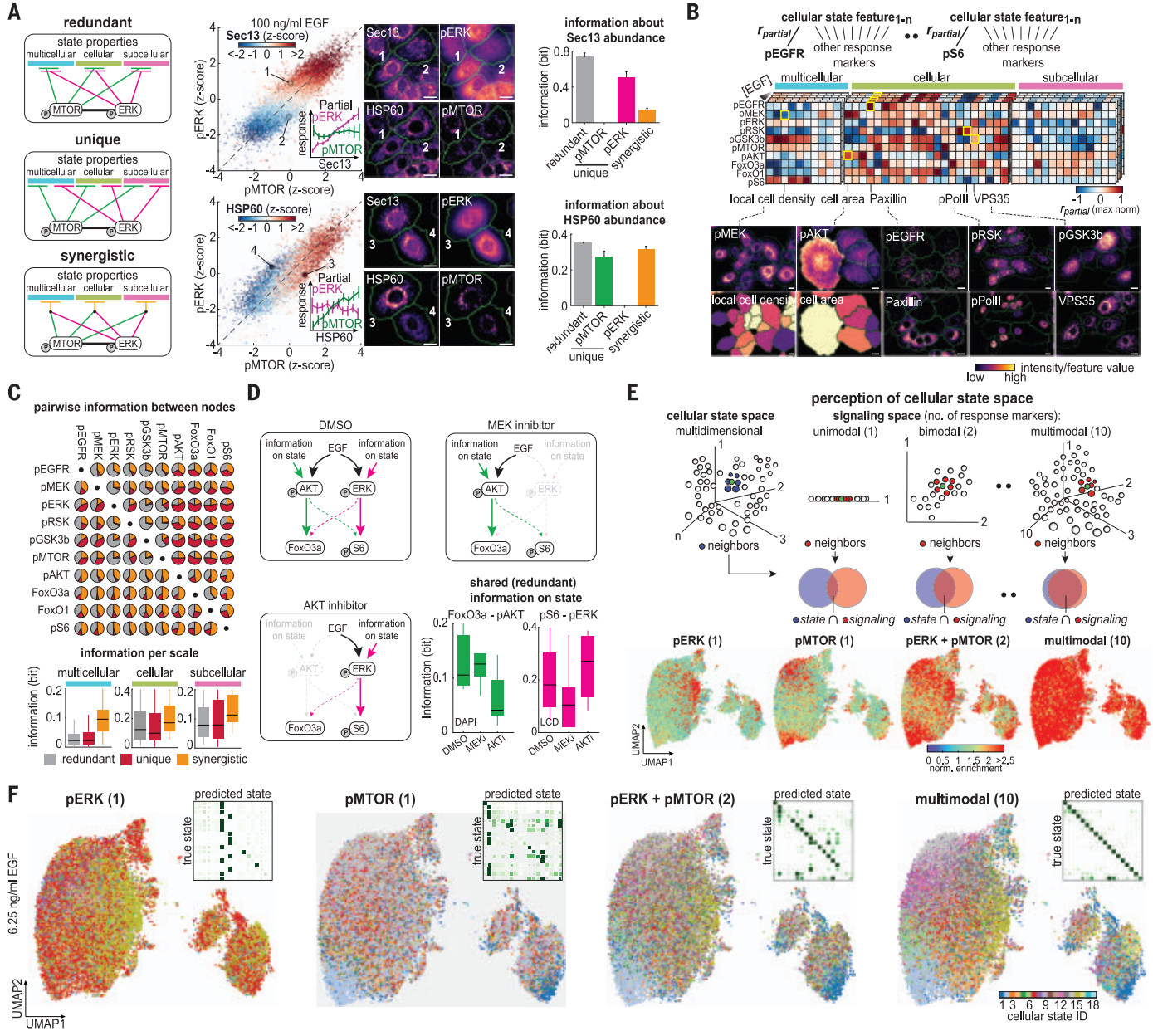


Fig 4. Multimodal perception captures the diversity of cellular states. (**A**) (Left) Encoding regimes of cellular state information by signaling nodes. (Middle left) Scatterplot of pMTOR and pERK abundances with Sec13 (top) or HSP60 (bottom) abundances visualized. (Inlets) Partial scaling of pERK and pMTOR with Sec13 and HSP60. (Middle right) Representative cells (green outlines) illustrating the partial scaling. Scale bars: 5 μm. (Right) Redundant, unique, and synergistic information on Sec13 and HSP60 encoded by pERK and pMTOR from PID. (**B**) (Top) Quantifying the unique scaling signaling nodes display with cellular state features. (Middle) Partial correlations (maximum normalized per signaling node) of each node with cellular state features. (Bottom) Representative cells illustrating the indicated scaling (yellow boxes). Scale bars: 2.5 μm. (**C**) (Top) Relative amount of redundant, unique, and synergistic information about the cellular state (averaged across all features) that signaling nodes encode (pairwise comparison). (Bottom) Encoding cellular state information as average across all pairwise combinations of signaling nodes. (**D**) Transfer of cellular state information from pAKT and pERK to FoxO3a and pS6 in presence of inhibitors and amount of redundant information on the indicated cellular state properties. (**E**) (Top) Calculating neighbor similarity between cellular state space and different signaling spaces. (Bottom) Similarity of cellular state space and different signaling neighborhoods. (**F**) Cellular state classes (as in Fig. 3A) predicted for each cell (logistic classifier) from indicated signaling spaces. (Inlets) Confusion matrices. Color indicates the fraction of cells assigned to a class. Square size indicates posterior probability of cells with that prediction.

To analyze whether collectively, as a multimodal percept, signaling nodes can comprehensively map cellular state space and inform an individual cell on its position in this space, we quantified the overlap between the statistical neighbors of an individual cell in cellular state space with its statistical neighbors in signaling space (Fig. 4E). Unimodal perception can only map neighborhoods within small regions of the cellular state space (Fig. 4E and fig. S10A) whereas a multimodal percept based on 10 nodes can map the neighborhood across the whole cellular state space (Fig. 4E and fig. S10B). The information gain in multimodal perception is particularly high for the first six nodes in the network (fig. S10C). Furthermore, only multimodal perception can accurately predict to which cellular state class an individual cell belongs for a large diversity of cellular states (Fig. 4F and fig. S10, D and E).

Multimodal perception links cellular state to decision-making

To test whether multimodal perception is used by cells to make context-aware decisions we exposed cells to longer EGF stimulation. To avoid problems resulting from the fact that

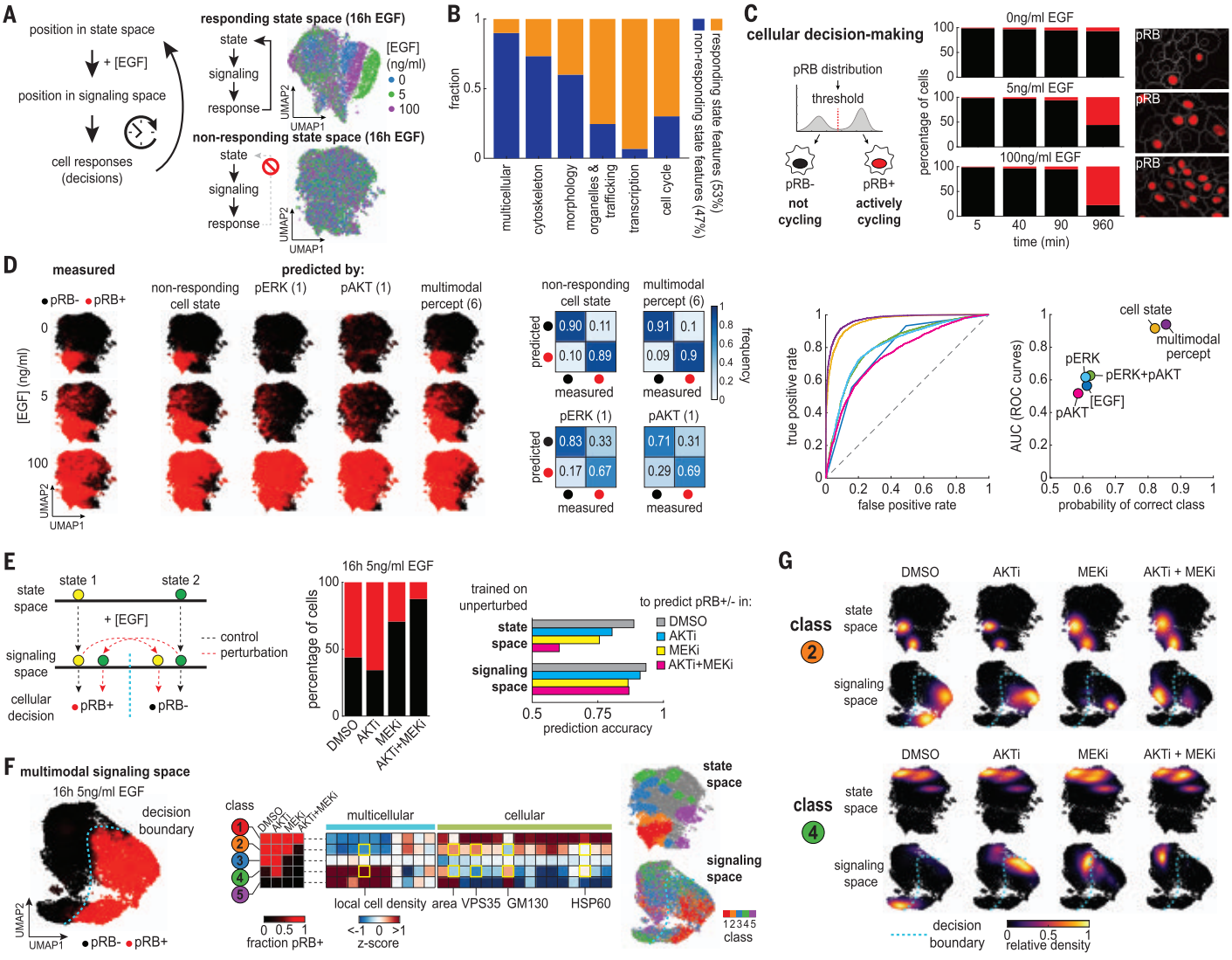


Fig. 5. Multimodal perception links cellular state to decision-making. (**A**) (Left) Position in cellular state space conditions signaling response. On long time scales, this changes the position in state space. (Top right) UMAP of the responding cellular state. (Bottom right) UMAP of the nonresponding cellular state. (**B**) Fractions of cellular state features related to properties and activities that change after 16 hours of EGF. (**C**) (Left) Classification of pRB status. (Middle) Fraction pRB status. (Right) Representative images. (**D**) (Left) Measured and predicted (logistic regression using indicated variables) pRB status on the cellular state landscape. Accuracy (middle), and ROC curves (left) of the classifier. (**E**) (Left) Perturbations that only affect position in signaling space and not in state space, changes the decision expected of distinct cellular states. (Middle) Fraction of pRB status with different inhibitors. (Right) Prediction accuracy of a logistic regression classifier trained on dimethyl sulfoxide (DMSO)–treated cells predicting pRB status in cells treated with inhibitors using the indicated spaces as variables. (**F**) (Left) pRB status visualized on a UMAP of the multimodal signaling space. Dashed blue line, decision boundary. (Middle left) Fraction of pRB status in different cellular state classes across inhibitors. (Middle right) Values of selected cellular state features. (Right) Cellular state class on UMAPs of state and signaling space. (**G**) Density of cellular state classes cells projected on both state and signaling space. Class 2 (top) and Class 4 (bottom).

some properties of the cellular state may change during longer stimulation times, we defined a nonresponding cellular state space (Fig. 5A and fig. S11A). This resulted in the exclusion of about half of all cellular state features and enrichment of features related to the transcriptional state and cell proliferation but also to cell morphology, the cytoskeleton, and membrane trafficking (Fig. 5B and fig. S11A). One of these is phosphorylated retinoblastoma protein (pRB), which is under the control of multiple signaling nodes (*11, 20*). RB is phosphorylated in actively proliferating cells and marks an EGF dose-dependent decision-making event to reenter the cell cycle (Fig. 5C). Projecting the status of individual cells as pRB-negative (pRB−) or pRB-positive (pRB+) after 16 hours of EGF induction on the cellular state landscape showed specific patterns across different doses of EGF, which were accurately predicted by using properties of the cellular state (Fig. 5D and fig. S11B). pRB status was also accurately predicted by the multimodal percept of single cells but less accurately by the responses of individual signaling nodes or their combination (Fig. 5D and fig. S11B). Thus, both the position in cellular state space as well as the position in multimodal signaling space allows for an accurate prediction of the response of single cells to reenter the cell cycle.

We next treated cells with inhibitors of AKT (AKTi) and MEK (MEKi). At low concentrations these inhibitors did not affect the positions of cells in cellular state space but did so in multimodal signaling space (fig. S11, C and D), both alone and in combination, resulting in altered fractions of pRB status (Fig. 5E). These altered responses were not accurately predicted by models based on the cellular state if trained on unperturbed cells but were accurately predicted by models based on the multimodal percept (Fig. 5E, right bar graph). Models based on the cellular state trained on perturbed cells were also accurate (fig. S11E). Thus, the position in multimodal signaling space couples cellular state to decision-making, which is altered by the inhibitors. To explore this we projected pRB status of both untreated and treated cells in the multimodal signaling landscape, revealing a sharp decision boundary (Fig. 5F). We then defined five classes of cells from cellular state space each with a distinct decision-making profile across perturbations. These classes occupy different regions in cellular state space and consequently different regions in multimodal signaling space (Fig. 5F). Although their positions in cellular state space remained the same during inhibitor treatment, their positions in multimodal signaling space changed, affecting the response of these cells (Fig. 5G and fig. S11F). For instance class 2 cells, which grew in regions of low local cell density and were large with abundant late endosomes, were located on the pRB+ side of the decision boundary (Fig. 5F). Individually inhibiting AKT or MEK affected their position in multimodal signaling space but did not result in crossing of the pRB decision boundary (Fig. 5G). Inhibiting both AKT and MEK resulted in altered pRB status (Fig. 5G). Thus, cells in this state were only prevented from re-entering the cell cycle by inhibiting both AKT and MEK. By contrast class 4 cells, which grew in regions of high local cell density, were small and had relatively few endosomes but an abundant Golgi complex, and were located on the pRB− side of the decision boundary in the unperturbed condition (Fig. 5F). Although all treatments affected their position in multimodal signaling space they only became pRB+ upon inhibition of AKT (Fig. 5G). Thus, cells in this state remained nonproliferative upon EGF stimulation but became aberrantly proliferative upon AKT inhibition.

Discussion

We have shown that the heterogeneity in acute signaling responses of individual cells contains partially nonredundant information about the cellular state that influences the growth factor response. The cellular state has a stronger effect on these responses than changes in growth factor concentration and thus represents an important source of information to predict these responses. Collectively, as a multimodal percept this enables individual cells to accurately sense a range of growth factor concentrations and to integrate this with their cellular state to make cellular state–dependent decisions. The cellular state is thus at least as relevant as growth factor concentration in determining cellular responses. This suggests that one purpose of cellular structures in controlling the activation of signaling nodes (*8, 9*) is to inject information about the cellular state into the decision-making process. It may also help explain why signaling responses are heterogeneous and signaling networks have a certain complexity. Although the redundant elements in network complexity can counteract uncertainty in individual responses (*5*) the nonredundant and synergistic elements can enable adaptive responses of multiple nodes to act as a multimodal percept that captures a large amount of information about the cellular state. This suggests that a collective of cells repeatedly generates the same spectrum of single-cell responses through generating the same landscape of cellular states. This may be important during development in which spatial effects and self-organization could drive the robust formation of such landscapes, providing the appropriate context for morphogens to induce a range of cellular decisions even in the absence of well-defined gradients (*21, 22*). That the cellular state determines the heterogeneous response of signaling nodes to clinically tested inhibitors resulting in different and sometimes unwanted state-dependent decisions may also be relevant for the treatment of diseases such as cancer (*23*).

REFERENCES AND NOTES

1. T. Hashimshony, M. Feder, M. Levin, B. K. Hall, I. Yanai, *Nature* **519**, 219–222 (2015).
2. D. Serra *et al*., *Nature* **569**, 66–72 (2019).
3. J. Yao, A. Pilko, R. Wollman, *Mol. Syst. Biol.* **12**, 894 (2016).
4. B. Snijder *et al*., *Nature* **461**, 520–523 (2009).
5. R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, A. Levchenko, *Science* **334**, 354–358 (2011).
6. S. Sampattavanich *et al*., *Cell Syst.* **6**, 664–678.e9 (2018).
7. J. Selimkhanov *et al*., *Science* **346**, 1370–1373 (2014).
8. M. Frechin *et al*., *Nature* **523**, 88–91 (2015).
9. A. V. Vieira, C. Lamaze, S. L. Schmid, *Science* **274**, 2086–2089 (1996).
10. G. Gut, M. D. Herrmann, L. Pelkmans, *Science* **361**, eaar7042 (2018).
11. D. T. Worster *et al*., *Sci. Signal.* **5**, ra19 (2012)
12. E. Becht *et al*., *Nat. Biotechnol.* **37**, 38–44 (2018).
13. S. A. Cepeda-Humerez, J. Ruess, G. Tkačik, *PLOS Comput. Biol.* **15**, e1007290 (2019).
14. S. Zmigrod, B. Hommel, *Multisens. Res.* **26**, 143–157 (2013).
15. N. M. Timme, C. Lapish, *eNeuro* **5**, ENEURO.0052-18.2018 (2018).
16. S. Ito, N. Kioka, K. Ueda, *Biosci. Biotechnol. Biochem.* **83**, 463–471 (2019).
17. T. S. K. Eisinger-Mathason *et al*., *Mol. Cell* **31**, 722–736 (2008).
18. V. F. Taelman *et al*., *Cell* **143**, 1136–1148 (2010).
19. M. Saxena *et al*., *Nat. Mater.* **16**, 775–781 (2017).
20. J. Y. Chen, J. R. Lin, K. A. Cimprich, T. Meyer, *Mol. Cell* **45**, 196–209 (2012).
21. B. Ewen-Campen, T. Comyn, E. Vogt, N. Perrimon, *Cell Rep.* **32**, 108121 (2020).
22. C. Gerri *et al*., *Nature* **587**, 443–447 (2020).
23. M. Fallahi-Sichani, S. Honarnejad, L. M. Heiser, J. W. Gray, P. K. Sorger, *Nat. Chem. Biol.* **9**, 708–714 (2013).
24. B. Kramer, Data_Multimodal_Perception_Cells, Version 2, figshare (2022).

ACKNOWLEDGMENTS

We thank all members of the Pelkmans lab for discussions. We further thank P. Liberali, W. Kramer, R. Klemm, and D. Gilmour for critical suggestions to improve the manuscript. **Funding:** Swiss National Science Foundation grant 310030_192622 (to L.P.); the European Research Council advanced grant CROSSINGSCALES-885579 (to L.P.); the Chan Zuckerberg Initiative grant CZF2019-002440 (to L.P.); and the University of Zurich (to L.P.). **Author contributions:** Conceptualization: B.A.K. and L.P. Methodology: B.A.K. and L.P. Experiments: B.A.K. and J.S.C. Data analysis: B.A.K. Visualization: B.A.K. and L.P. Writing: B.A.K. and L.P. **Competing interests:** L.P. has filed a patent on the 4i technology. **Data and materials availability:** All scripts and code used, and detailed description of the data can be accessed at https://github.com/hlardirian/Multimodal_Signaling. All data used can be downloaded at (*24*) **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf4062
Materials and Methods
Figs. S1 to S11
Tables S1 to S5
References (*25–27*)

Submitted 23 October 2020; resubmitted 3 September 2021
Accepted 1 July 2022
10.1126/science.abf4062

REPORTS

ORGANIC CHEMISTRY

Carbene reactivity from alkyl and aryl aldehydes

Lumin Zhang, Bethany M. DeMuynck†, Alyson N. Paneque†, Joy E. Rutherford†, David A. Nagib*

Carbenes are highly enabling reactive intermediates that facilitate a diverse range of otherwise inaccessible chemistry, including small-ring formation and insertion into strong σ bonds. To access such valuable reactivity, reagents with high entropic or enthalpic driving forces are often used, including explosive (diazo) or unstable (*gem*-dihalo) compounds. Here, we report that common aldehydes are readily converted (via stable α-acyloxy halide intermediates) to electronically diverse (donor or neutral) carbenes to facilitate >10 reaction classes. This strategy enables safe reactivity of nonstabilized carbenes from alkyl, aryl, and formyl aldehydes via zinc carbenoids. Earth-abundant metal salts [iron(II) chloride ($FeCl_2$), cobalt(II) chloride ($CoCl_2$), copper(I) chloride (CuCl)] are effective catalysts for these chemoselective carbene additions to σ and π bonds.

In organic synthesis, several important classes of chemical reactivity are exclusively mediated by carbenes and carbenoids (Fig. 1A) (*1*–*5*). However, access to these versatile intermediates is often limited by the need for highly energetic diazoalkane precursors. Given the strong entropic and enthalpic driving forces inherent in the release of N_2 from these reagents, strict precautions are necessary to prevent explosive, uncontrolled chain reactions (*6*). Strategies to address this underlying safety issue include flow chemistry and in situ diazotization (*7*–*9*). Carbonyl and arene stabilizing groups are also frequently used (*10*), which influence carbene polarity and selectivity (*11*, *12*). Yet nonstabilized alkyl carbenes (with eliminable α protons) are rarely accessed by diazoalkanes (*13*).

In a complementary approach, Simmons-Smith cyclopropanations by Zn carbenoids (from *gem*-dihalides) enable incorporation of the smallest divalent carbon, CH_2 (*14*, *15*). However, α protons are rarely tolerated owing to 1,2-H migration of the Zn carbenoid (*16*). Despite elegant solutions for Matteson rearrangement by Li-boronate carbenoids (*17*–*20*), Mo-mediated ketone deoxygenation (*21*, *22*), and Au-catalyzed alkyne cyclizations (*23*, *24*), there remains no general approach to access nonstabilized alkyl carbenes or for their use in the wide range of carbene reactivity available to stabilized diazo reagents (*25*).

The carbonyl is an ideal carbene precursor, because it is highly accessible, both commercially and synthetically. Toward this goal, we were inspired by the classic Clemmensen reduction of carbonyls with Zn(Hg) and HCl (*26*, *27*). Motherwell also harnessed this Zn carbenoid reactivity to deoxygenate ketones by 1,2-H migration of siloxy analogs (*28*). However, this tendency toward α-elimination precludes access to the broad range of diazo-based carbene reactivity (*29*).

In designing a general catalytic strategy to convert carbonyls to electronically diverse carbenes (Fig. 1B), we sought to develop a safe, stable carbenoid precursor that does not rely on activation by highly eliminable diazo, dihalo, or siloxy groups. Instead, we were cognizant that acyl chloride (AcCl) readily adds to carbonyls, yielding stable α-acyloxy halides, even with enolizable aldehydes (*30*). We previously showed that acyl iodide (AcI) adducts of carbonyls enable distinct ketyl radical reactivity by atom- or electron-transfer reduction mechanisms (*31*, *32*). By contrast, we hypothesized that the more stable pivaloyl chloride (PivCl) adduct **A** may prevent the radical pathway and permit chemoselective formation of α-acyloxy Zn carbenoid **B** (*33*). Importantly, by accessing this key intermediate in the absence of strong acids, we proposed that the controlled α-acyloxy elimination by base metal catalysts could form reactive, metal carbene **C**. We anticipated that catalyst influence could also impart distinct reactivity compared with simple Zn carbenoids. Ideally, the resulting (i) carbene dimerization, (ii) σ bond insertion, or (iii) small-ring formation would be chemoselectively dictated by these catalysts and

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA.
*Corresponding author. Email: nagib.1@osu.edu
†These authors contributed equally to this work.

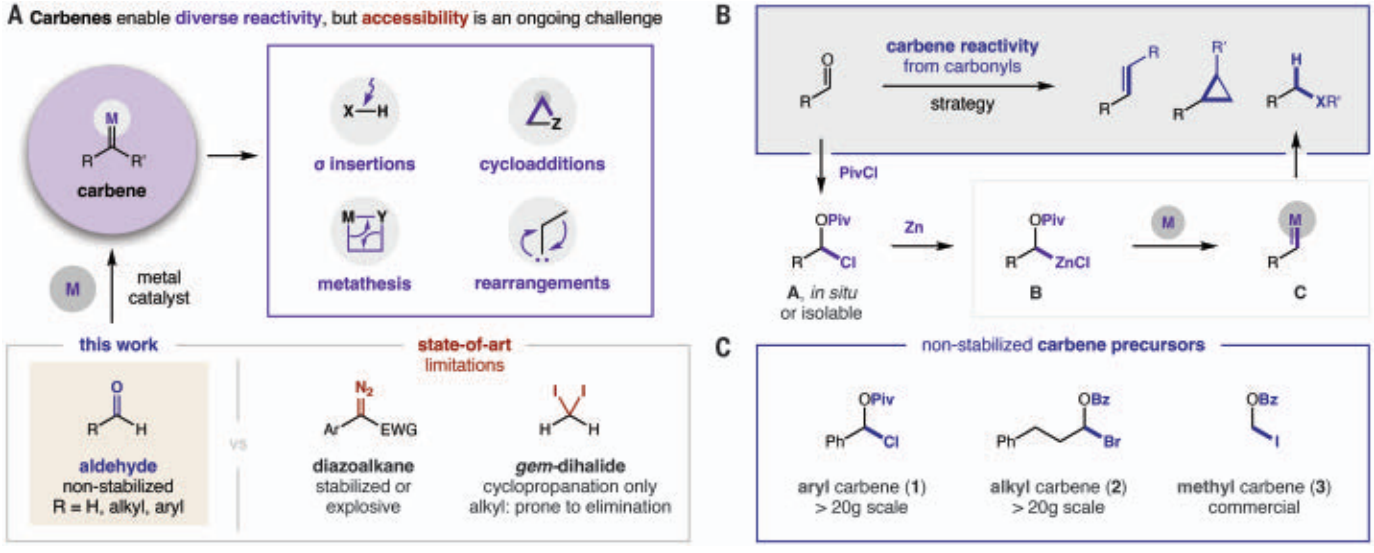


Fig. 1. Strategies to harness carbene reactivity. (**A**) Synthetic approaches to carbenes rely on either loss of N_2 from diazoalkanes or reduction of *gem*-dihalides. Both reagent classes exhibit limited scope, stability, and reactivity in the absence of aryl or α-carbonyl stabilization. (**B**) Our design converts carbonyls to nonstabilized carbenes via Zn insertion into α-acyloxy halides and activation by Earth-abundant metal catalysts. (**C**) Three classes of carbene precursors are readily prepared and exhibit substantially improved safety profiles versus diazo reagents. Ar, aryl; Bz, benzoyl; EWG, electron-withdrawing group; M, metal catalyst; Ph, phenyl; Piv, pivaloyl; X, Y, and Z, generic atoms.

override innate 1,2-rearrangements of alkyl Zn carbenoids.

Another key design element is that several classes of these carbene precursors are easily prepared, stable for months in cold storage, and handled safely at large scale. As shown in Fig. 1C, we found that less-electrophilic carbonyls are best paired with more-electrophilic acyl halides (Cl < Br < I) to afford increasingly reactive carbene precursors (α-Cl < α-Br < α-I). For example, PivCl readily adds to benzaldehyde to generate bench-stable aryl carbene precursor **1**. Yet, less-electrophilic, aliphatic aldehydes are best combined with benzoyl bromide (BzBr) to form nonstabilized alkyl analog **2**, which is nonetheless stable to chromatography. Both reactions can be performed at >20-g scale without the safety concerns that are inherent to analogous diazo analogs, which violently decompose upon loss of N_2 near 100°C [with a decomposition enthalpy (ΔH_D) >200 kJ/mol, as measured by differential scanning calorimetry (DSC)] (*6*). Conversely, these α-acyloxy halides are considerably more stable, with drastically less decomposition enthalpies observed (ΔH_D < 30 kJ/mol in all cases) and only at much higher temperatures than typical reaction conditions (T_{init}: **1**, 155°C; **2**, >300°C; **3**, 235°C). Lastly, benzoyl iodide (BzI) readily combines with the least-reactive aldehyde, formaldehyde, to access methyl carbene precursor **3**, which is a common protecting group that is commercially available in kilogram quantities and provides a safe alternative to diazomethane.

To investigate our design, carbene dimerization was examined by subjecting benzaldehyde to the proposed strategy, a three-stage sequence entailing (i) PivCl addition, (ii) Zn insertion, (iii) and metal catalysis. This procedure can either be completed successively in one pot or by preisolation of PivCl adduct **1**. As shown in Fig. 2, several base metal salts (e.g., CuCl, $CoCl_2$) enable efficient dimerization to stilbene **4**, with notably high catalyst stereocontrol observed for $CoCl_2$ [90%, 18:1 *E:Z*, via **1**], which is surprising given the absence of bulky ligands that have been shown to be integral for selectivity in other carbenoid reactions (*16*). As further evidence of catalyst control, no reaction was observed with $NiCl_2$ or without catalyst. In addition to serving as a proof of concept of carbene reactivity, these mechanistic probes also demonstrate that electronically diverse substituents (**6** to **8**) are well tolerated in the aryl carbene component.

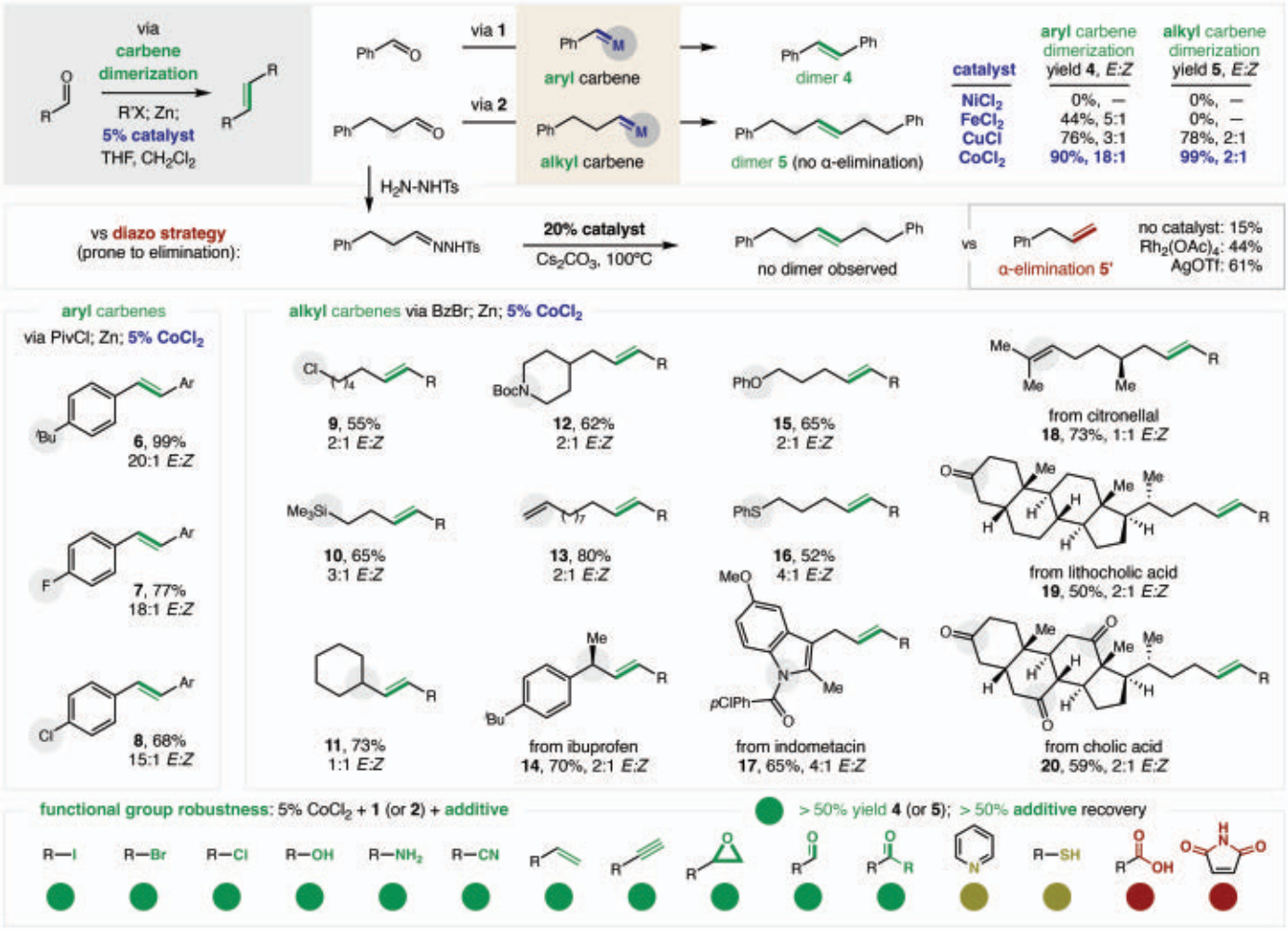


Fig. 2. Carbene dimerization. Catalytic dimerization of aldehyde-derived aryl and alkyl carbenes enabled by several base metal salts. The synthetic generality and broad functional group tolerance of this carbene dimerization are shown. Reaction conditions for the aryl carbene are as follows: aldehyde (1 equiv), PivCl (2 equiv), 2% $ZnCl_2$, $CHCl_3$, 2 hours; Zn (2 equiv), LiCl (2 equiv), THF, 12 hours; 5% $CoCl_2$, CH_2Cl_2:THF, 2 hours. Reaction conditions for the alkyl carbene are as follows: aldehyde (1 equiv), BzBr (1.2 equiv), 2% $ZnBr_2$, CH_2Cl_2, 2 hours; Zn (2 equiv), LiCl (2 equiv), THF, 12 hours; 5% $CoCl_2$, CH_2Cl_2:THF, 2 hours. To evaluate robustness, experiments were performed with additive and either **1** or **2** (1 equiv each). See the supplementary materials for full experimental details. Isolated yield and alkene stereochemistry (*E:Z*) are indicated. OTf, trifluoromethanesulfonate; *p*, *para*; ^{t}Bu, *tert*-butyl; Ts, toluenesulfonyl.

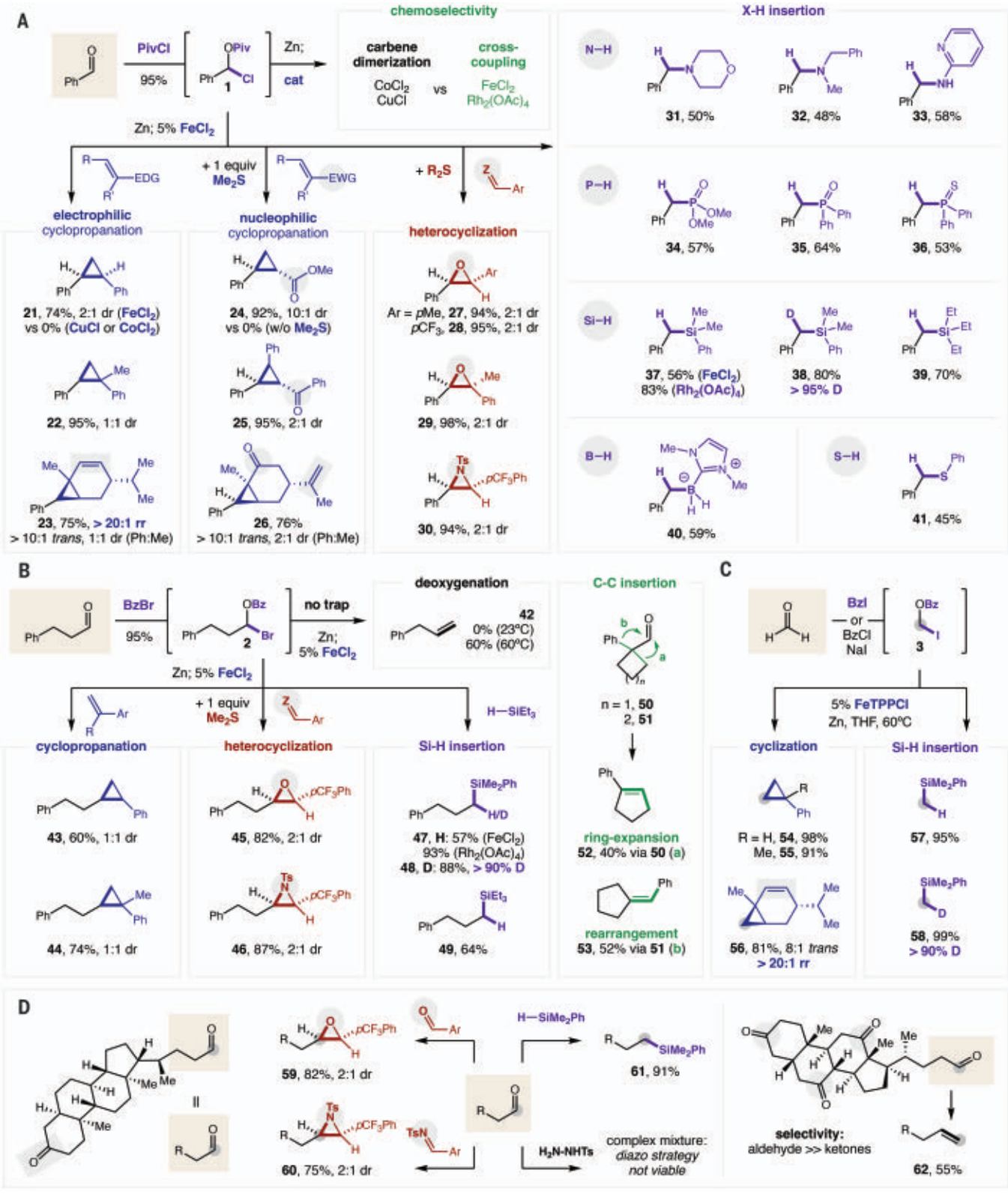


Fig. 3. Catalytic carbene reactivity. (**A**) Aryl carbene. Reaction conditions are as follows: aldehyde (2 equiv), PivCl (4 equiv), 4% $ZnCl_2$, $CHCl_3$, 2 hours; Zn (2 equiv), LiCl (2 equiv), THF, 6 hours; 5% $FeCl_2$ or 2.5% $Rh_2(OAc)_4$, carbene trap (1 equiv), CH_2Cl_2:THF, 12 hours. (**B**) Alkyl carbene. Reaction conditions are as follows: aldehyde (3 equiv), BzBr (3.6 equiv), 4% $ZnBr_2$, CH_2Cl_2, 2 hours; Zn (3 equiv), LiCl (3 equiv), THF, 6 hours; 5% $FeCl_2$, carbene trap (1 equiv), CH_2Cl_2:THF, 12 hours. (**C**) Methyl carbene. Reaction conditions are as follows: iodide (3 equiv), Zn (3 equiv), 5% FeTPPCl, carbene trap (1 equiv), CH_2Cl_2:THF, 4 hours. (**D**) Complex molecule applications. See the supplementary materials for full experimental details. Isolated yield, diastereomeric ratio (dr), and regioselectivity ratio (rr) are indicated. D, deuterium; EDG, electron-donating group; Et, ethyl; Me, methyl.

We next tested our hypothesis that alkyl carbene reactivity may be harnessed by this catalytic approach in a safer and more selective manner than by diazoalkanes or dihalides. The distinctively broad scope of aliphatic aldehydes (even with eliminable α protons) that could be converted to carbenes was the highlight of this initial study. To this end, we found that $CoCl_2$ effectively catalyzes carbene dimerization from hydrocinnamaldehyde (to alkene **5** via **2**: 99%, 2:1 *E:Z*). Notably, no α-elimination is observed, in contrast to the uncatalyzed siloxy variant, which is limited to α,β-unsaturated carbonyls to prevent such a pathway (*28*, *34*). For further comparison, the diazo strategy was examined by subjecting a hydrazone to various catalysts and bases to access its carbene. In all cases, only α-elimination is observed (up to 61% **5′**), likely owing to the high temperatures (100°C) required for reactivity.

By contrast, the valuable alkyl carbene reactivity enabled by this approach includes dimerization of aldehydes that contain synthetically tractable halides, alkenes, silanes, carbamates, ethers, sulfides, amides, ketones, and α substituents, including within natural product and drug scaffolds (**9** to **20**). In all cases, rapid dimerization preferentially occurs over other types of carbene reactivity, such as intramolecular cyclopropanation (**18**). Moreover, a robustness study (*35*) was conducted to further assess the broad functional group tolerance of this strategy, wherein chemically diverse additives were combined with catalyst and carbene precursors **1** (or **2**) to explore their viability in the dimerization step.

We then subjected these aldehyde-derived carbene precursors to several distinct classes of reactivity, including signature carbene reactions: cyclopropanation and X-H insertion (*36*, *37*) (Fig. 3A). While probing reactivity with several common carbene traps, we observed that the catalysts that promoted rapid dimerization (vide supra; $CoCl_2$, CuCl) are less efficient at chemoselective cross-coupling. Conversely, metallocarbenes generated from $FeCl_2$ or rhodium(II) acetate [$Rh_2(OAc)_4$]

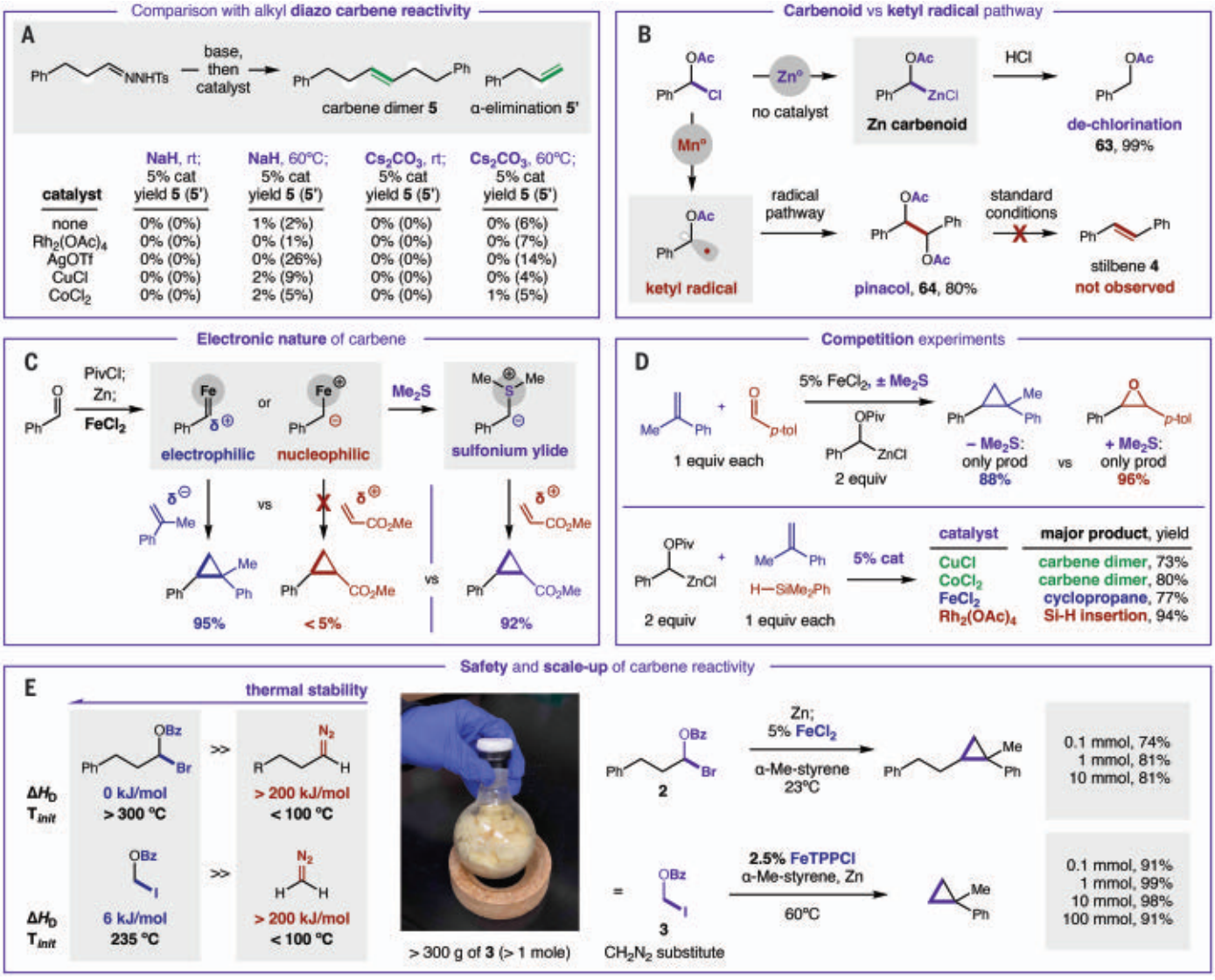


Fig. 4. Mechanistic experiments. (**A**) The carbene dimerization reactivity shown above is not accessible via diazoalkanes. (**B**) The proposed Zn carbenoid is validated by an acidic quench. Conversely, single-electron reduction by Mn affords ketyl radical reactivity but not carbene dimerization. (**C**) Electrophilic character of carbene is illustrated by selective reactivity with nucleophilic traps; sulfonium ylide affords polarity-reversed reactivity with electrophiles. (**D**) Competition experiments illustrate catalyst control of product selectivity. (**E**) The improved thermal stability of these carbene precursors relative to diazoalkanes allows a safer scale-up of batch chemistry. The photo shows the simple handling of a large quantity of **3**, which is a safer alternative to diazomethane. Ac, acetyl; cat, catalyst; rt, room temperature; tol, tolyl. [Photo credit: D.A.N.].

enable alternate modes of carbene reactivity aside from dimerization. For example, cyclopropanation of styrenes with aryl carbenes is efficiently catalyzed by $FeCl_2$ (**21** and **22**, up to 95%) but not $CoCl_2$. In contrast with typical Simmons-Smith selectivity (*38*, *39*), this Fe-catalyzed strategy is specifically selective for more-substituted, electron-rich alkenes (e.g., monoterpene **23**). Conversely, cyclopropanation of electron-deficient alkenes does not occur, as expected. However, if Me_2S is added as a cocatalyst to access the polarity-reversed sulfonium ylide (see Fig. 4C) as has been demonstrated with diazoalkanes (*40*), then electrophilic acrylates and enones can now be selectively benzylated (**24** and **25**), including in selective competition over other alkenes, as in the case of carvone (**26**). This sulfonium-mediated strategy also affords epoxides and aziridines by combination of these aldehyde-derived carbenes with electronically diverse aldehydes and imines (**27** to **30**), without the need for diazoalkanes and in a more modular fashion that relies on aldehyde rather than typical sulfonium ylide precursors.

Besides cyclopropanation, the most commonly used reactivity of metal carbenes is catalytic X-H insertion (*36*, *37*). Although Zn-carbenoids have not been useful in enabling such reactivity, we sought to examine if our metal-catalyzed activation could mirror the robust reactivity of diazo reagents. We found that a range of alkyl and aryl amines efficiently underwent N-H insertion to generate benzyl amines (**31** to **33**). Insertion into a less-polarized P–H bond also occurred with three classes of organophosphorus compounds, to access phosphates, phosphine oxides, and phosphine sulfides (**34** to **36**). Given the electrophilic nature of this carbene, other nucleophilic X-H donors were also examined. Si-H insertion into silanes occurred smoothly with either $FeCl_2$ (56% **37**) or a more typical carbene catalyst, $Rh_2(OAc)_4$ (83%). Insertion into Si-D yielded complete α-deuteration (>95% D, **38**) in tetrahydrofuran (THF), confirming the carbene nature of this catalytic mechanism, which is amenable to several classes of alkyl silanes (**37** to **39**). Insertion also occurred into σ bonds of varying polarity, including B–H and S–H bonds (**40** and **41**). Overall, we observed that more-nucleophilic H donors (e.g., phosphine, silane, borane) are highly suited to this carbene reactivity, which has now been applied to 10 reaction classes [dimerization, four (2+1) cyclizations, five X-H insertions].

Because none of these alkyl carbene reactions has been previously accessible by Zn carbenoid activation of carbonyls, we sought to test the limits of this reactivity with non-stabilized alkyl aldehydes (Fig. 3B). First, we noted that typical 1,2-H migration (*28*, *34*) did not occur in this mild system at room temperature. Instead, heating (60°C) was needed to promote catalytic deoxygenation (**42**). This lack of background H migration enables successful realization of several classes of alkyl carbene reactivity, including cyclopropanation (**43** and **44**), epoxidation and aziridination (**45** and **46**), and σ bond insertion (**47** to **49**), all with similar efficiency to aryl carbenes. Notably, alkyl carbene reactivity is typically challenging to access by other methods because diazoalkane and dihalide precursors are highly prone to H migration. Having harnessed this new substrate class, we also investigated C-C insertion, wherein ring expansion of α-cyclobutanes and rearrangement of α-cyclopentanes were each observed (**50** to **53**). Lastly, we confirmed viability of the simplest carbene (CH_2) for both reaction classes (Fig. 3C): π bond cycloaddition (**54** to **56**) and σ bond insertion (**57** and **58**). We expect this formaldehyde-based approach to provide a safer alternative to conventional diazomethane reactivity and to complement modern methods that entail in situ generation (*7*–*9*).

To demonstrate viability in more complex settings, the aldehydes of two bile acids (among the most complex carbonyls dimerized in Fig. 2) were subjected to four distinct classes of carbene reactivity (Fig. 3D). First, the lithocholic acid containing both a ketone and aldehyde was selectively elongated at the aldehyde by epoxidation (**59**), aziridination (**60**), and Si-H insertion (**61**). A comparison with the diazo strategy was not possible because hydrazone formation was unselective between the two carbonyls within this molecule. Additionally, this approach distinctively enables deoxygenation of an aldehyde in the presence of three ketones (**62**).

Our mechanistic understanding of this catalytic carbene reactivity from aldehydes is based on a collection of experiments, including intermediate characterization, reactivity comparisons, and kinetic data (Fig. 4). An investigation of various bases (NaH, Cs_2CO_3), temperatures (0° to 100°C), and catalysts (Fig. 4A) verifies that the carbene dimerization is not accessible via diazoalkanes. The proposed Zn carbenoid, generated in the absence of catalyst, was characterized by protodechlorination with HCl (**63**) (Fig. 4B). Conversely, a stronger Mn reductant affords pinacol coupling via ketyl radicals (**64**). This product remains unchanged and does not afford stilbene (**4**) when resubjected to reaction conditions. In probing the electronic nature of the catalytic carbene intermediate, we noted its electrophilicity, as evidenced by higher efficiency of reactivity with nucleophilic alkenes (α-Me-styrene; Me, methyl) versus electrophilic traps (acrylate) (Fig. 4C). However, upon introduction of a sulfide cocatalyst, a transient sulfonium ylide (observed by gas chromatography–mass spectrometry) enables inverted reactivity with such electrophiles (e.g., **24**), without the need for diazo intermediates (*40*). Furthermore, in a 1:1 competition of an alkene and aldehyde as potential carbene traps, exclusive selectivity is observed for epoxidation with Me_2S versus only cyclopropanation without Me_2S (Fig. 4D). Notably, the metal catalyst also dictates reaction chemoselectivity. For example, in a competition experiment among three classes of carbene traps, dimerization is exclusively observed with CuCl (73%) and $CoCl_2$ (80%), yet $FeCl_2$ predominantly affords cyclopropanation (77%) and $Rh_2(OAc)_4$ yields Si-H insertion (94%). Lastly, in situ infrared spectroscopy of the carbene dimerization reaction and variable time normalization analysis (VTNA) (*41*) indicate that the dimerization reaction is second order in catalyst (see supplementary materials).

Because this carbene generation mechanism does not rely on the evolution of N_2 gas, a substantial improvement in thermal stability was observed for these carbene precursors (ΔH_D = 0 kJ/mol for **2** and 6 kJ/mol for **3**) relative to diazoalkanes (ΔH_D > 200 kJ/mol) (Fig. 4E). To harness this superior safety profile, large-scale batch reactions were performed with quantities that would be prohibitively unsafe with diazoalkanes (e.g., CH_2N_2). In these cases, cyclopropanations succeeded at up to a 1000-times-larger scale without event.

We have introduced a catalytic strategy for harnessing carbene reactivity from carbonyls. We expect that the approach will have a threefold impact on the expanded development of carbene reactions, including through the use of base metal catalysts, the use of safe and scalable reagents, and improved synthetic access to nonstabilized carbenes.

REFERENCES AND NOTES

1. M. S. Sanford, J. A. Love, R. H. Grubbs, *J. Am. Chem. Soc.* **123**, 6543–6554 (2001).
2. H. M. L. Davies, J. R. Manning, *Nature* **451**, 417–424 (2008).
3. P. de Frémont, N. Marion, S. P. Nolan, *Coord. Chem. Rev.* **253**, 862–892 (2009).
4. V. H. Gessner, *Chem. Commun.* **52**, 12011–12023 (2016).
5. A. Caballero, P. J. Pérez, *Chem. Eur. J.* **23**, 14389–14393 (2017).
6. S. P. Green *et al*., *Org. Process Res. Dev.* **24**, 67–84 (2020).
7. L. D. Proctor, A. J. Warr, *Org. Process Res. Dev.* **6**, 884–892 (2002).
8. B. Morandi, E. M. Carreira, *Science* **335**, 1471–1474 (2012).
9. T. A. Hatridge, B. Wei, H. M. L. Davies, C. W. Jones, *Org. Process Res. Dev.* **25**, 1911–1922 (2021).
10. A. Ford *et al*., *Chem. Rev.* **115**, 9981–10080 (2015).
11. K. Liao, S. Negretti, D. G. Musaev, J. Bacsa, H. M. L. Davies, *Nature* **533**, 230–234 (2016).
12. Z. Liu *et al*., *Chem* **6**, 2110–2124 (2020).
13. D. Zhu, L. Chen, H. Fan, Q. Yao, S. Zhu, *Chem. Soc. Rev.* **49**, 908–950 (2020).
14. H. E. Simmons, R. D. Smith, *J. Am. Chem. Soc.* **80**, 5323–5324 (1958).
15. A. B. Charette, A. Beauchemin, in vol. 58 of *Organic Reactions*, L. E. Overman *et al*., Eds. (Wiley, 2001), pp. 1–415.
16. J. Werth, C. Uyeda, *Angew. Chem. Int. Ed.* **57**, 13902–13906 (2018).
17. D. Leonori, V. K. Aggarwal, *Acc. Chem. Res.* **47**, 3174–3183 (2014).
18. M. Burns *et al*., *Nature* **513**, 183–188 (2014).

19. L. Zhang *et al*., *Science* **351**, 70–74 (2016).
20. H. A. Sharma, J. Z. Essman, E. N. Jacobsen, *Science* **374**, 752–757 (2021).
21. S. Asako, S. Ishihara, K. Hirata, K. Takai, *J. Am. Chem. Soc.* **141**, 9832–9836 (2019).
22. L. Y. Cao *et al*., *Angew. Chem. Int. Ed.* **60**, 15254–15259 (2021).
23. M. R. Luzung, J. P. Markham, F. D. Toste, *J. Am. Chem. Soc.* **126**, 10858–10859 (2004).
24. D. Benitez *et al*., *Nat. Chem.* **1**, 482–486 (2009).
25. M. Jia, S. Ma, *Angew. Chem. Int. Ed.* **55**, 9134–9166 (2016).
26. E. Clemmensen, *Chem. Ber.* **46**, 1837–1843 (1913).
27. J. H. Brewster, *J. Am. Chem. Soc.* **76**, 6364–6368 (1954).
28. W. B. Motherwell, *J. Chem. Soc. Chem. Commun.* **1973**, 935 (1973).
29. W. B. Motherwell, *J. Organomet. Chem.* **624**, 41–46 (2001).
30. H. French, R. Adams, *J. Am. Chem. Soc.* **43**, 651–659 (1921).
31. L. Wang, J. M. Lear, S. M. Rafferty, S. C. Fosu, D. A. Nagib, *Science* **362**, 225–229 (2018).
32. S. M. Rafferty, J. E. Rutherford, L. Zhang, L. Wang, D. A. Nagib, *J. Am. Chem. Soc.* **143**, 5622–5628 (2021).
33. P. Knochel, R. D. Singer, *Chem. Rev.* **93**, 2117–2188 (1993).
34. A. K. Banerjee, M. C. Sulbaran De Carrasco, C. S. V. Frydrych-Houge, W. B. Motherwell, *J. Chem. Soc. Chem. Commun.* **1986**, 1803–1805 (1986).
35. K. D. Collins, F. Glorius, *Nat. Chem.* **5**, 597–601 (2013).
36. M. Brookhart, W. B. Studabaker, *Chem. Rev.* **87**, 411–432 (1987).
37. D. Gillingham, N. Fei, *Chem. Soc. Rev.* **42**, 4918–4931 (2013).
38. E. C. Friedrich, F. Niyati-Shirkhodaee, *J. Org. Chem.* **56**, 2202–2205 (1991).
39. J. Werth, C. Uyeda, *Chem. Sci.* **9**, 1604–1609 (2018).
40. V. K. Aggarwal, C. L. Winn, *Acc. Chem. Res.* **37**, 611–620 (2004).
41. J. Burés, *Angew. Chem. Int. Ed.* **55**, 2028–2031 (2016).

ACKNOWLEDGMENTS

C. Acarturk and L. Burris assisted with DSC measurements. J. Herbort and T. V. RajanBabu assisted with VTNA experiments. **Funding:** This work was funded by the National Institutes of Health (NIH R35 GM119812), the National Science Foundation (NSF CAREER 1654656), and the Sloan Foundation. J.E.R. is supported by an NIH Fellowship (F31). **Author contributions:** L.Z. and D.A.N. designed this strategy, and L.Z. demonstrated its viability. All authors contributed to designing, performing, and analyzing experiments, as well as to writing this manuscript. B.M.D., A.N.P., and J.E.R. contributed equally and are listed alphabetically. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.
License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo6443
Materials and Methods
Figs. S1 to S14
Tables S1 to S17
NMR Spectra

Submitted 16 February 2022; accepted 10 June 2022
10.1126/science.abo6443

OCEAN OXYGEN

Changes in North Atlantic deep-water oxygenation across the Middle Pleistocene Transition

Nicola C. Thomas*, Harold J. Bradbury, David A. Hodell

The oxygen concentrations of oceanic deep-water and atmospheric carbon dioxide (pCO_2) are intrinsically linked through organic carbon remineralization and storage as dissolved inorganic carbon in the deep sea. We present a high-resolution reconstruction of relative changes in oxygen concentration in the deep North Atlantic for the past 1.5 million years using the carbon isotope gradient between epifaunal and infaunal benthic foraminifera species as a proxy for paleo-oxygen. We report a significant (>40 micromole per kilogram) reduction in glacial Atlantic deep-water oxygenation at ~960 thousand to 900 thousand years ago that coincided with increased continental ice volume and a major change in ocean thermohaline circulation. Paleo-oxygen results support a scenario of decreasing deep-water oxygen concentrations, increased respired carbon storage, and a reduction in glacial pCO_2 across the Middle Pleistocene Transition.

During the past 800 thousand years (kyr) glacial carbon dioxide concentrations (pCO_2) in Earth's atmosphere averaged ~190 parts per million (ppm) by volume (*1*), lower than the preindustrial value by ~90 ppm (*1*, *2*). Aside from ice core records, atmospheric pCO_2 records are fragmentary but suggest that glacial pCO_2 may have been higher by 20 to 40 ppm prior to ~1000 to 800 thousand years ago (ka) during the early Pleistocene (*3*–*7*). Such a drop in glacial atmospheric pCO_2 concentrations is one of the proposed causes of increased continental ice volume during glacial periods across the Middle Pleistocene Transition (MPT) (*8*, *9*), which occurred between 1.25 and 0.64 million years ago (Ma) (*10*). During the MPT ice sheets grew larger and the duration of glacial cycles increased from primarily 41-kyr oscillations before the MPT to quasi–100-kyr cycles afterward (*8*, *10*, *11*). Lowering glacial pCO_2 across the MPT most likely involved increased carbon storage in the deep ocean through enhanced biological CO_2 uptake (*12*–*14*) and/or reduced CO_2 exchange between the atmosphere and surface ocean (*12*, *15*, *16*) especially in the Southern Ocean (*17*, *18*). These processes would not only increase carbon storage in the deep ocean but would also reduce the deep-sea concentration of dissolved oxygen.

We applied a paleoproxy to estimate past changes in oxygen concentration (hereafter expressed as $[O_2]$) in deep water of the North Atlantic. The proxy uses an empirical calibration between the $[O_2]$ and the carbon isotopic composition difference ($\Delta\delta^{13}C$) between surface and deep-dwelling benthic foraminifera, as first proposed by McCorkle *et al*. (*19*, *20*) and recalibrated by Hoogakker *et al*. (*21*) (see supplementary materials). We report $\Delta\delta^{13}C$ as a proxy for North Atlantic deep-water $[O_2]$ at International Ocean Discovery Program (IODP) Site U1385 on the Iberian Margin [37°34.285' N, 10°7.562' W, 2578 meters below sea level (mbsl)] (*22*, *23*) (Fig. 1; supplementary materials). We measured the epibenthic species *Cibicidoides wuellerstorfi* that provides a record of deep-water carbon isotopic composition ($\delta^{13}C$) (*21*, *24*), whereas the infaunal species *Globobulimina affinis* provides a record of $\delta^{13}C$ near the oxic-anoxic boundary in sediment porewaters (*21*, *24*, *25*). The carbon isotope gradient ($\Delta\delta^{13}C_{cib\text{-}aff}$) expresses the difference between the two.

The main assumption of the $\Delta\delta^{13}C_{cib\text{-}aff}$ derived $[O_2]$ method is that the carbon isotope gradient between the sediment water interface and the oxic-anoxic boundary is controlled by deep-water $[O_2]$ and hence total dissolved inorganic carbon (DIC) released during aerobic respiration. Addition of excess DIC through anaerobic processes such as nitrate reduction or sulfate reduction serves to lower $\delta^{13}C$ at the oxic-anoxic boundary (*20*, *26*) and causes an overestimation of $[O_2]$ (*26*). Thus, derived paleo-$[O_2]$ estimates represent maximum deep-water oxygen concentrations. The Hoogakker *et al*. (*21*) calibration includes release of DIC through anaerobic processes as demonstrated by the nonzero *y* intercept in their empirical calibration (see supplementary text, section 2).

Low organic carbon concentrations (<1%) at Site U1385 leads to relatively low rates of organic carbon oxidation and a deep sulfate-methane transition at ~50 meters below the seafloor (mbsf) (*27*). The Iberian Margin near-surface sediment column demonstrates low rates of sulfate reduction in the upper tens of centimeters (*28*), suggesting that $\Delta\delta^{13}C$ at Site U1385 is governed predominantly by the amount of aerobic respiration controlled by deep-water $[O_2]$. According to recommendations of Jacobel *et al*. (*26*), we compare the paleo-$[O_2]$ estimates derived from $\Delta\delta^{13}C_{cib\text{-}aff}$ with other redox/oxygenation proxies: U/Ca, U/Mn, and percent $C_{26}OH/(C_{26}OH + C_{29})$. We also assess relative changes in paleoproductivity that may lead to variable organic carbon fluxes using several proxies: $\Delta\delta^{13}C_{cib\text{-}uvig}$, *Uvigerina* spp. abundance, and sedimentary Ba/Al ratios (see supplementary text, sections 3 and 4).

The cores from different holes at Site U1385 were spliced to produce a composite section

Department of Earth Science, University of Cambridge, Cambridge, UK.
*Corresponding author. Email: nct30@cam.ac.uk; nct3051@gmail.com

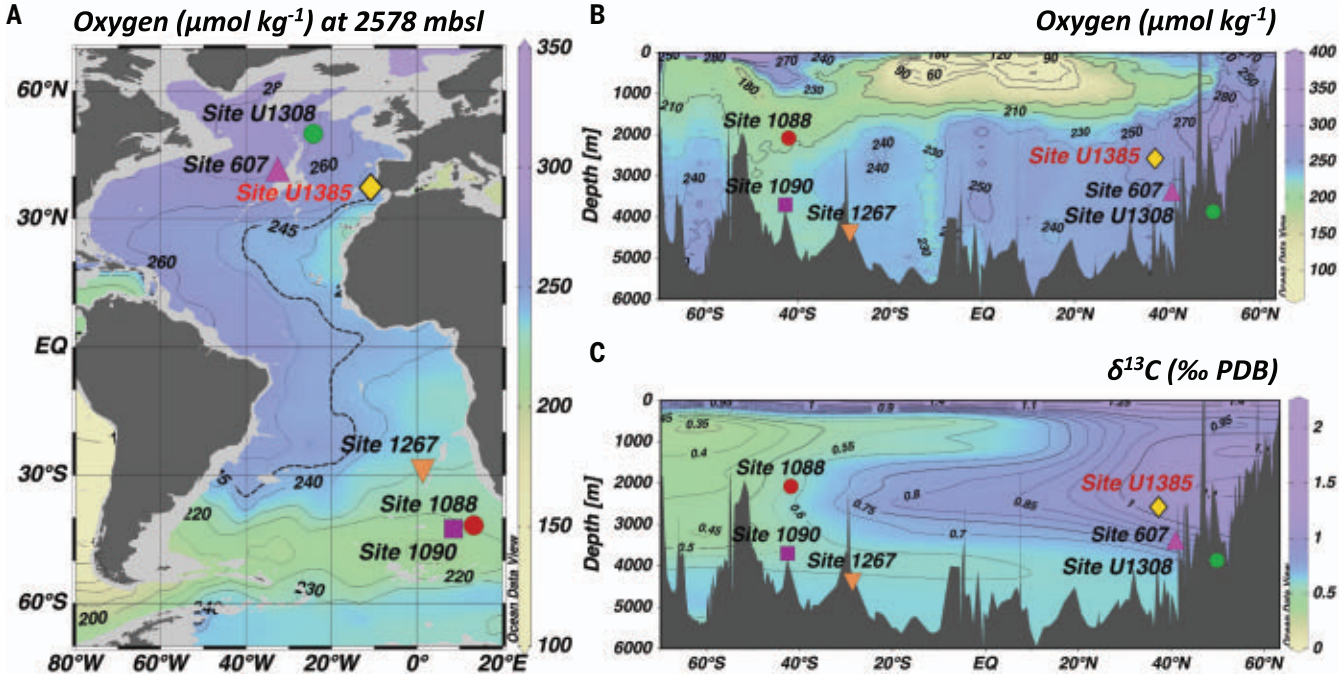


Fig. 1. Atlantic hydrographic profiles of [O_2] and $\delta^{13}C$ relative to site locations. (**A**) deep-water [O_2] profile at a depth of 2578 mbsl at IODP Site U1385 (this study; yellow diamond) in the Northeast Atlantic, along with the geographic position of other sites referred to in the text: North Atlantic DSDP Site 607 (*16*, *47*) (pink triangle); IODP Site U1308 (*68*) (green circle); South Atlantic ODP Sites 1267 (*14*) (orange triangle), and Sites 1090 and 1088 (*46*) (purple square and red circle, respectively). Meridional cross-sections showing the hydrographic profiles of (**B**) [O_2] and (**C**) $\delta^{13}C$ of DIC relative to the core locations. Maps were created in Ocean Data View (Schlitzer, 2019; http://odv.awi.de).

that extends from 0 to 166.5 meters composite depth (mcd), which is equivalent to Marine Isotope Stage (MIS) 47 at ~1.45 Ma. Sedimentation rates average ~11 cm kyr^{-1} (*23*) which translates into a mean sampling frequency of ~290 years with no significant difference across the MPT. Today the water depth of 2578 mbsl places Site U1385 in well-ventilated North Atlantic Deep Water (NADW) that is made up of varying proportions of Classical Labrador Sea Water (CLSW; ~45%) (see table S1 for [O_2]), Iceland-Scotland Overflow Water (ISOW; ~30%), and Denmark Strait Overflow Water (DSOW; ~5%), with the remainder (~20%) derived from the modified Antarctic Bottom Water (AABW) (*29*). Present-day [O_2] at the site ranges between values of ~237 and 245 µmol kg^{-1} depending on [O_2] profiles and databases considered (Fig. 1, fig. S1, and table S1) and the $\delta^{13}C$ value of DIC averages ~1‰ (Fig. 1). During glacial stages the relative proportion of northern-sourced deep water decreased at Site U1385 as the fraction of southern-sourced deep water increased (*30*–*33*).

Hoogakker *et al.* (*21*) applied the $\Delta\delta^{13}C_{cib\text{-}aff}$ method to estimate [O_2] for the past 160 kyr using piston core MD95-2042 (3146 mbsl), which is slightly deeper than nearby Site U1385 but has very similar modern-day oxygen concentrations (~245 µmol kg^{-1}). They concluded that deep-water [O_2] was lower during the last two glacial stages relative to today, with the lowest values (~100 µmol kg^{-1} lower) recorded during cold stadial conditions associated with Heinrich events. With Site U1385, we extend the [O_2] record of MD95-2042 beyond the last glacial cycle to span the past 1.5 Ma (Fig. 2, A to E). We use the calibration of Hoogakker *et al.* (*21*) to convert $\Delta\delta^{13}C_{cib\text{-}aff}$ to deep-water [O_2] for values <2.1‰, which correspond to [O_2] <235 µmol kg^{-1} in which a strong linear relationship exists: $\Delta\delta^{13}C_{cib\text{-}aff} = 0.0064 \times [O_2] + 0.555$; ($R^2 = 0.95$). Above 235 µmol kg^{-1} the relationship breaks down, so we truncate estimated [O_2] values at this concentration (Fig. 2E). For an alternative calibration, see supplementary materials (fig. S2) (*25*).

We assess the reliability of the $\Delta\delta^{13}C_{cib\text{-}aff}$ proxy using a transect of core-top sediment recovered from the Iberian Margin during Cruise JC089 (*22*) (fig. S1A). Core top $\Delta\delta^{13}C_{cib\text{-}aff}$ values demonstrate that the shape generally follows water column [O_2] with low values associated with Mediterranean Outflow Water (MOW) and higher values in NADW below (fig. S1B; supplementary materials). Because only the MOW has [O_2] <235 µmol kg^{-1} we could only test the calibration on two core tops under the influence of the MOW. The difference between CTD-measured [O_2] and predicted [O_2] by using measured $\Delta\delta^{13}C_{cib\text{-}aff}$ was 0 and 17 µmol kg^{-1} for the deep (~1100 m) and shallow (~600 m) MOW core tops, respectively, which is within the estimated error of the calibration (±17 µmol kg^{-1}) (*21*) (see supplementary materials).

From ~1435 to 960 ka during the early Pleistocene (MIS 47 to 27), deep-water [O_2] was relatively high and only once fell below ~180 µmol kg^{-1} (~1120 ka) (Fig. 2E and fig. S3). MIS 26 (~960 ka) marks the first time in the record when glacial [O_2] dropped to ~140 µmol kg^{-1}. Lower values were attained over the next three glacials (MIS 24 to 20) and throughout subsequent terminations before dropping to the lowest [O_2] value (~95 µmol kg^{-1}) during MIS 18 (~750 ka; Fig. 2E). The interval from ~960 to 750 ka (MIS 26 to 18) is marked by notable, persistent decreases in $\Delta\delta^{13}C_{cib\text{-}aff}$ (Fig. 2D), particularly throughout MIS 22 (~890 ka) when deep-water [O_2] exhibits a greater number of low-[O_2] events compared with those in the early Pleistocene. Over the MIS 27 to 21 (980 to 840 ka) interval, often referred to as the "900-ka event" (*10*) or erroneously as the first 100-kyr cycle (*34*), low-[O_2] values are supported by other recorders of paleo-[O_2]. High U/Ca and U/Mn (figs. S4 and S5B) (*35*, *36*) in foraminifera coatings measured in the same Site U1385 samples as those used for the $\Delta\delta^{13}C_{cib\text{-}aff}$ measurements (supplementary materials) support a drawdown in glacial deep-water [O_2]. Furthermore, qualitative indicators of low organic carbon flux during periods of low [O_2] provide confidence in reconstructed depleted [O_2] (*26*) events

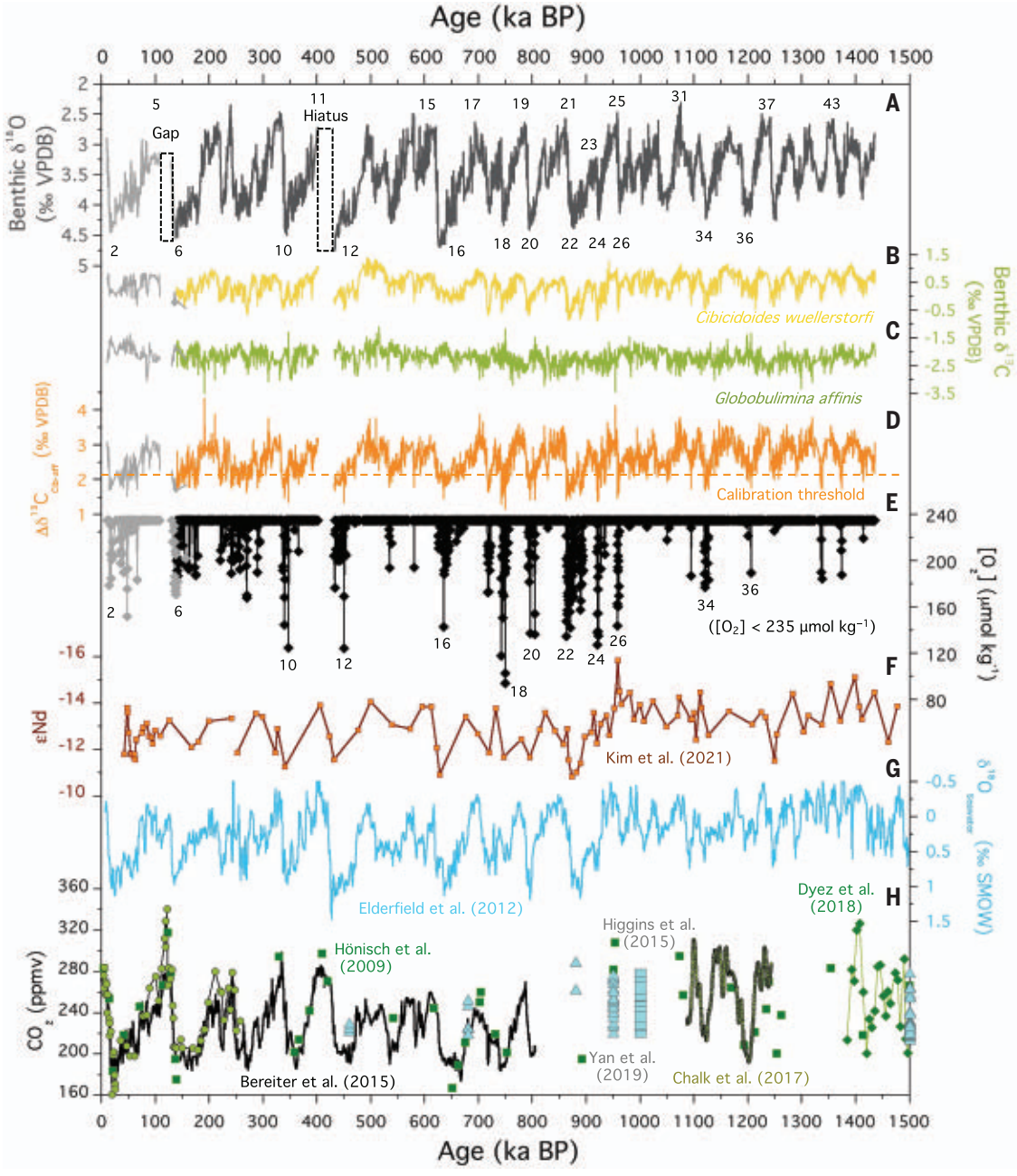


Fig. 2. North Atlantic IODP Site U1385 deep water records (this study) compared with Atlantic meridional overturning circulation, ice volume, and pCO_2 estimates across the MPT. Site U1385 records extend the record of piston core MD95-2042 (*21*) (light gray). (**A**) *Cibicidoides wuellerstorfi* $\delta^{18}O$ (dark gray); MIS numbers denote interglacials (odd) and glacials (even); black dashed boxes highlight a ~30-kyr hiatus and gap. (**B** and **C**) Benthic foraminiferal $\delta^{13}C$: *C. wuellerstorfi* (yellow) and *G. affinis* (green). (**D**) $\Delta\delta^{13}C_{cib\text{-}aff}$ (orange); the dashed line at 2.1‰ (orange) marks the calibration threshold. (**E**) $[O_2]$ (black line with diamonds), truncated at 235 µmol kg^{-1}, above which the calibration is unreliable; MIS numbers denote important glacials. (**F**) DSDP Site 607/V30-97 ε_{Nd} proxy for North Atlantic Ocean circulation changes (brown line with orange squares) (*47*). (**G**) ODP Site 1123 ice volume $\delta^{18}O_{seawater}$ (blue) (*62*). (**H**) Atmospheric pCO_2 records: Antarctic 800-kyr ice core data (*1*) (black line); Allan Hills Blue Ice (*6*, *7*) (blue triangles and squares, respectively); and $\delta^{11}B$-based reconstructed pCO_2 [green line with (*5*), solid green squares (*3*), early Pleistocene light green line and small circles, and late Pleistocene light green circles (*4*)]. Chronology is available in the supplementary materials, materials and methods.

at this time (fig. S5, C to E; supplementary materials).

Between ~620 and 470 ka (MIS 15 to early MIS 12) deep-water $[O_2]$ decreases were weaker and less frequent than those during the preceding period (MIS 26 to 16) and were similar to those observed during the early Pleistocene (Fig. 2E). MIS 14 was a weak glacial period which may explain why oxygen remained relatively high at this time. From ~470 to 10 ka (late MIS 12 to the early Holocene), the magnitude of deep-water $[O_2]$ depletions intensify supported by the alcohol preservation proxy for oxygenation $[C_{26}OH/(C_{26}OH + C_{29})]$ (*33*) (fig. S6) but remain higher than those centered on 900 ka.

The relationship between changes in deep-water $[O_2]$ and atmospheric pCO_2 arises from the biological carbon pump through the production and consumption of organic matter (*18*, *37*–*39*). Whereas many marine processes are involved in pCO_2 variations on glacial-interglacial time scales [see (*40*) for a review], changes in the efficiency of the ocean's soft-tissue pump directly affect dissolved deep-water oxygen through the regeneration of respired carbon. Weaker overturning circulation during glacial stages leads to a more efficient biological pump (i.e., an increase in the ratio of regenerated versus preformed nutrient content in the interior ocean) (*18*) by stemming the ventilation "leak" of CO_2 from the ocean to the atmosphere and increasing its oceanic residence time (*41*). At the same time a stronger biological pump (i.e., an increased rate of organic carbon exported through greater productivity in the surface ocean) (*42*) will result in greater organic carbon regeneration in the ocean's interior, enhancing deep-sea respired carbon storage with a corresponding decrease in oxygen. Widespread reduction in deep-sea $[O_2]$ (without compensatory change at intermediate water depths)

(*43*) implies increased carbon storage and decreased atmospheric pCO_2 (*18*, *38*, *44*). Reconstruction of deep-water paleo-[O_2] is therefore a valuable tool for estimating carbon transfers between the atmosphere and abyssal ocean, as well as for estimating glacial-interglacial changes in respired carbon storage and atmospheric pCO_2 (*21*, *26*).

We compare estimated changes in [O_2] from Site U1385 with other deep Atlantic paleoceanographic records to determine how widespread the observed oxygen depletions were. Comparison with the U/Ca record from the Deep Sea Drilling Project (DSDP) Site 607—which is bathed in a water mass similar to that of Site U1385 (Figs. 1 and 3D)—reveals strong correlation between low-[O_2] events at Site U1385 and increases in U/Ca at Site 607 (*16*), suggesting that intervals of reduced oxygenation were widespread in the deep North Atlantic. Tightly coupled changes in U/Ca and benthic $\delta^{13}C$ at Site 607 (fig. S7) suggest a link to changes in deep-water circulation (*16*), with both proxies showing a change at MIS 26 (~970 ka).

Decreases in deep-water [O_2] values are expected to mirror increases in deep-water [PO_4^{3-}] according to stoichiometric ratios involved in organic matter respiration (*38*, *45*). ODP Site 1267 in the Southeast Atlantic (Fig. 1) provides evidence of an increase of 0.5 μmol kg^{-1} in [PO_4^{3-}] at ~960 ka (Fig. 3E) (*14*) coeval with decreases in [O_2] and decreasing P/Al ratios at Site U1308 in the North Atlantic (figs. S8 and S9B; see supplementary text section 5). Both proxy-derived [PO_4^{3-}] at Site 1267 and [O_2] at Site U1385 show statistically significant changes across the MPT (figs. S3 and S9A; see table S2 for statistics), suggesting a greater nutrient content and respired carbon pool in the deep Atlantic Ocean after 960 ka.

The change in nutrient content of the deep South Atlantic also corresponds to a major change in thermohaline circulation (THC) inferred from changes in neodymium isotopes (*14*, *46*, *47*). Neodymium isotope (ε_{Nd}) records from North Atlantic DSDP Site 607/V30-97 (Figs. 1 and 2F) (*47*) and ODP Sites 1267 (*14*), 1088, and 1090 (*46*) in the Southeast Atlantic (Figs. 1 and 3, F and G) document significant reduction in NADW contribution and/or an increased influence of AABW (*14*) between 950 and 900 ka (*14*, *46*). These THC changes occurred at times of reduced upwelling and degassing of southern-sourced deep water under expanded sea ice cover in the Southern Ocean, thereby contributing to decreasing benthic foraminifer $\delta^{13}C$ (*16*).

Changes in deep-water circulation can affect [O_2] in several ways: Preformed [O_2] may decrease at sites of deep water formation if equilibration of oxygen with the atmosphere is reduced under sea ice leading to undersaturation in preformed oxygen (*48*, *49*). An increase in deep water residence time results in decreased oxygen concentrations through organic matter oxidation with or without attendant changes in carbon flux from the surface (*41*, *50*). Lastly, a change in the proportion of northern- versus southern-derived deep water masses (*16*, *51*, *52*) bathing Site U1385 can affect [O_2]; for example, NADW has [O_2] values that are ~50 μmol kg^{-1} higher than those of AABW

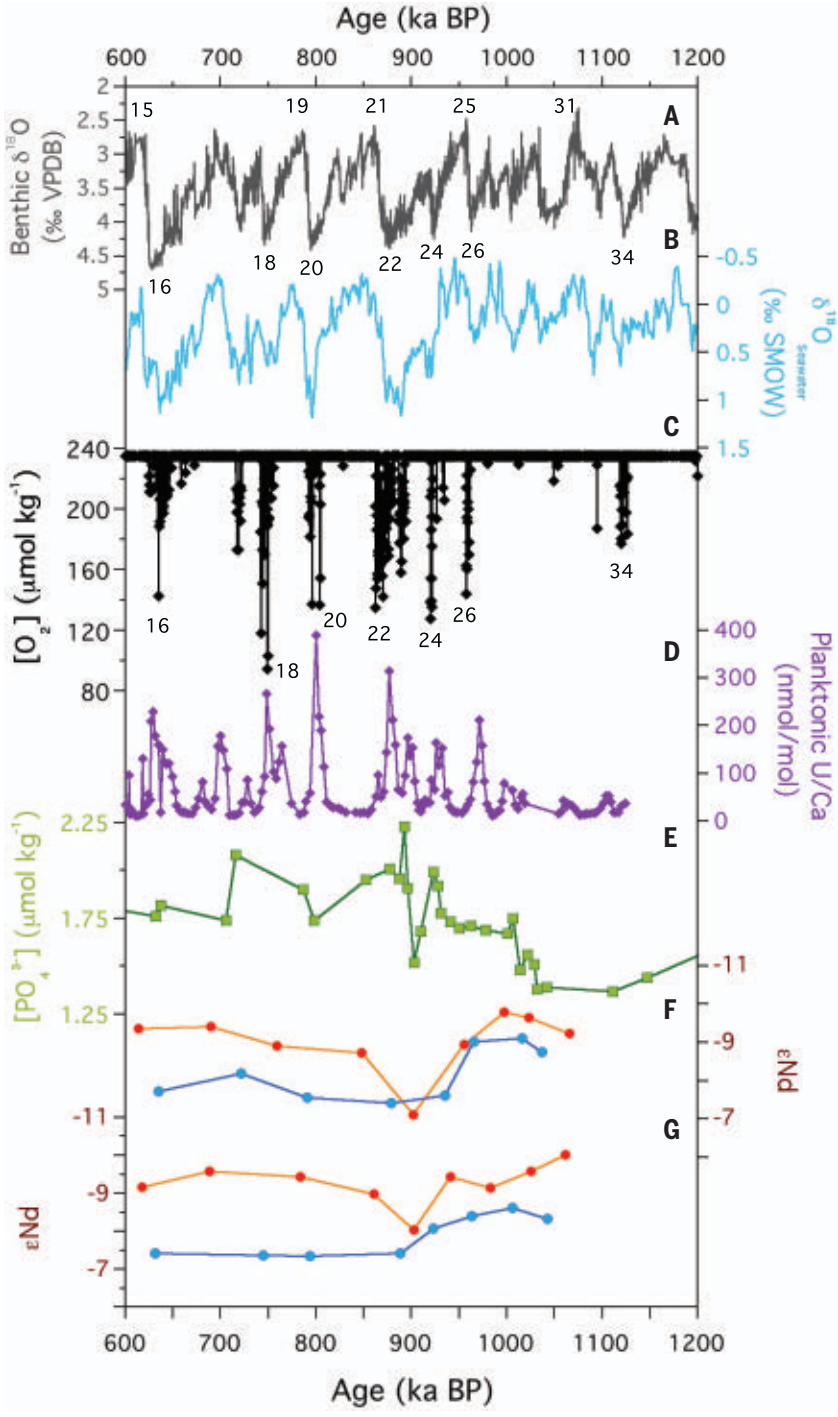


Fig. 3. Comparison of Atlantic Ocean deep-water oxygenation, nutrient and circulation proxies, and South Pacific ice volume changes through the MPT. (**A**) IODP Site U1385 *Cibicidoides wuellerstorfi* $\delta^{18}O$ (dark gray, this study); MIS numbers denote interglacials (odd) and glacials (even). (**B**) ODP Site 1123 ice volume $\delta^{18}O_{seawater}$ (blue) (*62*). (**C**) [O_2] estimated from $\Delta\delta^{13}C_{cib\text{-}aff}$ (this study; black diamonds). MIS numbers denote important glacials. (**D**) DSDP Site 607 U/Ca redox proxy (purple diamonds) (*16*). (**E**) ODP Site 1267 Cd/Ca-derived [PO_4^{3-}] (green squares) (*14*). (**F** and **G**) ODP Sites 1088 (mid-depth) and 1090 (deep) ε_{Nd}, respectively (*46*); interglacial (orange circles) and glacial (blue circles) maxima.

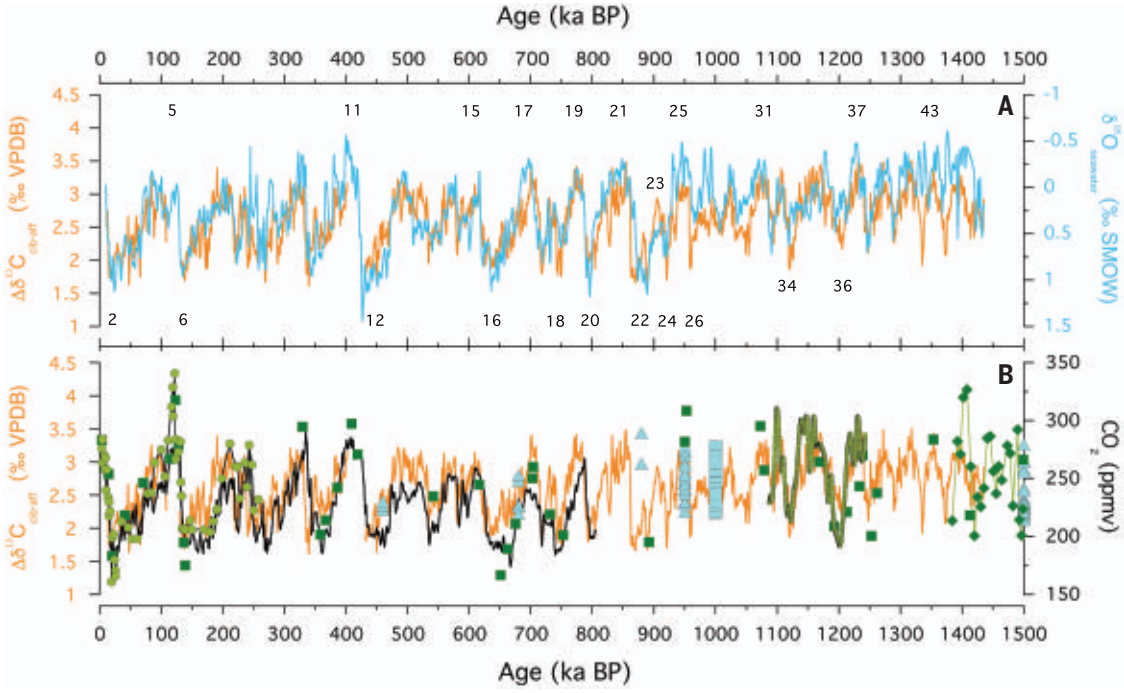


Fig. 4. Comparison of deep-water oxygen, atmospheric pCO_2, and ice volume proxies. (**A**) IODP Site U1385 $\Delta\delta^{13}C_{cib\text{-}aff}$ including piston core MD95-2042 data (*21*) (Gaussian interpolated with a 500-year interval, orange line); the oxygen calibration is only valid for $\Delta\delta^{13}C_{cib\text{-}aff}$ <2.1‰; ODP Site 1123 ice volume $\delta^{18}O_{seawater}$ (*62*) (Gaussian interpolated with a 500-year interval, blue line); MIS numbers denote major interglacials and glacials. (**B**) IODP Site U1385 $\Delta\delta^{13}C_{cib\text{-}aff}$ [as in (A)]; Atmospheric pCO_2 records: Antarctic 800-kyr ice core data (*1*) (Gaussian interpolated with a 500-year interval, black line); Allan Hills Blue Ice (*6*, *7*) (blue triangles and squares, respectively); and $\delta^{11}B$-based reconstructed pCO_2 [green line with diamonds (*5*), solid green squares (*3*), early Pleistocene light green line and small circles, and late Pleistocene light green circles (*4*)].

and 75 μmol kg^{-1} higher than Circumpolar Deep Water (Fig. 1 and table S1) (*53*).

Changes in source areas of NADW formation could substantially affect $[O_2]$ in the deep North Atlantic. CLSW is a well-ventilated water mass that distributes oxygen throughout the entire North Atlantic (*54*) including Site U1385 on the Iberian Margin that today is bathed by ~45% CLSW (*29*). Oxygen saturation of CLSW is highly sensitive to winter conditions in the Labrador Sea through bubble-mediated air-sea transfer associated with intensive winds, cooling, and deep convection (*55*). The Labrador Sea has been described as a "trapdoor" through which the flux of oxygen ventilates much of the deep Atlantic basin (*55*). During Heinrich stadials of the last glaciation, the Labrador Sea was covered by extensive sea ice which would have reduced the ventilation of the deep Atlantic by CLSW (*21*, *56*). Sea ice also expanded over the Nordic Seas (*57*–*59*) (source areas of DSOW and ISOW) (*53*) and the Irminger Sea where these water masses mix to form NADW (*53*, *60*, *61*).

The decrease in glacial deep-water $[O_2]$ at ~960 ka coincides with an increase in continental ice volume and lowered sea level, inferred from changes in $\delta^{18}O$ of seawater at Site 1123 in the Southwest Pacific (*51*, *62*, *63*) (Figs. 2G, 3B, and 4A). Available evidence suggests Northern Hemisphere glaciation intensified during MIS 22 [(*51*) and references therein], which resulted in changes to Atlantic deep-water circulation (*52*).

The timing of the initial decrease in deep-water $[O_2]$ at ~960 ka during MIS 26 is associated with a short-lived excursion to very negative ε_{Nd} values at Site 607 (Fig. 2F) (*47*). This event is interpreted as a result of extensive weathering and erosion of the North American craton between MIS 27 to 25 (~980 to 950 ka) and the possible transition from terrestrial- to marine-terminating ice sheets in the Northern Hemisphere (*64*). MIS 26 also represents the first time the polar front shifted to a zonal position south of Site 980/981 in the Northeast Atlantic (55.5°N, 14.7°W) as inferred from an increase in the percent of the polar planktonic foraminifera *Neogloboquadrina pachyderma* (percent NPS), a proxy for cold surface temperatures (fig. S10) (*65*, *66*). These precursor changes heralded the major changes in ice volume, thermohaline circulation, and pCO_2 associated with the "900-ka event" during MIS 24 to 22 in less highly resolved records than those of Sites U1385 and 980/981 (*3*, *4*, *6*, *10*, *14*, *46*, *47*, *62*, *65*).

The deep-water oxygen depletions may have been driven by increased freshwater inputs and ice rafting to the source areas of NADW formation. The lowest $[O_2]$ values during the last glacial cycle are associated with cold millennial events in the North Atlantic (fig. S11B) (*21*). Millennial $[O_2]$ depletion events occur throughout the entire record and are most common toward the latter part of glacial cycles and during terminations. These times are also associated with enhanced sea ice extent in the North Atlantic (*67*), cold stadial periods, and increases in ice-rafted detritus (IRD) at both the North Atlantic Site U1308 (*68*) (Fig. 1 and figs. S11A and S12) and the Labrador Sea Site U1302/03 (fig. S13) (*69*). A close connection exists today between North Atlantic deep-water oxygenation and winter surface conditions in the Labrador, Irminger, and Nordic Seas (*55*–*58*). We suggest that enhanced surface stratification and reduced deep-water convection in NADW source areas may have caused episodic reduction of North Atlantic deep-water ventilation, leading to reduced deep-water oxygen and increased carbon storage (*70*). Incursions of glacial southern-sourced deep water into the North Atlantic associated with weaker NADW production (*51*, *52*, *61*, *71*, *72*) could also partly account for $[O_2]$ changes observed at Site U1385 over the past 1.5 Ma given the $[O_2]$ difference (~50 μmol kg^{-1}) between northern and southern-sourced deep water today (Fig. 1 and table S1).

The changes in deep-water $[O_2]$ at Site U1385 could equally be driven by surface processes in the Southern Ocean such as changes in productivity (*13*), surface stratification (*15*, *73*), vertical mixing, and sea ice extent. These changes would be transmitted to the deep sea through an expanding southern-sourced deep water mass such as Lower Circumpolar Deep Water (*74*) in which oxygen was considerably reduced during the last glacial period (*26*, *36*, *44*). Reduction in ventilation due to circulation changes, Southern Ocean stratification, and sea ice expansion would have contributed to the inferred increase

in deep-sea carbon storage over the MPT (*14*–*16*, *18*, *73*, *75*–*78*).

Decreased glacial deep-water [O_2] and increased deep-sea carbon storage across the MPT have implications for atmospheric pCO_2. The $\Delta\delta^{13}C_{cib\text{-}aff}$ proxy of [O_2] at Site U1385 is consistent with changes in pCO_2 measured directly in ice cores and blue ice in Antarctica (*1*, *6*, *7*) and inferred indirectly from boron isotope data (*3*–*5*) (Fig. 4B). Prior to ~900 ka, minimum glacial pCO_2 values were ~24 ppm higher than they were afterward (*3*, *5*–*7*). If applied to the whole Atlantic Ocean influenced by an expanding southern-sourced deep water mass, Hoogakker *et al.* (*21*) estimated at least 15% of the reduction in atmospheric pCO_2 during the last glacial maximum (LGM) could be accounted for by the increases observed in their respired carbon pool. Furthermore, over the same period, the Pacific Ocean is reported to have had old, oxygen-depleted deep-water (*26*, *43*, *44*). The reconstructed glacial [O_2] values reported here for the MPT are at least 20 μmol kg^{-1} lower than the LGM (*21*), implying greater carbon storage than during the LGM. It remains to be seen from other basins whether this was a global phenomenon.

In summary, inferred changes in North Atlantic deep-water [O_2] for the past 1.5 Ma reveal a significant (>40 μmol kg^{-1}) reduction in glacial deep-water [O_2] at ~900 ka suggesting increased storage of respired carbon, which is consistent with a drawdown of glacial atmospheric pCO_2 values (Fig. 4B) (*1*, *3*, *4*, *6*, *7*). The inferred change in [O_2] is supported by trace metal records (*16*) and nutrient (*14*) proxy records in other Atlantic sites associated with a critical change in glacial THC at ~900 ka (*46*) (Fig. 3, C to G). The close association between [O_2] depletions and IRD events (*68*, *69*, *79*) suggests that increased stratification and sea ice cover in NADW source regions (*56*–*59*) reduced the oxygen supply to much of the deep North Atlantic (*55*). In addition, northward expansion of southern-sourced deep water into the North Atlantic and processes in the Southern Ocean (e.g., productivity, surface water stratification, vertical mixing, and sea ice extent) also contributed to reduced ventilation associated with a major change in deep-water circulation (*46*). Our results support a set of internally consistent changes in Atlantic deep water beginning at ~960 ka across the MPT, which included a decrease in oxygen concentrations, increased nutrient concentrations (*14*), and storage of respired carbon that led to a reduction in glacial pCO_2 (*3*–*7*, *47*) and an associated increase in global ice volume (*62*).

REFERENCES AND NOTES

1. B. Bereiter *et al.*, *Geophys. Res. Lett.* **42**, 542–549 (2015).
2. D. Lüthi *et al.*, *Nature* **453**, 379–382 (2008).
3. B. Hönisch, N. G. Hemming, D. Archer, M. Siddall, J. F. McManus, *Science* **324**, 1551–1554 (2009).
4. T. B. Chalk *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 13114–13119 (2017).
5. K. A. Dyez, B. Hönisch, G. A. Schmidt, *Paleoceanogr. Paleoclimatol.* **33**, 1270–1291 (2018).
6. Y. Yan *et al.*, *Nature* **574**, 663–666 (2019).
7. J. A. Higgins *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6887–6891 (2015).
8. N. J. Shackleton, *Science* **289**, 1897–1902 (2000).
9. A. Berger, X. S. Li, M. F. Loutre, *Quat. Sci. Rev.* **18**, 1–11 (1999).
10. P. U. Clark *et al.*, *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
11. J. D. Hays, J. Imbrie, N. J. Shackleton, *Science* **194**, 1121–1132 (1976).
12. S. L. Jaccard *et al.*, *Science* **339**, 1419–1423 (2013).
13. A. Martínez-Garcia *et al.*, *Nature* **476**, 312–315 (2011).
14. J. R. Farmer *et al.*, *Nat. Geosci.* **12**, 355–360 (2019).
15. A. P. Hasenfratz *et al.*, *Science* **363**, 1080–1084 (2019).
16. C. H. Lear *et al.*, *Geology* **44**, 1035–1038 (2016).
17. L. Menviel, *Science* **363**, 1040–1041 (2019).
18. D. M. Sigman, M. P. Hain, G. H. Haug, *Nature* **466**, 47–55 (2010).
19. D. C. McCorkle, S. R. Emerson, P. D. Quay, *Earth Planet. Sci. Lett.* **74**, 13–26 (1985).
20. D. C. McCorkle, S. R. Emerson, *Geochim. Cosmochim. Acta* **52**, 1169–1178 (1988).
21. B. A. A. Hoogakker, H. Elderfield, G. Schmiedl, I. N. McCave, R. E. M. Rickaby, *Nat. Geosci.* **8**, 40–43 (2015).
22. D. A. Hodell, H. Elderfield, M. Greaves, I. N. McCave, L. Skinner, A. Thomas, N. White, The JC089 scientific party, JC089 cruise report: IODP site survey of the Shackleton sites, SW Iberian Margin (British Oceanographic Data Centre, 2014); https://www.bodc.ac.uk/resources/inventories/cruise_inventory/report/13392/.
23. D. Hodell *et al.*, *Global Planet. Change* **133**, 49–64 (2015).
24. D. C. McCorkle, L. D. Keigwin, B. H. Corliss, S. R. Emerson, *Paleoceanography* **5**, 161–185 (1990).
25. G. Schmiedl, A. Mackensen, *Paleoceanography* **21**, PA4213 (2006).
26. A. W. Jacobel *et al.*, *Quat. Sci. Rev.* **230**, 106065 (2020).
27. A. V. Turchyn, G. Antler, D. Byrne, M. Miller, D. A. Hodell, *Global Planet. Change* **141**, 82–90 (2016).
28. H. J. Bradbury *et al.*, *Front. Earth Sci.* **9**, 652960 (2021).
29. W. J. Jenkins, W. M. Smethie Jr., E. A. Boyle, G. A. Cutter, *Deep Sea Res. Part II Top. Stud. Oceanogr.* **116**, 6–20 (2015).
30. N. J. Shackleton, M. A. Hall, E. Vincent, *Paleoceanography* **15**, 565–569 (2000).
31. L. C. Skinner, H. Elderfield, M. Hall, *Geophys. Monogr.* **173**, 197–208 (2007).
32. V. Margari *et al.*, *Nat. Geosci.* **3**, 127–131 (2010).
33. B. Martrat *et al.*, *Science* **317**, 502–507 (2007).
34. P. Bajo *et al.*, *Science* **367**, 1235–1239 (2020).
35. R. Boiteau, M. Greaves, H. Elderfield, *Paleoceanography* **27**, PA002335 (2012).
36. J. Gottschalk *et al.*, *Nat. Commun.* **7**, 11539 (2016).
37. S. L. Jaccard, E. D. Galbraith, A. Martínez-García, R. F. Anderson, *Nature* **530**, 207–210 (2016).
38. W. S. Broecker, *Prog. Oceanogr.* **11**, 151–197 (1982).
39. S. L. Jaccard, E. D. Galbraith, *Nat. Geosci.* **5**, 151–156 (2012).
40. D. M. Sigman, E. A. Boyle, *Nature* **407**, 859–869 (2000).
41. L. C. Skinner *et al.*, *Nat. Commun.* **8**, 16010 (2017).
42. M. P. Hain, D. M. Sigman, G. H. Haug, in *Treatise on Geochemistry: Second Edition*, H. D. Holland, K. K. Turekian, Eds. (Elsevier, 2014), chap. 8.18, 485–517.
43. B. A. A. Hoogakker *et al.*, *Nature* **562**, 410–413 (2018).
44. R. F. Anderson *et al.*, *Global Biogeochem. Cycles* **33**, 301–317 (2019).
45. L. A. Anderson, J. L. Sarmiento, *Global Biogeochem. Cycles* **8**, 65–80 (1994).
46. L. D. Pena, S. L. Goldstein, *Science* **345**, 318–322 (2014).
47. J. Kim *et al.*, *Quat. Sci. Rev.* **269**, 107146 (2021).
48. T. Ito, M. J. Follows, *J. Mar. Res.* **63**, 813–839 (2005).
49. E. Cliff, S. Khatiwala, A. Schmittner, *Nat. Geosci.* **14**, 43–50 (2021).
50. L. C. Skinner, C. Waelbroeck, A. E. Scrivner, S. J. Fallon, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5480–5484 (2014).
51. D. A. Hodell, J. E. T. Channell, *Clim. Past* **12**, 1805–1828 (2016).
52. D. C. Lang *et al.*, *Nat. Geosci.* **9**, 375–379 (2016).
53. M. Liu, T. Tanhua, *Ocean Sci.* **17**, 463–486 (2021).
54. I. Stendardo, N. Gruber, *J. Geophys. Res.* **117** (C11), C11004 (2012).
55. D. Atamanchuk, J. Koelling, U. Send, D. W. R. Wallace, *Nat. Geosci.* **13**, 232–237 (2020).
56. A. De Vernal, C. Hillaire-Marcel, *Quat. Sci. Rev.* **19**, 65–85 (2000).
57. T. M. Dokken, K. H. Nisancioglu, C. Li, D. S. Battisti, C. Kissel, *Paleoceanography* **28**, 491–502 (2013).
58. H. Sadatzki *et al.*, *Sci. Adv.* **5**, eaau6174 (2019).
59. U. Hoff, T. L. Rasmussen, R. Stein, M. M. Ezat, K. Fahl, *Nat. Commun.* **7**, 12247 (2016).
60. F. Fröb *et al.*, *Nat. Commun.* **7**, 13244 (2016).
61. B. P. Flower *et al.*, *Paleoceanography* **15**, 388–403 (2000).
62. H. Elderfield *et al.*, *Science* **337**, 704–709 (2012).
63. H. L. Ford, M. E. Raymo, *Geology* **48**, 113–117 (2020).
64. M. Yehudai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, 1–11 (2021).
65. S. Barker *et al.*, *Paleoceanogr. Paleoclimatol.* **36**, 1–23 (2021).
66. S. Barker *et al.*, *Nature* **520**, 333–336 (2015).
67. E. L. McClymont, A. Rosell-Melé, G. H. Haug, J. M. Lloyd, *Paleoceanography* **23**, PA001622 (2008).
68. D. A. Hodell, J. E. T. Channell, J. H. Curtis, O. E. Romero, U. Röhl, *Paleoceanography* **23**, PA001591 (2008).
69. J. E. T. Channell *et al.*, *Earth Planet. Sci. Lett.* **317-318**, 218–230 (2012).
70. F. Pöppelmeier *et al.*, *Geology* **48**, 826–829 (2020).
71. M. E. Raymo, W. F. Ruddiman, N. J. Shackleton, D. W. Oppo, *Earth Planet. Sci. Lett.* **97**, 353–368 (1990).
72. M. E. Raymo, D. W. Oppo, W. Curry, *Paleoceanography* **12**, 546–559 (1997).
73. K. Billups, K. York, L. I. Bradtmiller, *Paleoceanogr. Paleoclimatol.* **33**, 432–442 (2018).
74. L. C. Skinner, *Clim. Past* **5**, 537–550 (2009).
75. D. A. Hodell, K. A. Venz, C. D. Charles, U. S. Ninnemann, *Geochem. Geophys. Geosyst.* **4**, GC000367 (2003).
76. H. Gildor, E. Tziperman, *Philos. Trans. R. Soc. A* **361**, 1935–1942 (2003).
77. H. Gildor, E. Tziperman, *Geophys. Res. Lett.* **28**, 2421–2424 (2001).
78. H. Gildor, E. Tziperman, J. R. Toggweiler, *Global Biogeochem. Cycles* **16**, GB001446 (2002).
79. J. E. T. Channell, D. A. Hodell, S. J. Crowhurst, L. C. Skinner, R. Muscheler, *Quat. Sci. Rev.* **191**, 57–72 (2018).
80. N.C. Thomas, H. J. Bradbury, D. A. Hodell, Reconstructed paleo-oxygen concentration, stable isotope measurements and Gaussian interpolated $\Delta\delta_{13}C$ obtained from the IODP Site 339-U1385 for the past 1.5 Ma, including trace element analyses and Uvigerina spp. abundance over MIS 27-21. PANGAEA (2022); https://doi.org/10.1594/PANGAEA.945872

ACKNOWLEDGMENTS

We thank members of the Godwin laboratory for Paleoclimate Research in the Department of Earth Science, University of Cambridge, UK for technical assistance; J. Booth, M. Vautravers, J. Nicolson, and J. Rolfe for stable isotope analysis; S. Crowhurst for support with age modelling, statistical analysis, and helpful discussions as well as contributing to the collection of the XRF data; and M. Greaves for technical support and guidance in trace element analyses. We thank all those who helped on board the R/V *James Cook* during cruise JC089 to recover sediments, porewaters, and water column measurements used to assess the proxy data. This research used samples provided by the International Ocean Discovery Program (IODP). We thank the drilling crew, ship's crew, and scientific and technical staff of the *JOIDES Resolution* during IODP Expedition 339 without whom core recovery from Site U1385 would not have been possible. **Funding:** The Natural Environment Research Council (NERC) provided funding for all laboratory analytical costs and funded D.A.H. to collect the samples for both IODP and JC089; NERC grants NE/J00653X/1, NE/K005804/1, and NE/J017922/1 (to D.A.H.). N.C.T. has been privately funded by R.M. Thomas. NERC grant NE/R013519/1 was given to H.J.B. **Author contributions:** D.A.H. conceived the study and collected the samples. N.C.T. and D.A.H. conducted the stable isotopic analyses and N.C.T. performed trace element analyses. H.J.B. contributed to and wrote significant sections of the methodology in the revised version of the manuscript. N.C.T. and D.A.H. wrote the original version and both authors significantly contributed to the final manuscript. Conceptualization: D.A.H. Methodology: N.C.T., H.J.B., and D.A.H. Investigation: N.C.T. Visualization: N.C.T. and D.A.H. Funding acquisition: D.A.H. Project administration: D.A.H. and N.C.T. Writing – original draft: N.C.T. and D.A.H. Writing – review and editing: N.C.T., H.J.B., and D.A.H. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data not previously reported are available online at the World Data Center PANGAEA repository (www.pangaea.de) (*80*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj7761
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (*81*–*92*)

Submitted 1 June 2021; accepted 6 July 2022
10.1126/science.abj7761

IMMUNOLOGY

A conserved Bacteroidetes antigen induces anti-inflammatory intestinal T lymphocytes

Djenet Bousbaine[1,2,3]†, Laura I. Fisch[2]‡, Mariya London[4]‡, Preksha Bhagchandani[2,3], Tiago B. Rezende de Castro[4,5], Mark Mimee[3,6,7], Scott Olesen[3,8], Bernardo S. Reis[4], David VanInsberghe[1,9], Juliana Bortolatto[5], Mathilde Poyet[3,8], Ross W. Cheloha[2], John Sidney[10], Jingjing Ling[2], Aaron Gupta[4], Timothy K. Lu[3,6,7], Alessandro Sette[10,11], Eric J. Alm[3,8], James J. Moon[12], Gabriel D. Victora[5], Daniel Mucida[4,13], Hidde L. Ploegh[2,3]*, Angelina M. Bilate[3,4]*

The microbiome contributes to the development and maturation of the immune system. In response to commensal bacteria, intestinal CD4$^+$ T lymphocytes differentiate into functional subtypes with regulatory or effector functions. The development of small intestine intraepithelial lymphocytes that coexpress CD4 and CD8αα homodimers (CD4IELs) depends on the microbiota. However, the identity of the microbial antigens recognized by CD4$^+$ T cells that can differentiate into CD4IELs remains unknown. We identified β-hexosaminidase, a conserved enzyme across commensals of the Bacteroidetes phylum, as a driver of CD4IEL differentiation. In a mouse model of colitis, β-hexosaminidase–specific lymphocytes protected against intestinal inflammation. Thus, T cells of a single specificity can recognize a variety of abundant commensals and elicit a regulatory immune response at the intestinal mucosa.

The microbiota contributes to functional specification of adaptive immunity, both through direct interactions and through soluble mediators released into the environment (*1*–*3*). Colonic bacteria such as *Helicobacter hepaticus* promote differentiation of antigen-specific CD4$^+$ T cells into Foxp3$^+$ regulatory T cells (T_{reg} cells) in the colon, whereas segmented filamentous bacteria (SFB) induce quasi-clonal pro-inflammatory T helper 17 (T_H17) cells in the ileum (*4*–*10*). Such interactions are not only specific for the bacterial species concerned but, depending on location and context, can also influence T cell fates (*4*, *11*). Because fate decisions are made at the clonal level, different T cell receptor (TCR) specificities ought to drive distinct developmental and functional outcomes.

[1]Microbiology Graduate Program, Massachussetts Institute of Technology (MIT), Cambridge, MA, USA. [2]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA, USA. [3]Center for Microbiome Informatics and Therapeutics, MIT, Cambridge, MA, USA. [4]Laboratory of Mucosal Immunology, The Rockefeller University, New York, NY, USA. [5]Laboratory of Lymphocyte Dynamics, The Rockefeller University, New York, NY, USA. [6]Synthetic Biology Center, MIT, Cambridge, MA, USA. [7]Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA. [8]Department of Biological Engineering, MIT, Cambridge, MA, USA. [9]Department of Civil and Environmental Engineering, MIT, Cambridge, MA, USA. [10]Division of Vaccine Discovery, La Jolla Institute for Immunology, La Jolla, CA, USA. [11]Department of Medicine, University of California, San Diego, CA, USA. [12]Center for Immunology and Inflammatory Diseases and Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [13]Howard Hughes Medical Institute, The Rockefeller University, New York NY, USA.
***Corresponding author. Email: abilate@mail.rockefeller.edu (A.M.B.); hidde.ploegh@childrens.harvard.edu (H.L.P.)**
†Present address: Department of Bioengineering, Department of Microbiology and Immunology, ChEM-H Institute, Stanford University, Stanford, CA 94305, USA.
‡These authors contributed equally to this work.

Two subsets of CD4$^+$ T lymphocytes are known to regulate adaptive immunity at the intestinal mucosa: peripherally induced regulatory T cells (pT_{reg} cells) and CD8αα-expressing intraepithelial lymphocytes (CD4IELs) (*12*–*16*). Depletion of CD4IELs aggravates intestinal inflammation in mice that lack expression of Foxp3 (*14*). CD4IELs and pT_{reg} cells thus cooperate in the regulation of local intestinal inflammation. In specific pathogen–free (SPF) mice, CD4IELs are present mostly in the small intestinal epithelium (SIE). Their abundance varies with age, diet, and microbiota (fig. S1, A and B, and data S1, A and B) (*17*, *18*). CD4IELs can develop in a microbiota-dependent manner, either from conventional CD4$^+$ T cells or from T_{reg} cell precursors (*11*, *14*). Germ-free (GF) mice have few if any CD4IELs (fig. S1C) (*14*, *19*). Mice treated with antibiotics show a similar decline in CD4IELs (fig. S1, D and E) (*14*, *19*), whereas fecal transplants from SPF mice restore the CD4IEL population in GF mice to varying degrees (fig. S1F) (*17*, *18*). The microbiota therefore contributes not only to the development but also to the maintenance of CD4IELs. However, the microbial antigens recognized by CD4IELs are unknown. Most IEL populations, including CD4IELs, display a restricted TCR repertoire (*17*, *20*, *21*). A limited array of antigens might thus suffice to shape this compartment. We used a microbiota-specific transnuclear (TN) monoclonal T cell model (*11*) to identify naturally occurring TCR ligands of CD4IELs in SPF mice (fig. S2A). T cells from the TN mouse carry a monoclonal TCR, cloned from a pT_{reg} cell obtained from a mesenteric lymph node (mLN) of a SPF mouse (*11*). Naïve monoclonal T cells isolated from TN mice populate not only the recipient's mLNs, where they differentiate into pT_{reg} cells, but also the small intestinal epithelium, where they develop into CD4IELs in a microbiota-dependent fashion (*11*).

We thus sought to identify the commensal member(s) recognized by TN T cells by selectively growing fecal bacteria derived from Taconic "excluded flora" (EF) mice, which carry a microbiota enriched for the antigen recognized by the TN TCR (*11*). TN T cells proliferated strongly in the presence of bacterial extracts derived from Bacteroides bile esculin (BBE) agar plates or aerotolerant bacteria recovered from Schaedler blood agar (SBA) plates. 16*S* ribosomal RNA (rRNA) sequencing showed enrichment for operational taxonomic units (OTUs) corresponding to the *Parabacteroides* genus in activating extracts (Fig. 1, A to C, and data S1C). TN T cells proliferated robustly in the presence of *Parabacteroides goldsteinii* extracts (Fig. 1D), the predominant member of the altered Schaedler flora (ASF) (*22*). This proliferation was dependent on antigen presentation by dendritic cells (fig. S2B). Neither the related *Parabacteroides distasonis* nor any other species tested induced proliferation of TN T cells in vitro (Fig. 1D and fig. S2C). Next, we transferred CD4$^+$ TN T cells into congenic recipients, which were then immunized with bacterial extracts. Proliferation of TN cells occurred in the draining lymph nodes in response to *P. goldsteinii* extract but not *P. distasonis* extract or phosphate-buffered saline (Fig. 1E). Thus, *P. goldsteinii*, a commensal abundant in Taconic mice but less so in Jackson (Jax) mice (fig. S3, A to C, and data S1, D and E), can engage the TN TCR in vivo. TN T cells failed to proliferate and differentiate into CD4IELs in *Rag1*$^{-/-}$ recipients treated with antibiotics or when colonized with unrelated bacteria (*Clostridium tertium*) but expanded and differentiated into CD4IELs in mice colonized with *P. goldsteinii* (Fig. 1F and fig. S3D). Thus, *P. goldsteinii* promotes the in vivo development of CD4IELs from naïve T cells carrying the TN TCR.

We next tested whether *P. goldsteinii* could induce CD4IELs in wild-type (WT) mice with a polyclonal T cell repertoire. SPF mice obtained from Jax, which naturally harbor few CD4IELs (fig. S1A), developed more CD4IELs upon colonization with *P. goldsteinii* (Fig. 1G). By contrast, colonization with the unrelated commensal SFB, also absent in Jax mice (*10*), did not boost CD4IEL frequencies in these mice (fig. S3, E and F). Monocolonization of GF animals with *P. goldsteinii* was not sufficient to induce CD4IELs, however (fig. S4A). One possible explanation is that monocolonized mice have low tissue expression of interferon-γ (IFNγ) and lack class II major histocompatibility complex (MHCII) products on intestinal epithelial cells (fig. S4, B to D), both of which are required for CD4IEL development (*20*, *23*).

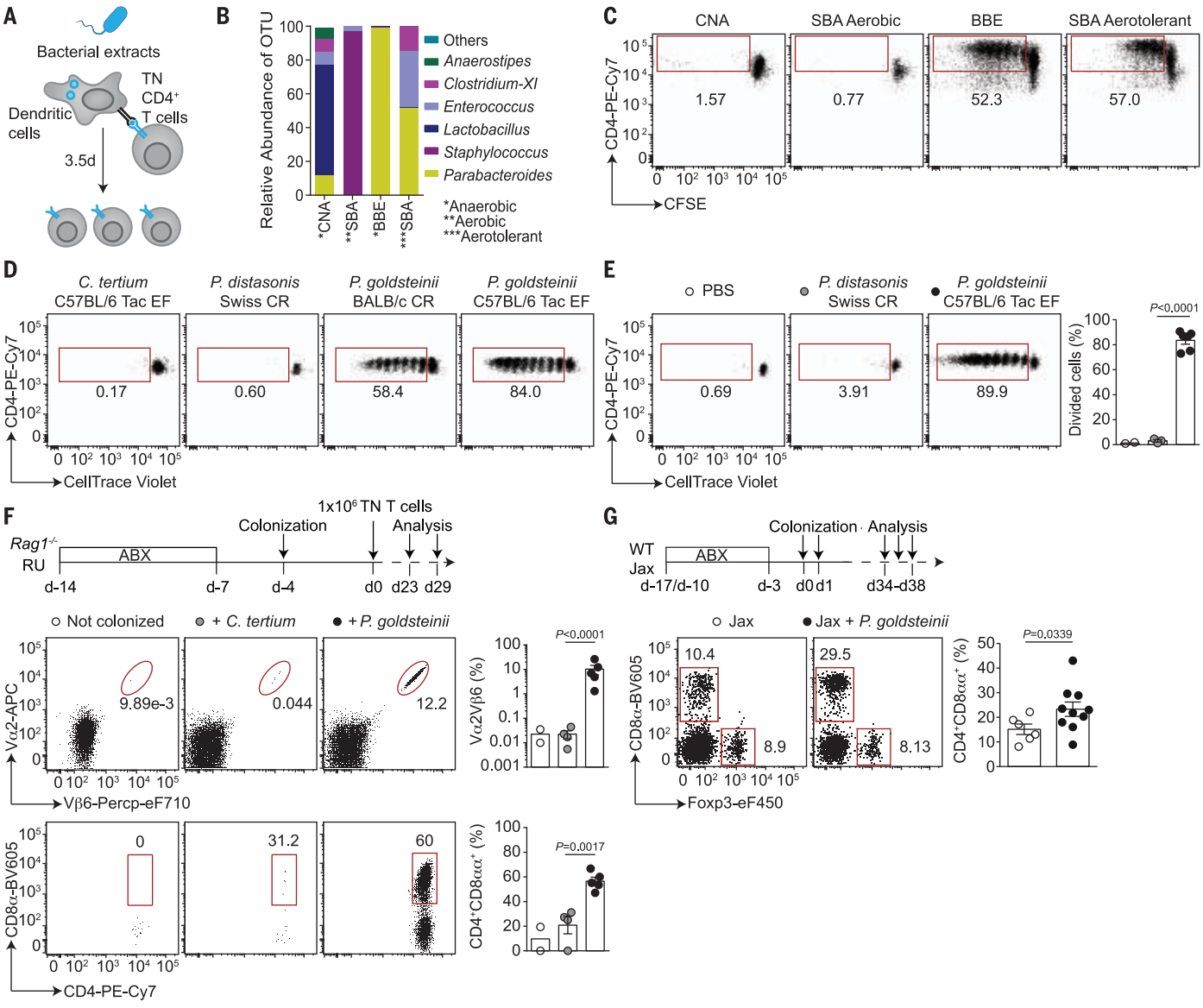


Fig. 1. *P. goldsteinii* induces CD4IELs in both monoclonal and polyclonal SPF mice. (**A**) Schematic of the in vitro proliferation assay. (**B**) 16S rRNA sequencing of isolated bacterial extracts used in (C). (**C**) Carboxyfluorescein diacetate succinimidyl ester (CFSE) dilution among TN cells in response to the indicated bacterial extracts derived from Taconic EF fecal bacteria, isolated by using the indicated growth conditions. (**D**) CellTrace Violet dilution among $CD4^+$ TN T cells in response to the indicated bacteria. (C) and (D) are representative of flow cytometry dot plots of three experiments. (**E**) Representative flow cytometry dot plots showing CellTrace Violet dilution (left) and frequency of divided cells (right) among $CD4^+$ TN T cells harvested from the draining inguinal LNs of mice immunized with the indicated bacterial extracts (n = 2 to 6 mice per group). (**F**) *Rag1*$^{-/-}$ hosts were treated with antibiotics (ABX) then either colonized or not colonized with the indicated bacteria before receiving $CD4^+$ TN T cells. Cells from the SIE were analyzed by means of flow cytometry (n = 2 to 5 mice per group). (Top) Experimental design. (Middle) Representative flow cytometry dot plots of Vα2 and Vβ6 expression among $CD45^+$ cells and CD4 and CD8α expression among TN cells ($V\alpha2^+V\beta6^+$) in SIE (bottom). (Right) Frequency of (top) TN cells among $CD45^+$ cells and (bottom) CD4IELs ($CD4^+CD8\alpha\alpha^+$) in the SIE among TN $CD4^+$ T cells in all mice analyzed in (F). (**G**) WT Jax mice were treated with ABX and colonized with *P. goldsteinii*. Cells from the SIE were analyzed by means of flow cytometry (n = 6 to 10 mice per group). (Top) Experimental design. (Bottom) Representative dot plots showing (left) Foxp3 and CD8α expression and (right) frequency of CD4IELs among $CD4^+$ T cells. The graphs in (F) and (G) show the means ± SEM, and each symbol indicates a mouse from two to three experiments. P values were calculated with unpaired two-tailed Student's t test in (E) and (F) and two-tailed Student's t test with Welch correction in (G). Tac, Taconic; CR, Charles River; SBA, Schaedler blood agar; BBE, Bacteroides bile esculin; CNA, colistin–nalidixic acid.

Thus, *P. goldsteinii* promotes the accumulation of CD4IELs from both TN and WT precursors in SPF mice.

We next fractionated *P. goldsteinii* lysates to identify potential TCR ligands derived from this species (fig. S5, A to D). This procedure yielded fractions that activated TN T cells in vitro (fig. S5E). Liquid chromatography with tandem mass spectrometry (LC-MS/MS) analysis of fraction 17 identified 33 proteins, the top 15 of which were expressed recombinantly in *Escherichia coli* (table S1). Their sonicates served as a source of candidate antigens tested in the proliferation assay. We identified β-*N*-acetylhexosaminidase (β-hex) as the protein recognized by the TN TCR (Fig. 2A). TN T cells proliferated in mice immunized with protein extract of *E. coli*

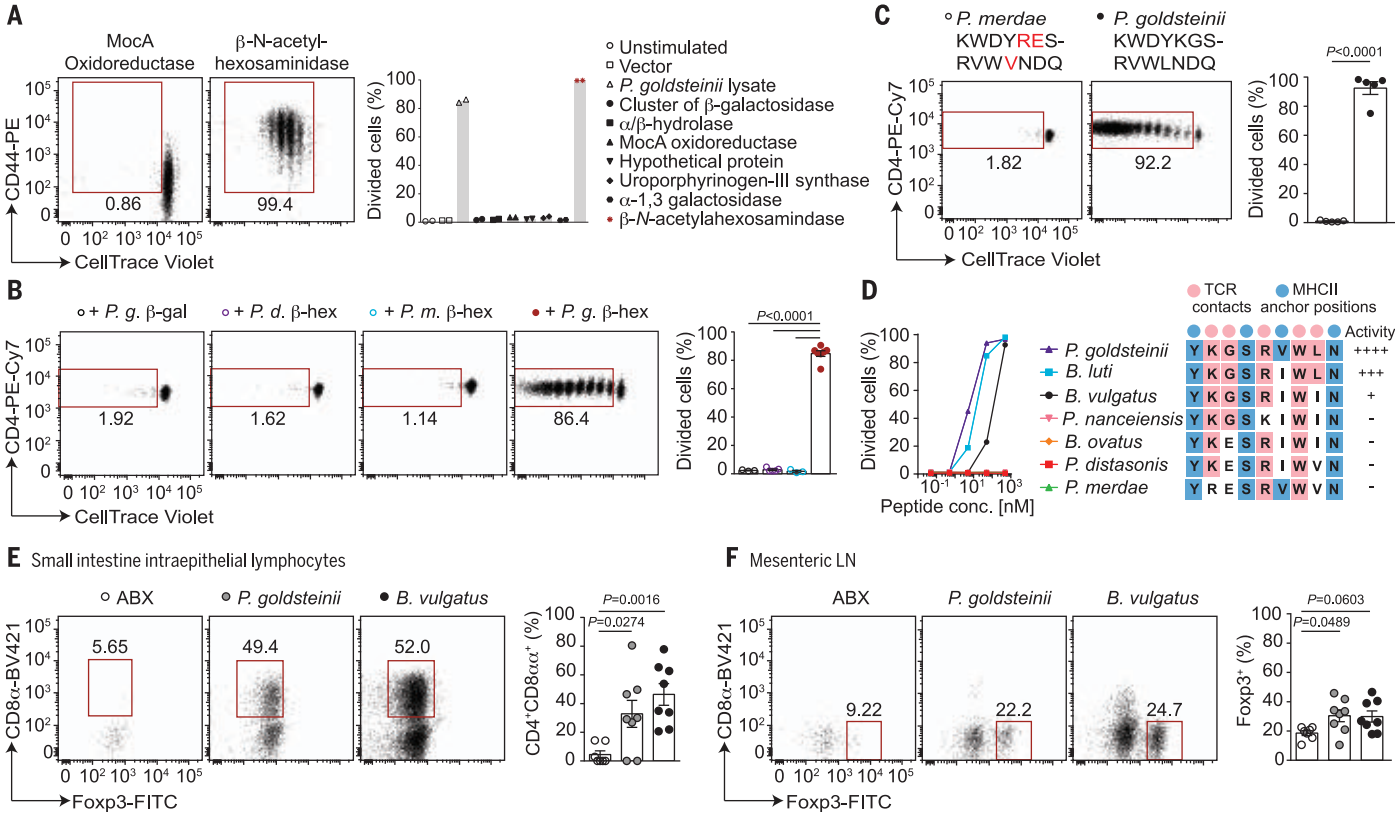


Fig. 2. The TN TCR recognizes epitopes from Bacteroidetes β-*N*-acetylhexosaminidase in complex with I-A^b. (**A**) Candidate polypeptides identified with mass spectrometry were recombinantly expressed in *E. coli*, and bacterial extracts were tested in vitro for their ability to activate TN T cells. (Left) Representative dot plots. (Right) Frequency of divided TN T cells in the presence of tested extracts. Graph shows means of two technical replicates. (**B** and **C**) In vivo proliferation of CD4$^+$ TN T cells in response to 25 μg of bacterial extracts [(B), *n* = 3 to 7 mice per group] or 2 μg of peptide [(C), *n* = 5 mice per group)]. Mutations in the core epitope are indicated in red. (**D**) (Left) Same as (A), but by using the indicated peptides at concentrations of 500 nM to 50 pM in serial 10-fold dilutions. (Right) Alignment of sequences homologous to the TN epitope. "Activity" is the ability of each peptide to induce proliferation of TN T cells in vitro (left). (**E** and **F**) CD45.1$^+$ recipients were treated with ABX and received CD45.2$^+$CD4$^+$ TN T cells. The mice were then either colonized or not colonized with the indicated bacteria (*n* = 7 or 8 mice per group). Four weeks after colonization, (E) small intestine intraepithelial lymphocytes and (F) mLNs were analyzed. Dot plots show CD8α and Foxp3 expression, and graphs show the frequency of CD4IELs (CD4$^+$CD8αα$^+$) in the (E) SIE and (F) mLN T_{reg} cells (Foxp3$^+$) among TN CD4$^+$ T cells. In vitro experiments are representative of three experiments [(A) and (D)]. Dot plots for in vivo experiments show one representative mouse, and the graphs show all mice analyzed in two experiments. Each symbol represents a mouse, and graphs show means ± SEM [(B), (C), (E), and (F)]. *P* values were calculated with one-way analysis of variance (ANOVA) with Dunnett's post hoc test in (B), (E), and (F) or unpaired two-tailed Student's *t* test in (C). *P.g.*, *P. goldsteinii*; *P.d.*, *P. distasonis*; *P.m.*, *P. merdae*.

expressing *P. goldsteinii* β-hex but not when immunized with β-hex derived from closely related species (Fig. 2B). By analyzing the activity of truncated versions of the protein, we defined a ~70-residue stretch that contained the cognate epitope of the TN TCR. Using overlapping peptides, we identified YKGSRVWLN as the minimal epitope (fig. S6, A to D). The corresponding β-hex peptide from *Parabacteroides merdae* failed to stimulate TN T cells, even though it bound to I-A^b (Fig. 2C; fig. S6, E to G; and table S2). Thus, the *P. goldsteinii* YKGSRVWLN β-hex epitope is a natural ligand of the TN TCR. Homologous β-hex peptides derived from *Bacteroides vulgatus* and *Bacteroides luti* also induced proliferation of TN cells in vitro, but only at higher concentrations (Fig. 2D). Using a combination of BlastP and Jackhmmer analyses (*24*), we found putative β-hex immunostimulatory sequences in many Bacteroidetes members—the most abundant phylum found exclusively in the gastrointestinal tract (*25*, *26*)—as well as in more distantly related species, such as *Spirosoma panaciterrae* (fig. S7). Thus, the TN TCR likely recognizes an even broader collection of bacterial species that belong to the Bacteroidetes phylum.

Colonization of SPF mice with *P. goldsteinii* promoted the expansion and development of CD4IELs and pT$_{reg}$ cells from TN CD4$^+$ T cells. Likewise, *B. vulgatus*, which also encodes a TN β-hex epitope (Fig. 2D), supported the expansion and development of TN CD4IELs (Fig. 2, E and F, and fig. S8, A to C). By contrast, colonization of mice with a β-hex–deficient strain of *B. vulgatus* failed to do so (fig. S8, D to H). Despite similar expression of β-hex along the intestinal tract (fig. S9, A and B), TN cells transferred into SPF mice proliferated predominantly in the jejunum/ileum–draining mLN (fig. S9, C to E). Thus, the β-hex epitope is presented in the distal mLN and is necessary for the differentiation of CD4IELs from TN precursors.

To determine whether this prevalent β-hex epitope is also recognized by intestinal CD4$^+$ lymphocytes of WT SPF mice, we designed *P. goldsteinii* β-hex–MHCII tetramers spanning the β-hex epitope (table S3) and enumerated β-hex–specific CD4$^+$ T cells by means of flow cytometry (Fig. 3, A to C, and fig. S10, A to L). Splenic CD4$^+$ T cells showed equally low staining with control and β-hex tetramers, whereas β-hex–specific cells were readily identified in both gut-draining mLNs and small intestinal epithelium (Fig. 3, A to C, and fig. S10, A to F). The frequency of β-hex–specific cells was significantly higher among mLN T$_{reg}$ cells of mice housed at both Boston Children's

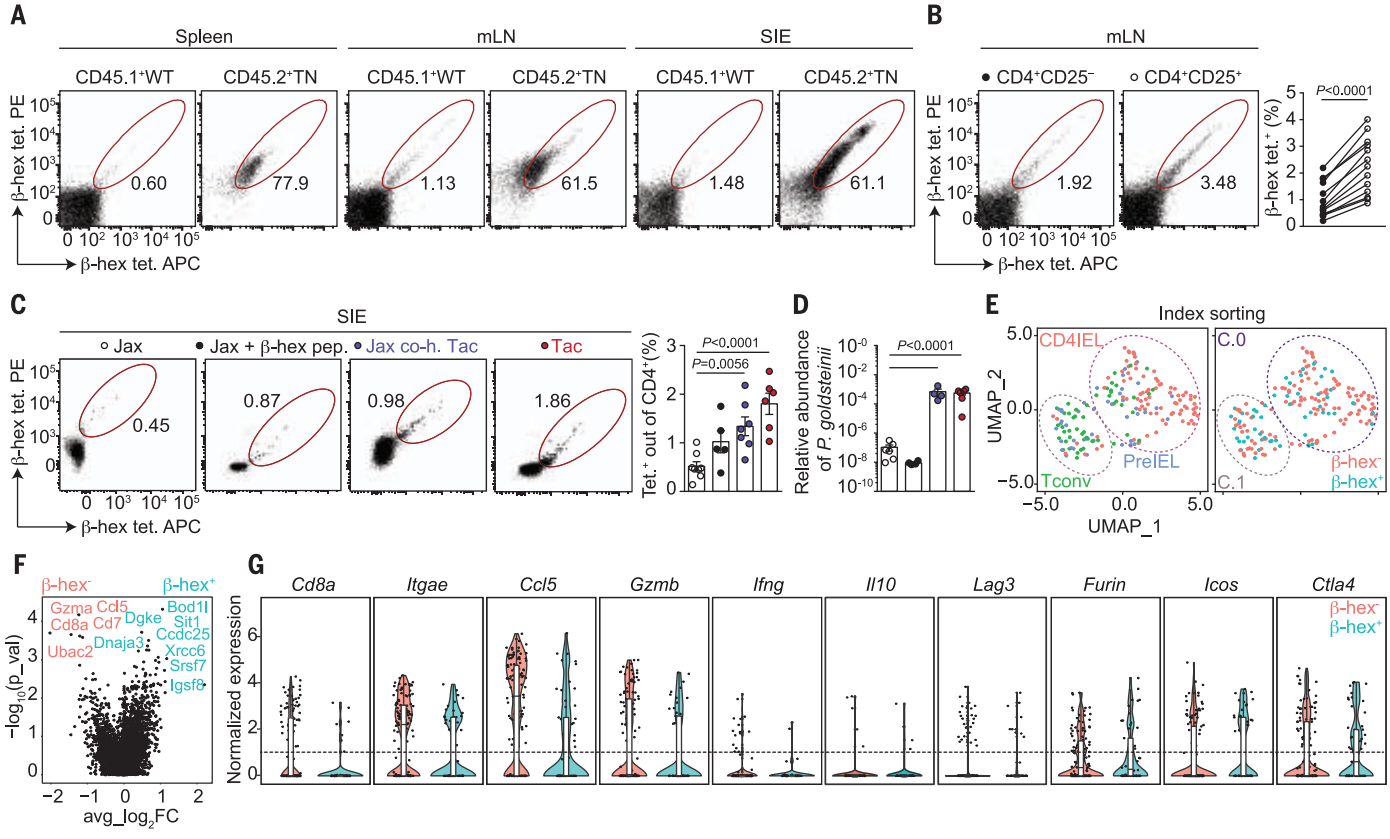


Fig. 3. The β-hexosaminidase epitope is recognized by intraepithelial lymphocytes and T_{reg} cells. (**A**) Cells from spleen, mLNs, and the SIE were harvested from SPF mice housed at Boston Children's Hospital. Cells from CD45.1$^+$ WT mice and CD45.2$^+$ TN mice were mixed at 10:1 (WT:TN) ratio and stained with β-hex tetramers. Gates show the frequency of tetramer$^+$ CD45.1$^+$ or CD45.2$^+$ cells among CD4$^+$ T cells. (**B**) Representative dot plots of mLNs from Tac SPF mice (n = 13) stained with the indicated tetramers (left). Frequencies of tetramer$^+$ cells in the indicated populations among CD4$^+$ T cells isolated from mLNs (right). (**C**) Tetramer analysis among SIE CD4$^+$ T cells of WT mice sourced from the indicated facilities, cohoused or treated orally with β-hex peptide as indicated (n = 5 to 8 mice per group). (**D**) Quantification of *P. goldsteinii* abundance by means of quantitative polymerase chain reaction in fecal samples derived from the mice shown in (C). (**E** to **G**) scRNA-seq analysis of β-hex–specific (β-hex$^+$) and β-hex$^-$ CD4$^+$ T cells sorted from SIE of three Taconic SPF mice. (E) UMAP plots according to gene expression analysis showing the distribution of conventional T cells (Tconv, CD103$^-$CD8α$^-$), PreIEL (CD103$^+$CD8α$^-$), and CD4IELs (CD103$^+$CD8α$^+$) according to the index sorting. (F) Volcano plot comparing the gene expression of β-hex$^+$ and β-hex$^-$ cells. (G) Expression of selected genes by β-hex$^+$ and β-hex$^-$ cells. All mice shown of [(E) to (G)] two, (A) three, (B) five, or [(C) and (D)] six experiments. Error bars indicate means ± SEM. P values were calculated with paired two-tailed Student's t tests in (B), one-way ANOVA with Dunnett's post-hoc test in (C) and (D), and Wilcoxon rank sum test in (F) and (G). Jax, Jackson; Tac, Taconic; co-h, co-housed.

Hospital and Rockefeller University animal facilities (Fig. 3B and fig. S10G). In mLNs, β-hex–specific cells accounted for up to 4% of all CD25$^+$CD4$^+$ T cells, a population highly enriched for T_{reg} cells (Fig. 3B) (*27*). Despite the intrinsically reduced ability of CD4IELs to bind MHCII tetramers (fig. S10H), 40% of β-hex tetramer$^+$ cells in the epithelium were CD4IELs (fig. S10, I and J). The frequencies of β-hex–specific CD4$^+$ T cells varied by mouse, with age, and even between rooms within the same facility, ranging from 1 to 20% in the epithelium and 0.2 to 2% in the mLN (fig. S10, C to F, I, and K). We then examined the extent to which the commensal-derived β-hex antigen shapes the epithelial lymphocyte compartment (Fig. 3C and fig. S10L). In Jax mice, which carry low frequencies of CD4IELs (fig. S1A) and are virtually devoid of *P. goldsteinii* (fig. S3, A and B), the proportion of β-hex–specific IELs was lower than in Taconic mice (Fig. 3, C and D). Taconic mice had high frequencies of CD4IELs and were natural carriers of intestinal bacteria encoding the β-hex antigen, including *P. goldsteinii* and other Bacteroidetes (fig. S1E, and fig. S3, A to C). Similarly, Jax mice cohoused with Taconic mice were efficiently colonized with *P. goldsteinii* and showed a higher proportion of β-hex–specific IELs, close to that observed in the Taconic mice themselves (Fig. 3, C and D). Treatment of Jax mice with β-hex peptide slightly increased the frequencies of β-hex–specific IELs, although this difference did not reach statistical significance (Fig. 3C). Thus, the *P. goldsteinii* β-hex epitope is a prominent natural T cell receptor ligand for mLN T_{reg} cells and intraepithelial CD4$^+$ T cells. We conclude that a microbiota rich in *P. goldsteinii* supports the expansion of commensal-specific IELs, using as example the β-hex antigen derived from abundant intestinal commensals such as *Parabacteroides* and *Bacteroides* spp.

To define the transcriptional signature of β-hex–specific IELs, we index-sorted β-hex$^+$ and β-hex$^-$ IELs from Taconic mice (fig. S11A) and profiled them by means of single-cell RNA-sequencing (scRNA-seq) using the SMART-Seq2 methodology (*28*). This approach enabled us to analyze the gene expression profile of commensal-induced β-hex–specific IELs at homeostasis. We identified two major clusters visualized with uniform manifold approximation and projection (UMAP) (Fig. 3E and fig. S11B). Cluster 0 contained mostly *Cd8a* and *Itgae* (encoding CD103)–expressing cells, and a gene signature typical of CD4IELs, which includes both cytotoxic and regulatory profiles as defined by expression of granzymes and *Ctla4* (*20*, *29*). Cluster 1 was composed

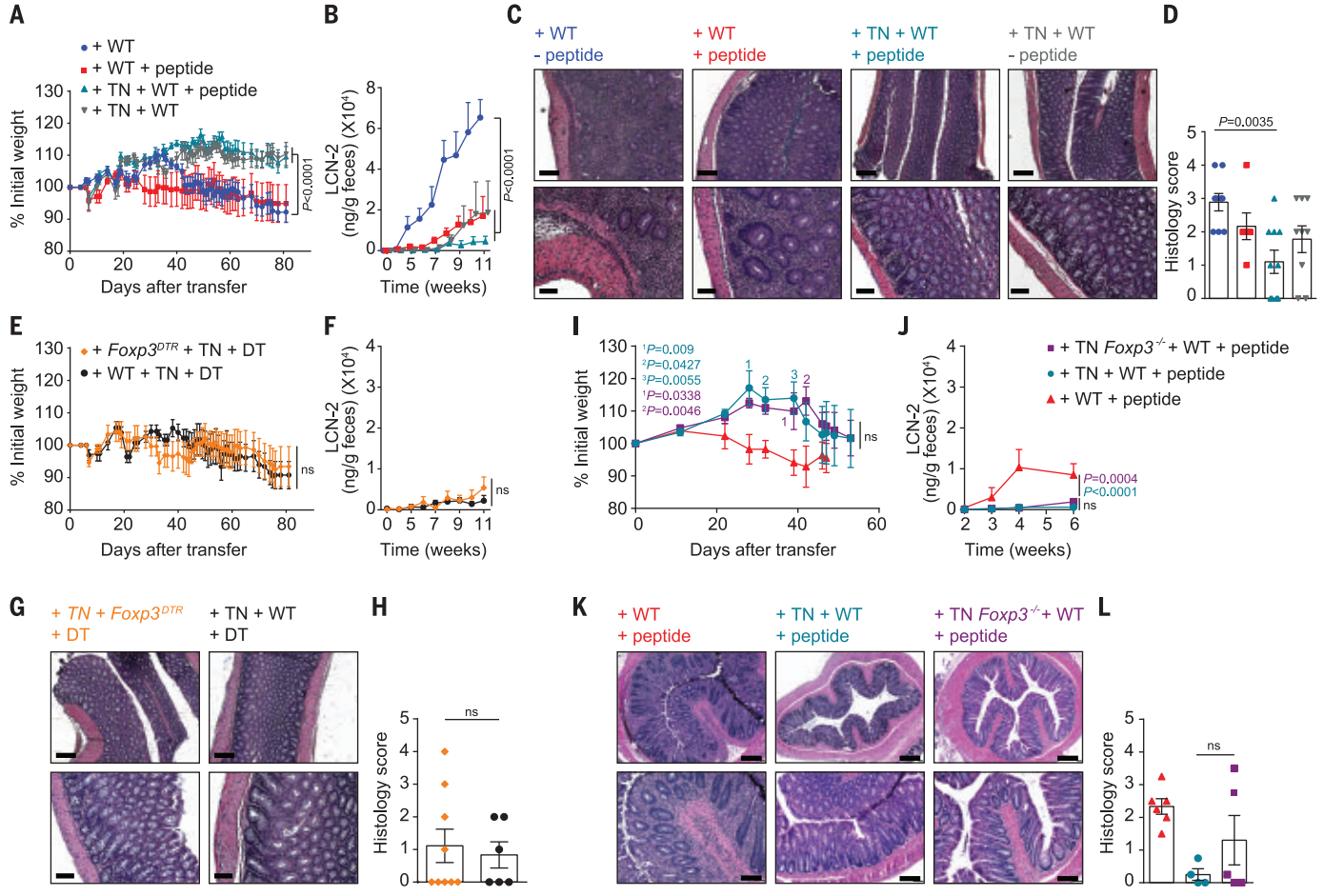


Fig. 4. β-hexosaminidase–specific T cells confer protection against intestinal inflammation. (**A** to **D**) Taconic SPF *Rag2*$^{-/-}$ either received or did not receive CD45.2^{+}CD4^{+} TN T cells (day 0) and naïve CD45.1^{+}CD4^{+} T cells from WT mice (day 12). On days 2, 4, and 13, some mice received β-hex peptide orally (n = 8 to 10 mice). (A) Percent of initial weight over time. (B) Fecal lipocalin (LCN-2) levels throughout the experiment (n = 5 to 8 mice). (C) Representative photomicrographs of hematoxylin and eosin (H&E)–stained colonic sections from indicated groups. (D) Histology score of H&E-stained sections of colon. (**E** to **H**) Same as (A) to (D) but with either WT (n = 7 mice) or *Foxp3*^{DTR} (n = 10 mice) CD4^{+} T cells (on day 12 after transfer of TN T cells). These mice received DT intraperitoneally once per week throughout the experiment. (E) Percent of initial weight over time. (F) Fecal LCN-2 levels throughout the experiment. (G) Representative photomicrographs of H&E-stained colonic sections from indicated groups. (H) Histology score of H&E-stained sections of colon. (**I** to **L**) Same as (A) to (D) but by using either WT (n = 5 mice) or *Foxp3*$^{-/-}$ TN CD4^{+} cells (n = 5 mice, day 0). (I) Percent of initial weight over time. (J) Fecal LCN-2 levels. (K) Representative photomicrographs of H&E-stained sections of colon of indicated groups. (L) Histology scores of H&E-stained sections of colon. (A) to (L) show all mice analyzed in two experiments. Graphs show means ± SEM. P values were calculated with two-way ANOVA for repeated measures with the Tukey post hoc test in (A), (B), (E), (F), (I), and (J); Mann-Whitney test in (H); and Kruskal-Wallis test followed by Dunn's multiple comparisons post-hoc test in (D) and (L). Scale bar, (C) and (G) (top) 200 μm and (bottom) 60 μm; (K) (top) 200 μm and (bottom) 100 μm. Low- and high-magnification photomicrographs show two representative mice in each group in (C), (G), and (K). ns, not significant.

predominantly of *Cd8α*$^{-}$ IELs, including pre-IELs (fig. S11C) (*20*, *29*). β-hex^{+} and β-hex^{-} IELs were largely indistinguishable from each other and were evenly distributed among the two clusters (Fig. 3E and fig. S11, D to F). There were no differentially expressed genes between β-hex^{+} and β-hex^{-} IELs, and both subsets expressed typical IEL markers (Fig. 3, F and G, and fig. S11, E to G), associated with both regulatory (*Ctla4*) and cytotoxic (*Gzmb*) profiles. We observed a slight enrichment of β-hex^{+} IELs in cluster 1 compared with cluster 0 (Fig. 3, F and G, and fig. S11D). This enrichment may have been related to the reduced ability of CD4IELs to bind MHCII-tetramers (fig. S10H), presumably because of the inhibitory interactions between the thymus leukemia (TL) antigen on epithelial cells and CD8αα on CD4IELs (*30*). Thus, the gene expression profile of β-hex–specific IELs is representative of all IELs. Consequently, β-hex likely represents one of many possible specificities that lead CD4^{+} T cells to acquire an IEL phenotype after initial priming by a commensal antigen.

Although CD4IELs express IFNγ and granzyme B, they display a regulatory phenotype—characterized by the expression of interleukin-10 and Lag3 (Fig. 3G) (*11*, *20*, *29*)—and can protect against intestinal inflammation (*14*–*16*). To test the anti-inflammatory potential of β-hex–specific T cells, we transferred CD4^{+} TN T cells into immunodeficient *Rag2*$^{-/-}$ mice and then induced colitis through WT CD4^{+} T cells (fig. S12A) (*31*). WT cells expanded in the intestine, irrespective of whether TN cells were present or not (fig. S12B). TN cells efficiently differentiated into CD4IELs (fig. S12C) but poorly differentiated into pT$_{reg}$ cells (fig. S12D), similar to what has been observed for other commensal-specific T cells transferred into *Rag1*$^{-/-}$ recipients (*32*). TN cells are found predominantly in the small intestine and mLNs (*11*), but in this colitis model, TN cells were also present in the large intestine, where they acquired expression of CD8αα (fig. S12, B and C). Mice

that received TN cells were partially protected from colitis induced by colitogenic WT cells. We observed reduced weight loss, improved histopathological scores, and lower levels of fecal lipocalin-2 (LCN-2), a marker of intestinal inflammation (Fig. 4, A to D) (*33*). Oral administration of β-hex peptide—a procedure that expands TN cells and leads to their differentiation into CD4IELs (fig. S12E)—further reduced intestinal inflammation in recipients of TN and WT $CD4^+$ T cells (Fig. 4, B to D). Similarly, mice that received only colitogenic WT cells and oral β-hex peptide also exhibited decreased intestinal inflammation and increased frequency of total CD4IELs in the small intestine (Fig. 4B and fig. S12C). At an earlier time point, we also observed an increased frequency of total CD4IELs as well as of β-hex–specific IELs, which was accompanied by a decreased frequency of T_{reg} cells in the epithelium (fig. S12, F to H). The frequency of WT-derived T_{reg} cells was low in all groups, suggesting little contribution by pT_{reg} cells to such protection (fig. S12D), as has been observed previously (*34*). To definitively exclude a role for WT T_{reg} cells, we induced colitis by transferring T cells from *Foxp3*^{DTR} mice and depleted T_{reg} cells with diphtheria toxin (DT) (Fig. 4, E to H, and fig. S12, I to J). Depletion of T_{reg} cells in this colitis model leads to severe weight loss, intestinal inflammation, and death (*35*). Despite this depletion, TN T cells still protected mice against intestinal inflammation. Although mice lost weight because of toxicity associated with repeated injections of DT (*36*, *37*), there was no difference between the two groups (recipients of *Foxp3*^{DTR} versus *Foxp3*^{WT} naïve T cells) (Fig. 4, F to H). Similar to WT T_{reg} cells, the proportion of TN T cells that differentiated into T_{reg} cells was low but not null in this model (fig. S12D). To evaluate a possible contribution of TN-derived T_{reg} cells to the anti-inflammatory response induced by β-hex–specific TN T cells, we ablated Foxp3 directly in pT_{reg} cell TN/*Rag*1$^{-/-}$/*Foxp3*^{eGFP} zygotes by using CRISPR/Cas 9 genome editing. We generated mice with either a 1–base pair (bp) deletion or 1-bp insertion frameshifts in exon 8 of the *Foxp3* gene (fig. S13, A to C). We adoptively transferred splenic $CD4^+$ T cells from these Foxp3-deficient TN mice into *Rag2*$^{-/-}$ recipients and then induced colitis by using WT naïve $CD4^+$ T cells as outlined above (fig. S13D). Foxp3-deficient TN T cells (fig. S13E) conferred partial protection against intestinal inflammation (Fig. 4, I to L). Thus, the anti-inflammatory effect exerted by TN T cells in this model can occur in the absence of Foxp3 expression. Foxp3-deficient and -sufficient TN T cells were equally able to expand in all intestinal sites surveyed (fig. S13F). In this colitis model, TN T cells differentiated into CD4IELs (up to 80%) and secreted IFNγ and granzyme B, which are hallmarks of IELs, irrespective of the presence of a functional *Foxp3* gene in the TN donors (fig. S13, G to I). Thus, β-hex–specific T cells that migrate to, expand within, and differentiate into CD4IELs in response to commensal antigens can protect against intestinal inflammation.

Commensal bacteria shape the differentiation and function of intestinal T cells (*1*), including CD4IELs. We show that *P. goldsteinii* and other Bacteroidetes can induce antigen-specific differentiation of CD4IELs. It is possible that a fraction of β-hex–specific pT_{reg} cells generated in the mLNs migrate to the epithelium. There, they can further differentiate into preIELs or CD4IELs after reencountering their cognate antigen (*14*, *20*). Because $CD4^+$ T cells require local engagement with MHCII to differentiate into CD4IELs (*20*), recognition of an antigen such as β-hex, present across a range of abundant commensals, may provide CD4IEL precursors with a competitive advantage to populate the intestinal epithelium. The same β-hex–specific TCR rearrangement found in the TN strain is present among CD4IELs and their epithelial precursors in mice that carry only the Vβ6 chain of the TN TCR and a polyclonal TCRα chain (*20*). Although we identified β-hex as a natural ligand of intraepithelial $CD4^+$ T cells, it is likely that other commensal-derived antigens are also recognized by these cells. Because activation of CD4IELs is attenuated by interactions between TL antigen expressed by intestinal epithelial cells and CD8αα on IELs (*38*), antigen abundance and TCR affinity may play a role in the development of CD4IELs. In addition, commensals such as *P. goldsteinii* and *B. vulgatus* may not only provide TCR ligands but also metabolites essential for the creation of an environment that supports the development and maintenance of regulatory IELs. In this work, we showed that β-hex–specific T cells that reside in the epithelium can protect against intestinal inflammation. The TCR-titratable regulation imposed by the interaction between the TL antigen and CD8αα (*38*) may control the anti-inflammatory and cytotoxic functions of CD4IELs. Deciphering the rules that govern the mutualism between commensals, pathobionts, and immune cells is essential for a better understanding of homeostasis and inflammation at the intestinal mucosa.

REFERENCES AND NOTES

1. Y. Belkaid, O. J. Harrison, *Immunity* **46**, 562–576 (2017).
2. I. I. Ivanov, K. Honda, *Cell Host Microbe* **12**, 496–508 (2012).
3. M. B. Geuking, R. Burkhard, *Mucosal Immunol.* **13**, 855–866 (2020).
4. E. Ansaldo *et al.*, *Science* **364**, 1179–1184 (2019).
5. E. Sefik *et al.*, *Science* **349**, 993–997 (2015).
6. S. K. Lathrop *et al.*, *Nature* **478**, 250–254 (2011).
7. Y. Yang *et al.*, *Nature* **510**, 152–156 (2014).
8. M. Xu *et al.*, *Nature* **554**, 373–377 (2018).
9. J. N. Chai *et al.*, *Sci. Immunol.* **2**, eaal5068 (2017).
10. I. I. Ivanov *et al.*, *Cell* **139**, 485–498 (2009).
11. A. M. Bilate *et al.*, *Sci. Immunol.* **1**, eaaf7471 (2016).
12. A. M. Bilate, J. J. Lafaille, *Annu. Rev. Immunol.* **30**, 733–758 (2012).
13. M. Shale, C. Schiering, F. Powrie, *Immunol. Rev.* **252**, 164–182 (2013).
14. T. Sujino *et al.*, *Science* **352**, 1581–1586 (2016).
15. B. S. Reis, A. Rogoz, F. A. Costa-Pinto, I. Taniuchi, D. Mucida, *Nat. Immunol.* **14**, 271–280 (2013).
16. J. Basu *et al.*, *Nat. Immunol.* **22**, 969–982 (2021).
17. L. Cervantes-Barragan *et al.*, *Science* **357**, 806–810 (2017).
18. Y. Umesaki, H. Setoyama, S. Matsumoto, Y. Okada, *Immunology* **79**, 32–37 (1993).
19. D. Mucida *et al.*, *Nat. Immunol.* **14**, 281–289 (2013).
20. A. M. Bilate *et al.*, *Immunity* **53**, 1001–1014.e20 (2020).
21. L. Wojciech *et al.*, *Sci. Rep.* **8**, 10848 (2018).
22. F. E. Dewhirst *et al.*, *Appl. Environ. Microbiol.* **65**, 3287–3292 (1999).
23. B. S. Reis, D. P. Hoytema van Konijnenburg, S. I. Grivennikov, D. Mucida, *Immunity* **41**, 244–256 (2014).
24. S. R. Eddy, *PLOS Comput. Biol.* **7**, e1002195 (2011).
25. A. G. Wexler, A. L. Goodman, *Nat. Microbiol.* **2**, 17026 (2017).
26. Human Microbiome Project Consortium, *Nature* **486**, 207–214 (2012).
27. J. D. Fontenot, M. A. Gavin, A. Y. Rudensky, *Nat. Immunol.* **4**, 330–336 (2003).
28. J. J. Trombetta *et al.*, *Curr. Protoc. Mol. Biol.* **107**, 22.1–17, 17 (2014).
29. M. London, A. M. Bilate, T. B. R. Castro, T. Sujino, D. Mucida, *Nat. Immunol.* **22**, 449–459 (2021).
30. D. Olivares-Villagómez *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17931–17936 (2008).
31. F. Powrie, M. W. Leach, S. Mauze, L. B. Caddle, R. L. Coffman, *Int. Immunol.* **5**, 1461–1471 (1993).
32. K. Nutsch *et al.*, *Cell Rep.* **17**, 206–220 (2016).
33. B. Chassaing *et al.*, *PLOS ONE* **7**, e44328 (2012).
34. A. Izcue *et al.*, *Immunity* **28**, 559–570 (2008).
35. F. Boehm *et al.*, *BMC Gastroenterol.* **12**, 97 (2012).
36. Y. Dombrowski *et al.*, *Nat. Neurosci.* **20**, 674–680 (2017).
37. A. Goldwich, A. Steinkasserer, A. Gessner, K. Amann, *Lab. Invest.* **92**, 1674–1685 (2012).
38. D. Olivares-Villagómez, L. Van Kaer, *Immunol. Lett.* **134**, 1–6 (2010).

ACKNOWLEDGMENTS

We thank J. Leube for helping with culture experiments; P. Silver for access to anaerobic chambers; C. McClune for assistance with bioinformatic analyses; S. Jain, J. Bhagat, and M. Suarez-Farinas for statistical consultation; and O. Venezia from the Moon laboratory for assistance with tetramer production. We are grateful to M. Fischbach for supporting D.B. in performing experiments during revision of this manuscript and to M. Laub and D. Irvine for their guidance throughout the project. We thank S. Kolifrath, A. Rogoz, and S. Gonzalez for mouse husbandry; R. Mathieu from the flow cytometry facilities of BCH and K. Gordon from RU; E. Spooner (WIBR) for mass spectrometry; R. Bronson (Harvard) and Y. Alvarez (RU) and the NYU Histology core facility for histology and colitis scoring; and the NIH tetramer core facility for providing tetramers. We thank members of the Ploegh, Victora, and Mucida laboratories for discussions and suggestions. **Funding:** H.L.P. was supported by CMIT, and D.M. was supported by FASI/FARE. D.B. and H.L.P. were supported by NIH R01AI087879 and DP1AI150593-01. A.M.B., M.L., and D.M. were supported by NIH R01DK093674-07, R01DK113375, and R01DK093674. T.B.R.d.C., J.B., and G.D.V. were supported by NIH DP1AI144248. J.J.M. was supported by NIH R01DK126910 and P30DK043351, and MGH ECOR. D.B. is currently funded by an Early Postdoc Mobility Fellowship from the Swiss National Science Foundation. **Author contributions:** A.M.B., D.B., and H.L.P. conceived the study and wrote the manuscript. A.M.B. and D.B. designed and performed the experiments. D.B. performed the in vitro experiments, with help from A.M.B. and P.B.; D.B. and A.M.B. performed the in vivo experiments, with help from L.I.F., M.L., P.B., B.R., and A.G.; T.B.R.d.C. performed scRNA-seq analysis. M.M. and M.P. provided bacterial isolates, and M.M. generated all β-hex–deficient strains, under the supervision of T.K.L.; S.O. analyzed the 16*S* sequencing data, under the supervision of E.J.A.; D.V., R.W.C., and J.L. provided support and guidance. J.B. and A.M.B. generated the *Foxp3*-deficient TN mouse line, under the supervision of G.D.V.; J.S. performed MHC II–peptide–binding assays, under the supervision of A.S.; J.J.M. generated tetramers for this study.

J.J.M., G.D.V., and D.M. assisted in data interpretation. A.M.B. and H.L.P. supervised the study. **Competing interests:** E.J.A. is a consultant and has equities at Finch Therapeutics. T.K.L. is a cofounder of Senti Biosciences, Synlogic, Engine Biosciences, Tango Therapeutics, Corvium, BiomX, Eligo Biosciences, Bota.Bio, Avendesora, and NE47Bio. T.K.L. also holds financial interests in nest.bio, Armata, IndieBio, CognitoHealth, Quark Biosciences, Personal Genomics, Thryve, Lexent Bio, MitoLab, Vulcan, Serotiny, Avendesora, Pulmobiotics, Provectus Algae, Invaio, and NSG Biolabs. The other authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Raw and processed scRNA-seq data can be accessed at the Gene Expression Omnibus (GEO) under accession no. GSE198037. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg5645

Materials and Methods
Figs. S1 to S14
Tables S1 to S5
References (*39*–*57*)
MDAR Reproducibility Checklist
Data S1

Submitted 13 January 2021; accepted 24 June 2022
10.1126/science.abg5645

MUSCLE REPAIR

JMJD3 activated hyaluronan synthesis drives muscle regeneration in an inflammatory environment

Kiran Nakka[1,2], Sarah Hachmer[1,2]†, Zeinab Mokhtari[1,2]†, Radmila Kovac[1,2], Hina Bandukwala[1], Clara Bernard[3], Yuefeng Li[1,2], Guojia Xie[4], Chengyu Liu[5], Magid Fallahi[1], Lynn A. Megeney[1,2], Julien Gondin[3], Bénédicte Chazaud[3], Marjorie Brand[1,2,6], Xiaohui Zha[1,7], Kai Ge[4], F. Jeffrey Dilworth[1,2,6]*

Muscle stem cells (MuSCs) reside in a specialized niche that ensures their regenerative capacity. Although we know that innate immune cells infiltrate the niche in response to injury, it remains unclear how MuSCs adapt to this altered environment for initiating repair. Here, we demonstrate that inflammatory cytokine signaling from the regenerative niche impairs the ability of quiescent MuSCs to reenter the cell cycle. The histone H3 lysine 27 (H3K27) demethylase JMJD3, but not UTX, allowed MuSCs to overcome inhibitory inflammation signaling by removing trimethylated H3K27 (H3K27me3) marks at the *Has2* locus to initiate production of hyaluronic acid, which in turn established an extracellular matrix competent for integrating signals that direct MuSCs to exit quiescence. Thus, JMJD3-driven hyaluronic acid synthesis plays a proregenerative role that allows MuSC adaptation to inflammation and the initiation of muscle repair.

Muscle stem cells (MuSCs) provide skeletal muscle with an efficient mode for regeneration of damaged tissue (*1*). After injury, the regenerative process is initiated by necrosis of damaged myofibers and the release of myokines that instruct recruitment of various tissue-resident and infiltrating cell types that coordinate muscle repair (*2*). Tight control of signal integration from the regenerative milieu promotes the expansion of muscle progenitors to mediate both myofiber repair and stem cell–niche repopulation (*3*, *4*). Recent work identified a population of infiltrating macrophages that establish a transient niche necessary for quiescent MuSCs to reenter the cell cycle (*5*). However, it remains unknown how MuSCs adapt to this modified niche to initiate regeneration.

In response to injury, MuSCs undergo a stress response that is accompanied by alteration of the epigenetic landscape (*6*). Among the different epigenetic changes induced by injury in MuSCs, global levels of trimethylated histone H3 lysine 27 (H3K27me3) are reduced as stem cells transition from a quiescent to a proliferative state (*7*). Removal of H3K27me3 marks is mediated by the KDM6 family of H3K27 demethylases, which includes JMJD3 and UTX proteins (*8*, *9*). Although H3K27me3 marks are tightly linked to transcriptional repression (*10*, *11*), the importance of active H3K27me3 removal to tissue development and repair has been called into question because mouse embryos that lack both JMJD3 and UTX survive to term (*12*). However, regeneration after injury is linked to inflammation, which differs from tissue development. Here, we examined the importance of H3K27 demethylation by JMJD3 and UTX for MuSC adaptation to the regenerative niche of injured muscle.

The *Utx* and *Jmjd3* genes are ubiquitously expressed, and we observed that both proteins are present in quiescent MuSCs (fig. S1). To determine how each H3K27 demethylase contributes to regeneration, we generated tamoxifen-inducible MuSC-specific knockouts of *Utx* (UTX^{scKO}) or *Jmjd3* ($JMJD3^{scKO}$) in mice (fig. S2). Upon acute cardiotoxin (CTX) injury of the tibialis anterior (TA) muscle, both UTX^{scKO} and $JMJD3^{scKO}$ mice showed impaired myofiber regeneration, demonstrating that KDM6 proteins cannot compensate for each other in MuSC-mediated regeneration (Fig. 1A and fig. S3). To investigate the specific roles for UTX and JMJD3, single-cell RNA sequencing (scRNA-seq) was performed on purified MuSCs that were isolated 40 hours after injury. Pseudotime trajectory analysis showed that $JMJD3^{scKO}$ MuSCs were enriched in clusters that expressed an immediate-early MuSC activation gene signature, whereas UTX^{scKO} cells proceeded to proliferate (Fig. 1, B and C; and fig. S4). These findings suggested that JMJD3 might be required for efficient transitioning of MuSCs from a quiescent to a proliferating state. Indeed, using the thymidine analog 5-ethynyl-2′-deoxyuridine (EdU) to identify cells undergoing DNA replication, we observed that MuSCs that lacked JMJD3 were impaired in their progression into the cell cycle (Fig. 1D). Thus, JMJD3, but not UTX, is required for efficient activation of MuSCs. At the same time, we observed that UTX, but not JMJD3, is required for MuSCs to exit the cell cycle and undergo terminal differentiation (fig. S5) (*13*). Thus, JMJD3 and UTX play nonredundant roles in MuSC-mediated regeneration.

To determine whether JMJD3 was acting through its enzymatic activity, a mutant mouse was generated that expresses an enzyme-dead JMJD3 protein ($JMJD3^{scDD}$) (fig. S6, A and B). Similar to $JMJD3^{scKO}$ cells, MuSCs that expressed mutant JMJD3 were inefficient at exiting quiescence (Fig. 1D) and, as a result, showed impaired regeneration after injury (fig. S6, C and D). This demonstrates that JMJD3 functions through H3K27 demethylation to facilitate the exit of quiescence for MuSCs in injured muscle.

Next, we used ex vivo experiments to explore the mechanism through which JMJD3 contributes to the activation process. Using MuSCs purified from uninjured mice, we were surprised to find that $JMJD3^{scKO}$ MuSCs reentered the cell cycle to the same extent as wild-type MuSCs (Fig. 2A and fig. S7). Thus, in the absence of muscle injury, JMJD3 was not necessary for cell cycle reentry. This indicated

[1]Sprott Center for Stem Cell Research, Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada. [2]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada. [3]Institut NeuroMyoGène, Unité Physiopathologie et Génétique du Neurone et du Muscle, Université Claude Bernard Lyon 1, CNRS 5261, INSERM U1315, Université de Lyon, Lyon, France. [4]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [5]Transgenic Core, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA. [6]LIFE Research Institute, University of Ottawa, Ottawa, ON, Canada. [7]Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, Ottawa, ON, Canada.
***Corresponding author. Email: jdilworth@ohri.ca**
†These authors contributed equally to this work.

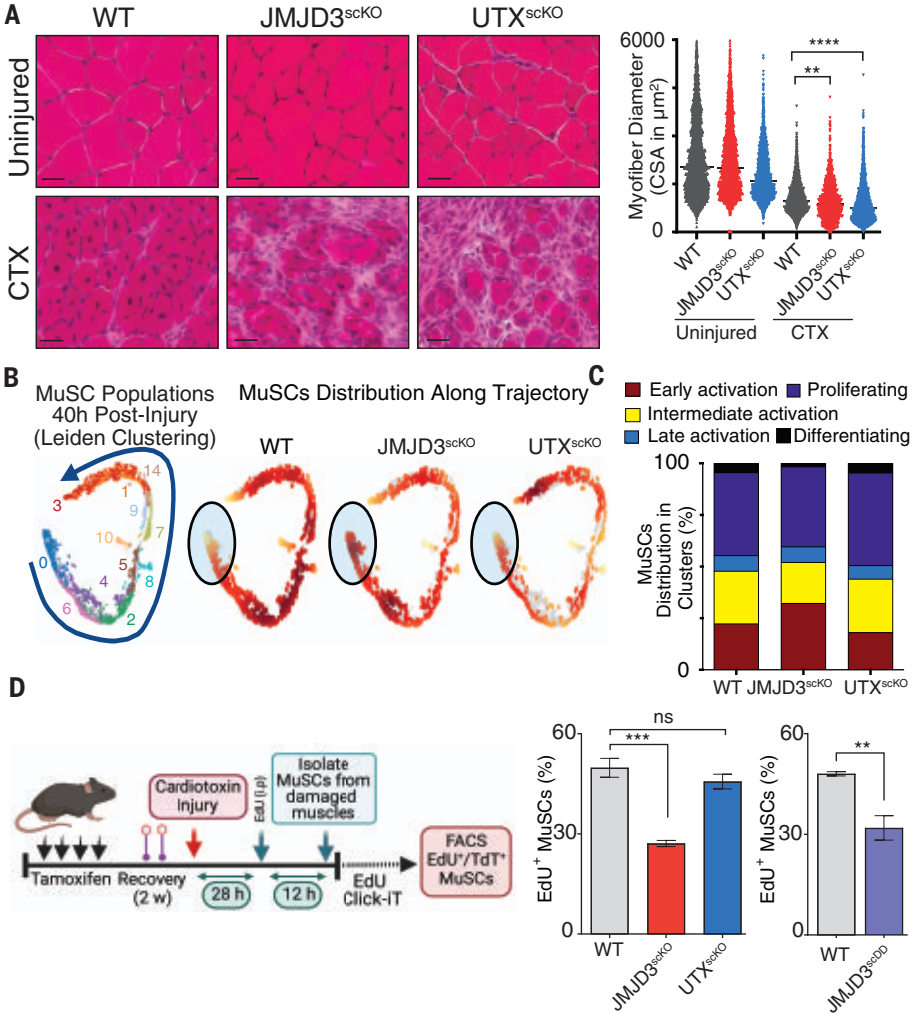


Fig. 1. JMJD3 and UTX play nonredundant roles in muscle regeneration where JMJD3 is required for injury-induced activation of satellite cells. (**A**) Hematoxylin and eosin staining of TA muscle cross sections from wild-type (WT), JMJD3^{scKO}, and UTX^{scKO} mice at 7 days after injury (CTX). Regeneration was quantified by measuring myofiber diameter. Data are means ± SD, and $N = 3$. $^{****}p < 0.0001$, and $^{**}p < 0.01$ [by analysis of variance (ANOVA)]. Scale bars are 60 µm. CSA, cross-sectional area. (**B**) Clustering and trajectory analysis of combined scRNA-seq data from lineage marked MuSCs [isolated based on TdTomato expression (TdT$^+$)] from the WT, JMJD3^{scKO}, and UTX^{scKO} mice isolated 40 hours after injury. Numbers indicate distinct MuSC clusters identified. The ovals highlight the enrichment of cells in the immediate-early activated MuSC state. (**C**) Distribution of MuSCs in clusters representative of different stages of the regenerative process. (**D**) Activation of MuSCs was measured by using in vivo EdU incorporation to measure the first passage of cells through the S phase of the cell cycle between 24 and 40 hours after injury. Fluorescence-activated cell sorting (FACS) analysis identified MuSCs (TdT$^+$) that were positive for EdU. Data are means ± SD, and $N = 3$; $^{***}p < 0.001$, $^{**}p < 0.01$, and ns is not significant (by Student's t test). i.p., intraperitoneally.

that JMJD3-mediated removal of H3K27me3 may be required for MuSCs to integrate signals from the regenerative environment. To test this, we performed TA muscle injury in one leg of JMJD3^{scKO} mice and then waited 36 hours before measuring cell cycle reentry of MuSCs isolated from the contralateral leg. MuSCs lacking JMJD3 were impaired in their ability to reenter the cell cycle when recovering from a distal injury (Fig. 2A and fig. S8). Thus, circulating factors that originate in the injured muscle must be responsible for preventing MuSCs that lack JMJD3 from reentering the cell cycle. A soluble extract prepared from an injured wild-type muscle (dMusEx) could also inhibit cell cycle reentry of purified MuSCs that lack JMJD3 (Fig. 2B and fig. S9), which shows that the circulating factors that inhibit MuSC activation are released from injured muscles independent of JMJD3. Thus, JMJD3 facilitates MuSC activation in a non–cell autonomous manner by overcoming an inhibition signal that originates in the regenerative niche. Using an extract prepared from mice with dystrophic muscle (MdxMusEx), we also found that JMJD3 is required for MuSC activation in conditions of chronic degeneration (Fig. 2B), highlighting the importance of JMJD3 for mediating activation of MuSCs in muscular dystrophy.

To identify JMJD3 target genes in MuSCs, we used a combination of genomics and transcriptomics (Fig. 3, figs. S10 and S11, table S1, and supplementary text). We defined JMJD3 direct targets based on three properties: (i) decreased expression in JMJD3^{scKO} MuSCs, (ii) JMJD3 binding within 2500 base pairs of the promoter, and (iii) accumulation of H3K27me3 marks near the gene upon ablation of JMJD3. Of the 41 genes that met these criteria (table S1), we focused on those coding for extracellular matrix (ECM)–related proteins because they may facilitate communication with the regenerative milieu. *Has2* (Fig. 3) was of particular interest because it encodes a key enzyme involved in the synthesis of hyaluronic acid (HA), a glycosaminoglycan polymer that incorporates into the ECM to alter rigidity (*14*) while acting as a rheostat for various signaling pathways (*15*). By reanalyzing published datasets (*16*, *17*), we found that *Has2* is not expressed in quiescent MuSCs (fig. S12, A and B) but is up-regulated in injured muscle as MuSCs reenter the cell cycle. In regenerating muscle 40 hours after injury, HA enveloped the activated MuSCs of wild-type or dystrophic mice while also contributing to the ECM of the necrotic myofibers (Fig. 4A and fig. S13, A to D). MuSCs that lacked JMJD3 were devoid of cell-enveloping HA, even though HA was present in the ECM of necrotic myofibers (Fig. 4A and fig. S13). This dynamic incorporation of HA into the ECM of MuSCs suggested that it may play an important role in regeneration of healthy muscle. To test this, we used the chemical inhibitor 4-methylumbelliferone (4-MU) to inhibit HA production (*18*) in wild-type MuSCs. Whereas purified wild-type MuSCs (from uninjured mice) were able to efficiently exit quiescence in the presence of 4-MU (Fig. 4B and fig. S14), the same MuSCs incubated with

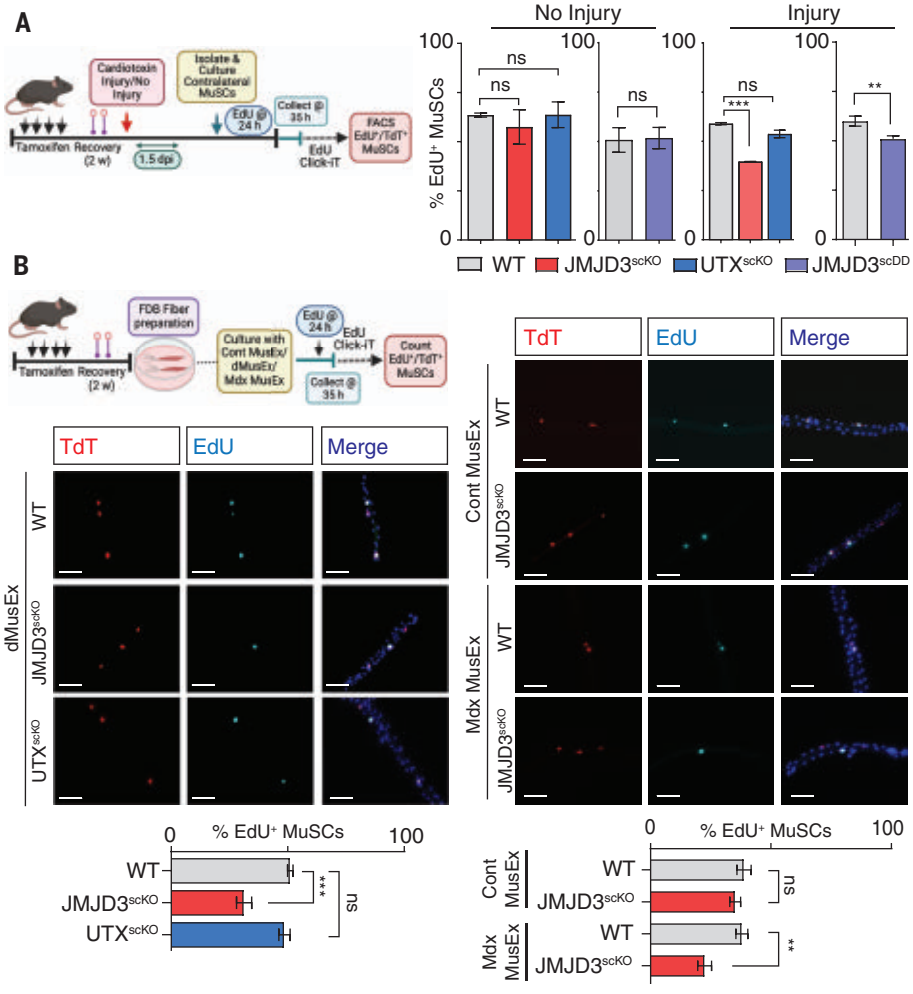


Fig. 2. JMJD3 mediates MuSC activation in a non–cell autonomous manner. (**A**) Purified MuSCs (TdT$^+$) from injured (CTX injury to contralateral leg 36 hours before MuSC isolation from uninjured muscles) or uninjured mice were assayed in vitro using EdU for cell cycle reentry between 24 and 40 hours after injury. FACS analysis identified MuSCs (TdT$^+$) that were positive for EdU. Data are means ± SD, and $N = 3$; ***$p < 0.001$, **$p < 0.01$, and ns is not significant (by Student's t test). dpi, days postinjury. (**B**) Ex vivo activation assay where flexor digitorum brevis (FDB) myofiber–associated MuSCs were treated with soluble extracts prepared from uninjured (ContMusEx), injured (dMusEx), or dystrophic (MdxMusEx) muscle. Immunofluorescence analysis identified MuSCs (TdT$^+$) that were positive for EdU. Data are means ± SD, and $N = 3$; ***$p < 0.001$, **$p < 0.01$, and ns is not significant (by Student's t test). Scale bars are 50 μm.

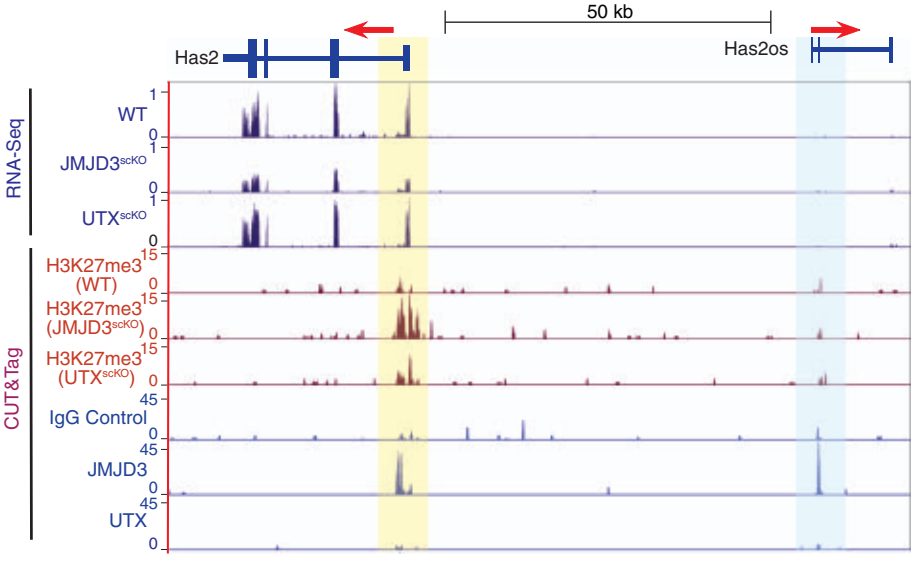


Fig. 3. *Has2* is a direct transcriptional target of JMJD3. UCSC (University of California Santa Cruz) browser track showing the *Has2* locus in MuSCs at 30 hours after CTX injury. Plots include RNA-seq data for WT, JMJD3^{scKO}, and UTX^{scKO} mice; CUT&Tag analysis shows enrichment of JMJD3, UTX, and H3K27me3.

an extract from damaged muscle showed impaired cell cycle reentry upon 4-MU treatment (Fig. 4B and supplementary text). Thus, HA synthesis is required for wild-type MuSCs to repair muscle after injury. Addition of exogenous HA to cultures was sufficient for JMJD3^{scKO} MuSCs to overcome an inhibition to cell cycle reentry (Fig. 4C and figs. S15 to S17), further emphasizing the importance of HA incorporation into the ECM of MuSCs to initiate population expansion in response to signals from the regenerative niche.

Finally, we explored the nature of the niche-derived signals that are responsible for impeding cell cycle reentry of MuSCs that lack

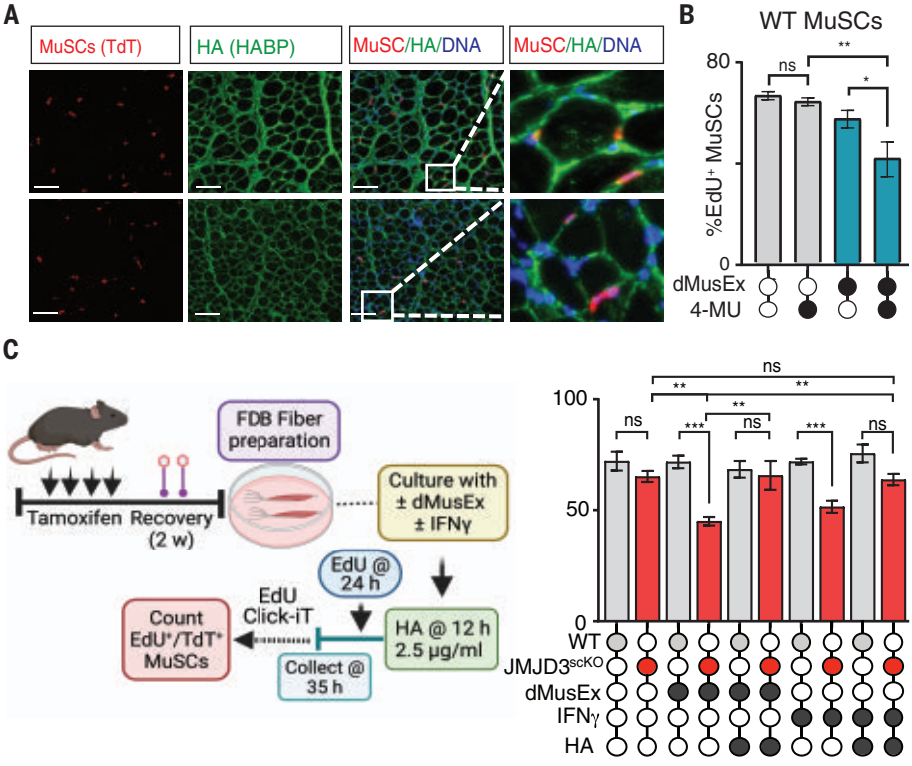


Fig. 4. HA incorporation into the ECM of MuSCs allows exit of quiescence. (**A**) Immunofluorescence analysis of TA muscle cross sections at 72 hours after CTX injury from WT and JMJD3^{scKO} mice. MuSCs (TdT$^+$), HA [hyaluronic acid binding protein (HABP)], and nuclear DNA [4′,6-diamidino-2-phenylindole (DAPI)] are shown. Scale bars are 75 μm. (**B**) Activation of MuSCs from WT mice was measured in the presence of 4-MU by in vitro EdU incorporation. An extract from damaged WT mouse muscle (dMusEx) was added as indicated. FACS analysis identified MuSCs (TdT$^+$) that were positive for EdU. Open circles indicate that the variable was not included in the experimental samples of this condition; filled circles indicate that the variable was included in the experimental samples of this condition. Data are means ± SD, and $N = 3$; $^{**}p < 0.01$, $^{*}p < 0.05$, and ns is not significant (by Student's t test). (**C**) Ex vivo activation assay of myofiber-associated MuSCs. FDB myofibers isolated from uninjured mice were incubated with either a damaged muscle extract (dMusEx) or recombinant IFN-γ protein. HA was added to cultures where indicated. Immunofluorescence analysis identified MuSCs (TdT$^+$) that were positive for EdU. Open circles indicate that the variable was not included in the experimental samples of this condition; filled circles indicate that the variable was included in the experimental samples of this condition. Data are means ± SD, and $N = 3$; $^{***}p < 0.001$, $^{**}p < 0.01$, and ns is not significant (by Student's t test).

JMJD3. Gene set enrichment analysis (GSEA) of RNA sequencing (RNA-seq) data identified the interferon-γ (IFN-γ), interleukin-6 (IL-6), and tumor necrosis factor–α (TNF-α) cellular responses as being up-regulated in MuSCs from JMJD3^{scKO} mice (fig. S18). Using recombinant cytokines, addition of exogenous IFN-γ or IL-6 [but not TNF-α, transforming growth factor–β (TGF-β), or IL-4] to the culture medium was sufficient to impair cell cycle reentry of JMJD3^{scKO} MuSCs (Fig. 4C, figs. S19 and S20, and supplementary text). Macrophages (Ly6C$^+$ cells) that are present within the regenerative niche at 24 hours after injury express both IFN-γ and IL-6 (fig. S21) and likely contribute to the signaling that impairs activation of MuSCs. This proinflammatory signaling is likely reinforced by neutrophils that are known to secrete IFN-γ and IL-6 in regenerating muscle (*19*). We propose that proinflammatory cytokines (including IFN-γ and IL-6) produced by immune cells from the regenerative environment play a critical role in preventing MuSCs from undergoing an untimely exit of quiescence. To overcome the cytokine-mediated block to stem cell function, MuSCs initiate *Has2* expression in a JMJD3-dependent manner, where incorporation of HA into the remodeled ECM renders the stem cells competent to receive proregenerative signals. Thus, HA has both anti-inflammatory effects on injured tissues (*20*) and proregenerative activity that stimulates MuSC-mediated regeneration.

Our study has revealed a role for the epigenetic enzyme JMJD3 in directing MuSC adaption to the regenerative niche by facilitating expression of JMJD3 target gene *Has2*. The resulting production of HA allows MuSCs to integrate proregenerative signaling from the local environment to facilitate the repair of injured muscle.

REFERENCES AND NOTES

1. F. Relaix *et al.*, *Nat. Commun.* **12**, 692 (2021).
2. M. Wosczyna, T. Rando, *Dev. Cell* **46**, 135–143 (2018).
3. B. Evano, S. Khalilian, G. Le Carrou, G. Almouzni, S. Tajbakhsh, *Cell Rep.* **30**, 3195–3206.e7 (2020).
4. D. C. L. Robinson *et al.*, *Dev. Cell* **56**, 1014–1029.e7 (2021).
5. D. Ratnayake *et al.*, *Nature* **591**, 281–287 (2021).
6. L. Machado *et al.*, *Cell Stem Cell* **28**, 1125–1135.e7 (2021).
7. S. Schwörer *et al.*, *Nature* **540**, 428–432 (2016).
8. K. Agger *et al.*, *Nature* **449**, 731–734 (2007).
9. S. Hong *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18439–18444 (2007).
10. R. Cao *et al.*, *Science* **298**, 1039–1043 (2002).
11. M. Brand, K. Nakka, J. Zhu, F. J. Dilworth, *Cell Stem Cell* **24**, 518–533 (2019).
12. K. B. Shpargel, J. Starmer, D. Yee, M. Pohlers, T. Magnuson, *PLOS Genet.* **10**, e1004507 (2014).
13. H. Faralli *et al.*, *J. Clin. Invest.* **126**, 1555–1565 (2016).
14. H. Ge, M. Tian, Q. Pei, F. Tan, H. Pei, *Front. Oncol.* **11**, 631991 (2021).
15. E. Puré, R. K. Assoian, *Cell. Signal.* **21**, 651–655 (2009).
16. C. A. Aguilar *et al.*, *Sci. Rep.* **5**, 13885 (2015).
17. L. Yue, R. Wan, S. Luan, W. Zeng, T. H. Cheung, *Dev. Cell* **53**, 661–676.e6 (2020).
18. N. Nagy *et al.*, *J. Biol. Chem.* **294**, 7864–7877 (2019).
19. W. Yang, P. Hu, *Cell. Immunol.* **326**, 2–7 (2018).
20. A. Marinho, C. Nunes, S. Reis, *Biomolecules* **11**, 1518 (2021).

ACKNOWLEDGMENTS

We thank D. Carragher, P. Gilbert, and S. Davoudi for insightful discussions; T. Magnuson for JMJD3$^{fl/fl}$ mice; O. Neyret for CUT&Tag and RNA-seq library sequencing; K. Sheikheleslamy for scRNA-seq library sequencing; uOttawa CBIA core for microscope assistance; F. Ortiz for FACS assistance; and L. Li for technical help. **Funding:** This work was supported by Canadian Institutes of Health Research grant FDN-143330 (F.J.D.), Canadian Institutes of Health Research grant PJT-156120 (L.A.M.), Canadian Institutes of Health Research grant MOP-343603 (M.B.), and the National Institutes of Health Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases (K.G.). **Author contributions:** Conceptualization: K.N., F.J.D.; Methodology: G.X., C.L., K.G., M.F.; Investigation: K.N., S.H., Y.L., C.B.; Formal analysis: Z.M., H.B., R.K.; Visualization: K.N., F.J.D., Z.M., H.B.; Supervision: F.J.D., K.G., M.B., X.Z., L.A.M., B.C., J.G.; Writing – original draft: F.J.D., K.N.; Writing – review and editing: F.J.D., K.N., M.B., K.G., L.A.M., X.Z., B.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The RNA-seq, scRNA-seq, and CUT&Tag datasets are available at the Gene Expression Omnibus using accession number GSE186833. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm9735
Materials and Methods
Supplementary Text
Figs. S1 to S21
Table S1
References (*21*–*36*)
MDAR Reproducibility Checklist

Submitted 1 November 2021; accepted 27 June 2022
10.1126/science.abm9735

QUANTUM DYNAMICS

Observation of a continuous time crystal

Phatthamon Kongkhambut[1], Jim Skulte[1,2], Ludwig Mathey[1,2], Jayson G. Cosme[3], Andreas Hemmerich[1,2]*, Hans Keßler[1]*

Time crystals are classified as discrete or continuous depending on whether they spontaneously break discrete or continuous time translation symmetry. Although discrete time crystals have been extensively studied in periodically driven systems, the experimental realization of a continuous time crystal is still pending. We report the observation of a limit cycle phase in a continuously pumped dissipative atom-cavity system that is characterized by emergent oscillations in the intracavity photon number. The phase of the oscillation was found to be random for different realizations, and hence, this dynamical many-body state breaks continuous time translation symmetry spontaneously. Furthermore, the observed limit cycles are robust against temporal perturbations and therefore demonstrate the realization of a continuous time crystal.

Time crystals are dynamical many-body states that break time translation symmetry in a spontaneous and robust manner (*1*, *2*). The original quantum time crystal envisaged by Wilczek involves a closed many-body system with all-to-all coupling that breaks continuous time translation symmetry by exhibiting oscillatory dynamics in its lowest-energy equilibrium state even though the underlying Hamiltonian is time-independent (*1*). This would constitute a startling state of matter in motion, fundamentally protected from bringing this motion to a standstill through energy removal. However, a series of no-go theorems have shown that nature prohibits the realization of such time crystals in isolated systems (*3*–*5*). The search for time crystals was thus extended to include equilibrium scenarios in periodically driven closed systems (*6*–*8*). This has led to realizations of discrete time crystals, which break the discrete time translation symmetry imposed by the external drive (*9*–*17*). In such discrete time crystals, during a short initial phase, the drive slightly excites the system until the system decouples from the drive, so that further energy or entropy flow is terminated. The system develops a subharmonic response, an intrinsic oscillation at a frequency slower than that of the drive. Initially, it was argued that dissipation, and hence the use of open systems, must be carefully avoided; then, so-called dissipative discrete time crystals were theoretically predicted (*18*) and experimentally realized (*19*–*21*). As shown in a number of theoretical works (*22*–*24*), the use of open systems comes with the unexpected consequence that continuous instead of periodic driving suffices to induce time crystal dynamics. These continuous time crystals (CTCs) realize the spirit of the original proposal more closely than discrete time crystals and circumvent the no-go theorems through their open character.

Here, we report the observation of a CTC in the form of a limit cycle phase in a continuously pumped dissipative atom-cavity system (Fig. 1A). In classical nonlinear dynamics, the term "limit cycle", coined by Poincaré in a mathematical context (*25*), denotes a closed phase space trajectory, asymptotically approached by at least one neighboring trajectory. Although limit cycles are well-established in classical nonlinear physics (*26*), there are two essential conditions for limit cycles in open quantum systems to form a CTC. First, the formation of the limit cycle must be associated with spontaneous breaking of continuous time translation symmetry. That is, the relative time phase of the oscillations for repeated realizations takes random values between 0 and 2π. Second, the limit cycle phase is robust against temporal perturbations of technical or fundamental character, such as quantum noise and, for open systems, fluctuations associated with dissipation. The characteristic signature of the CTC presented here is a persistent oscillation of the intracavity intensity and atomic density (Fig. 1, B and C), which complies with the robustness and spontaneous symmetry-breaking criteria (Fig. 1D).

Our experimental setup consists of a Bose-Einstein condensate (BEC) of $N_a \approx 5 \times 10^4$ ^{87}Rb atoms inside a high-finesse optical cavity. The

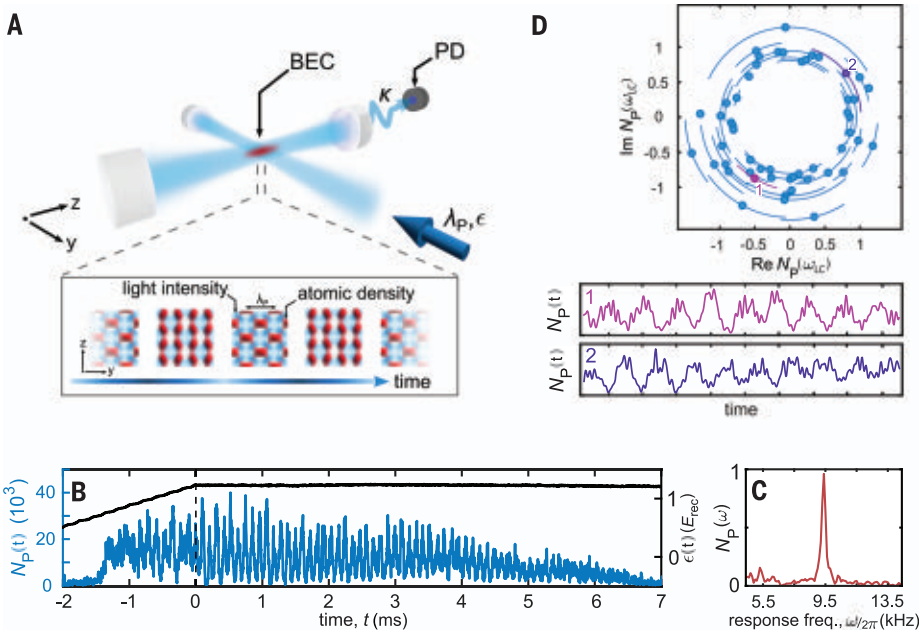


Fig. 1. CTC in an atom-cavity system. (**A**) Schematic drawing of the atom-cavity system pumped transversely with an optical pump lattice, blue detuned with respect to an atomic transition. (Inset) The photon field (blue) and the atomic density (red) of the limit cycle dynamics, based on simulations. The blue color shading of the time axis indicates the intracavity photon number. (**B**) Single experimental realization of the limit cycle phase for $\delta_{eff}/2\pi = -3.8$ kHz and $\varepsilon_f = 1.25\ E_{rec}$. The vertical dashed black line indicates the start of the 10 ms holding time, during which the pump strength is held constant. The black line indicates the time trace of the pump strength ε, and the blue line indicates the time evolution of the intracavity photon number $N_P(t)$. (**C**) Normalized and rescaled single-sided amplitude spectrum of N_P calculated from the data shown in (B). (**D**) (Top) Distribution of the time phase in the limit cycle phase for $\delta_{eff}/2\pi = -5.0$ kHz and $\varepsilon_f = 1.25\ E_{rec}$. The error bars indicate the phase uncertainty within our discrete Fourier transform resolution of 100 Hz. However, the uncertainty with regard to the radial dimension—the amplitude uncertainty—is negligibly small. For clarity, we removed the errors bars, around 30%, which are overlapping. (Bottom) The evolution of the intracavity photon number for two specific experimental realizations, marked with "1" and "2" at top, which have a time phase difference of almost π.

[1]Zentrum für Optische Quantentechnologien and Institut für Laser-Physik, Universität Hamburg, 22761 Hamburg, Germany. [2]The Hamburg Center for Ultrafast Imaging, 22761 Hamburg, Germany. [3]National Institute of Physics, University of the Philippines, Diliman, Quezon City 1101, Philippines.
*Corresponding author. Email: hemmerich@physnet.uni-hamburg.de (A.H.); hkessler@physnet.uni-hamburg.de (H.K.)

system is transversely pumped with a standing wave field with a wavelength λ_p = 792.55 nm (Fig. 1A). This wavelength is blue detuned with respect to relevant atomic D_1 transition of ^{87}Rb at a wavelength of 794.98 nm. The cavity operates in the recoil resolved regime (*27*)—its field decay rate $\kappa = 2\pi \times 3.4$ kHz is smaller than the recoil frequency $\omega_{rec} = 2\pi \times 3.7$ kHz. The cavity resonance frequency ω_c is shifted because of the refractive index of the BEC by an amount of $\delta_- = N_a U_0/2$, where $U_0 = 2\pi \times 1.3$ Hz is the maximal light shift per intracavity photon. We define the effective detuning as $\delta_{eff} \equiv \delta_c - \delta_-$, where $\delta_c \equiv \omega_p - \omega_c$ is the detuning between the pump field frequency ω_p and the resonance frequency of the empty cavity ω_c.

To determine the regime of the CTC, we measured the time dependence of the intracavity photon number $N_P(t)$ that emerges in the protocol given below. We show $N_P(t)$ in Fig. 2A and two derived quantities, the crystalline fraction Ξ and the limit-cycle frequency ω_{LC}, in Fig. 2, B and C, respectively. In our protocol, the intracavity photon number $N_P(t)$ was recorded as we linearly ramped the pump strength ε from 0 to 3.5 E_{rec} within 10 ms, while keeping δ_{eff} fixed. Initially, for weak pump intensities, the BEC phase was stable, and N_P was zero. Above a critical value of ε, the BEC became unstable toward the formation of a self-organized superradiant phase heralded by a nonzero N_P. This represents a many-body state as the cavity photons mediate a retarded infinite-range interaction between the atoms. Although this superradiant phase transition has been intensively studied for a red-detuned pump (*28*–*31*), it has only been realized recently for a blue-detuned pump after its theoretical prediction (*32*, *33*). For blue detuning, the atoms are low-field seeking and localize at the intensity minima of the light field. Nevertheless, the atoms can still self-organize into the superradiant phase, as evident from the large blue areas shown in Fig. 2A. However, the self-organized superradiant phase may become unstable for higher pump strengths because it costs energy for the atoms to localize away from the nodes of the pump lattice. This behavior leads to the disappearance of the self-organized phase for higher pump strengths (*32*). A phase diagram in fig. S1 in (*34*) shows a larger range of ε, demonstrating the disappearance of the self-organization for strong pumping. In the recoil-resolved regime, because of the retarded character of the cavity-mediated interaction, we additionally observed the emergence of a new dynamical phase or a limit cycle phase characterized by self-sustained oscillations of N_P as the atoms cycled through different density wave patterns (*33*, *35*). The resolution of the experimental imaging system is insufficient to observe the real-space density of the cloud; instead, simulations of the evolution of the single-particle density by use of a mean-field model are shown in fig. S3 (*34*). Physically, the limit cycles can be understood as a competition between opposing energy contributions: one coming from the pump lattice potential, and another coming from the cavity-induced all-to-all interaction between the atoms (*33*). In the superradiant phase, the cavity-induced interaction energy dominates, and the atoms localize at the antinodes. In the limit cycle phase for sufficiently strong pump intensities, localization of low-field–seeking atoms at the antinodes becomes energetically costly, resulting in a decrease in the density modulations and N_P as the system attempts to go back to the normal homogeneous phase. However, this is unstable toward self-organization because the chosen pump strength already exceeds the critical value, and thus, the cycle starts anew. The regime of recoil-resolution of the cavity, in which the dynamics of the atomic density and the light field evolve with similar time scales, has turned out to be the key ingredient to realize the limit cycle phase. This can be understood by noting that the delayed dynamics of the cavity field, with respect to the atomic density, leads to cavity

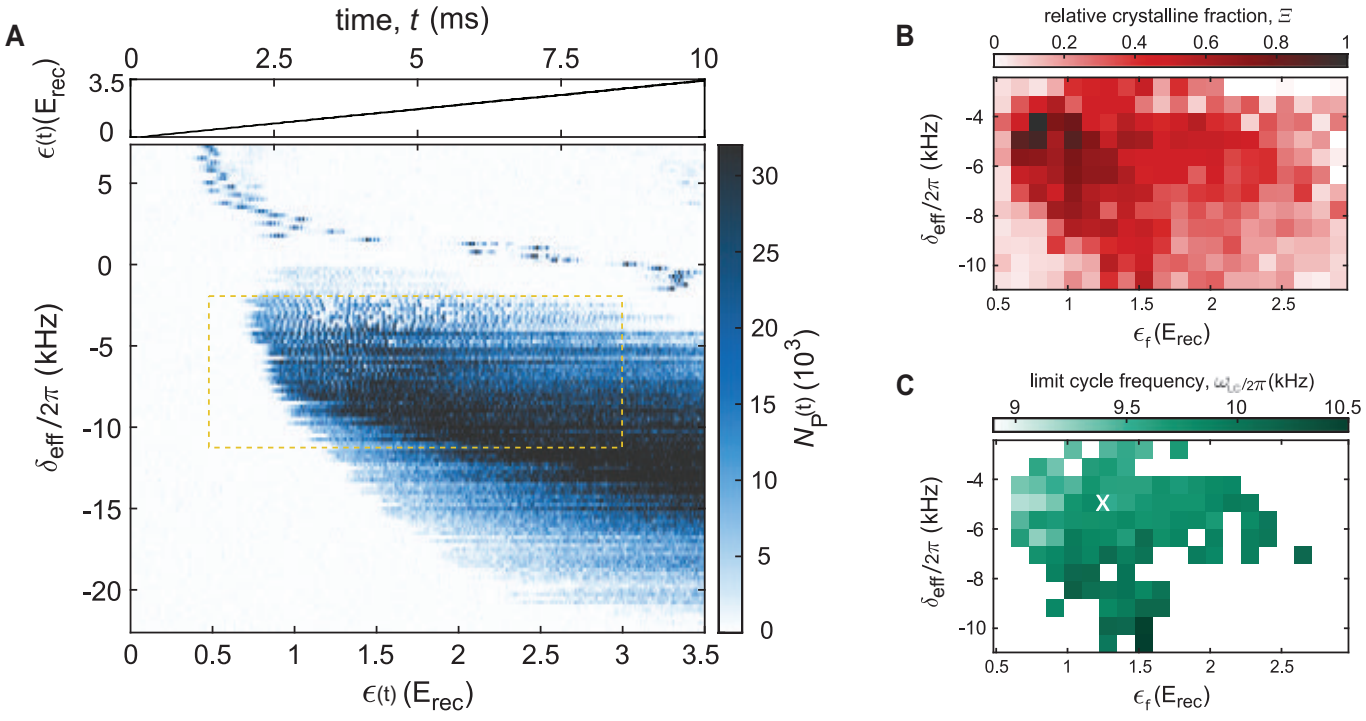


Fig. 2. Determining the time-crystalline regime. (**A**) (Top) Pump strength protocol. (Bottom) The corresponding intracavity photon number N_P as a function of δ_{eff} and ε. The area enclosed by the yellow dashed lines indicates the parameter space spanned in (B) and (C). (**B**) Relative crystalline fraction Ξ and (**C**) limit cycle frequency ω_{LC} plotted versus δ_{eff} and ε_f. To obtain (B) and (C), for fixed δ_{eff}, the pump strength is ramped to its final value ε_f and subsequently held constant for 10 ms. The relative crystalline fraction Ξ and the corresponding value of ω_{LC} identify the time-crystalline state. The parameter space is divided into 20 by 24 plaquettes and averages across 5 to 10 experimental implementations are produced. The white cross indicates the parameter values $\delta_{eff}/2\pi = -5.0$ kHz and $\varepsilon_f = 1.25\ E_{rec}$. The white area in (C) corresponds to data with Ξ below 1/*e*.

cooling, which in contrast to broadband cavity setups restricts the atoms to occupy only a small number of momentum modes. This prevents the system from heating up and entering chaotic dynamics. We observed the limit cycle phase in the region shown in Fig. 2A enclosed by the yellow dashed lines. To further highlight the dynamical nature of this phase, we show a typical single-shot realization in Fig. 1, B and C.

Next, we quantitatively identified the area in the parameter space, spanned by the pump strength ε and the effective detuning δ_{eff}, where limit cycles can be observed. For fixed δ_{eff}, we linearly ramped ε to the desired final value ε_{f}, using the same slope as for the measurement shown in Fig. 2A, and held ε constant for 10 ms. The protocol is depicted by the black curve in Fig. 1B. We show in Fig. 1C an example of the normalized and rescaled single-sided amplitude spectrum $N_{\mathrm{P}}(\omega) = \bar{N}_{\mathrm{P}}(\omega)/\bar{N}_{\mathrm{P,max}}(\omega_{\mathrm{LC}})$ obtained from $N_{\mathrm{P}}(t)$ within the holding time window [0,10] ms in Fig. 1B. $\bar{N}_{\mathrm{P}}(\omega)$ is the normalized single-sided amplitude spectrum, and $\bar{N}_{\mathrm{P,max}}(\omega_{\mathrm{LC}})$ is the maximum value of the measured limit cycle amplitude. In the case of pronounced limit cycle dynamics as in Fig. 1C, the single-sided amplitude spectrum shows a distinct peak, with a width associated with the limit cycle lifetime of several milliseconds. The narrowest peaks observed exhibit a e^{-2} width $\Delta\omega \approx 2\pi \times 1.4$ kHz: The limit cycle frequency ω_{LC}, plotted in Fig. 2C, is defined as the frequency of the dominant peak in the single-sided amplitude spectrum within the frequency interval $\Delta_{\mathrm{LC}} = [3.5,15.5] \times 2\pi$ Hz, chosen much larger than $\delta_{\mathrm{LC}} \in [\omega_{\mathrm{LC}} - \Delta\omega/2, \omega_{\mathrm{LC}} + \Delta\omega/2]$. The oscillation frequency of a CTC is not necessarily fixed, and robustness refers to the persistence of the CTC in the thermodynamic limit and for a wide range of system parameters [finite-size effects are discussed in the supplementary materials (*34*)] (*22*). We calculated a common measure for time crystallinity, the crystalline fraction Ξ' (*10*, *11*), as the ratio between the area under the single-sided amplitude spectrum within δ_{LC} and the total area within Δ_{LC}. That is, $\Xi' \equiv \sum_{\omega\in\delta_{\mathrm{LC}}} N_{\mathrm{P}}(\omega)/\sum_{\omega\in\Delta_{\mathrm{LC}}} N_{\mathrm{P}}(\omega)$. The relative crystalline fraction Ξ shown in Fig. 2B is normalized to the maximum crystalline fraction measured in the parameter space explored in this work. Because of the finite lifetime of the BEC, it is difficult to access the long-time behavior of the system, which makes it experimentally challenging to distinguish between the areas of stable limit cycle, chaos, and possible transient phases. Hence, we define a cut-off or threshold value for the relative crystalline fraction, $\Xi_{\mathrm{cut}} = 1/e$, to identify regions with observable limit cycle dynamics. In Fig. 2C, the frequency response of the limit cycle phase is only shown if its relative crystalline fraction is higher than the cut-off value: $\Xi > \Xi_{\mathrm{cut}}$. The experimental lifetime of our time crystal is limited by atom loss. Furthermore, the short-range contact interaction, due to collisions between the atoms, leads to dephasing of the system and hence melting of the time crystal. Simulations that include contact interactions and phenomenological atom loss can be found in the supplementary materials.

The spontaneous symmetry breaking of a many-body system indicates a phase transition. We demonstrated strong evidence that the limit cycle phase emerges through spontaneous breaking of continuous time translation symmetry, and thus, it is a CTC. We repeated the experimental pump protocol shown as the Fig. 1B black line more than 1500 times with fixed $\delta_{\mathrm{eff}}/2\pi = -5.0$ kHz and $\varepsilon_{\mathrm{f}} = 1.25\ E_{\mathrm{rec}}$. These parameter values are indicated in Fig. 2C with a white cross. Because of technical instabilities, the number of the atoms in the BEC N_{a} fluctuates by 5%. This leads to a fluctuating value of δ_{eff} and hence of ω_{LC}. Pictorially, this can be understood by observing that fluctuations in N_{a} effectively shift the CTC regime in Fig. 2C either up or down. For the parameter values indicated by a white cross in Fig. 2C, the median of ω_{LC} is $\bar{\omega}_{\mathrm{LC}} = 2\pi \times 9.69$ kHz. Our discrete Fourier transform resolution, set by the 10-ms time window, is 100 Hz. Thus, we only considered experimental runs, which yielded response frequencies of $\omega_{\mathrm{LC}} = \bar{\omega}_{\mathrm{LC}} \pm 2\pi \times (50\ \mathrm{Hz})$. For each single-shot measurement, we obtained the time phase defined as the principal argument $\arg[N_{\mathrm{P}}(\omega_{\mathrm{LC}})]$ of the Fourier transformed intracavity photon number $N_{\mathrm{P}}(\omega_{\mathrm{LC}})$ evaluated at the limit cycle frequency ω_{LC}. In Fig. 1D, we show the distribution of the observed time phases, which randomly covers the interval [0,2π]. This corroborates the spontaneous breaking of continuous time translation symmetry in the limit cycle phase. In the bottom of Fig. 1D, we show two specific experimental realizations, which have a time phase difference of almost π. Simulations representing the BEC as a coherent state show a range of the response frequency distribution of 300 Hz. Because we post-selected our data far below this limit, the origin of the spread over 2π in the time phase distribution is not due to technical noises but rather to quantum fluctuations. In the supplementary materials, we show a more detailed theoretical analysis to support this argument. The error bars along the angular direction in Fig. 1D indicate the phase uncertainty within 100 Hz of our Fourier limit. The average phase uncertainty is around 0.25π. The uncertainty in the radial direction corresponding to the oscillation amplitude is, however, negligible. Moreover, we removed 30% of the error bars for clarity in Fig. 1D.

Last, we demonstrated the robustness of the limit cycle phase against temporal perturbations, which is a defining feature of time crystals. We introduced white noise onto the pump signal with a bandwidth of 50 kHz. The noise strength is quantified by $n \equiv \sum_{\omega=0}^{2\pi\times 50\ \mathrm{kHz}} |\mathcal{A}_{\mathrm{noisy}}(\omega)| / \sum_{\omega=0}^{2\pi\times 50\ \mathrm{kHz}} |\mathcal{A}_{\mathrm{clean}}(\omega)| - 1$, where $\mathcal{A}_{\mathrm{noisy}}$ and $\mathcal{A}_{\mathrm{clean}}$ are the single-sided amplitude spectrum of the pump in the presence and absence of white noise, respectively. We chose the parameters $\delta_{\mathrm{eff}}/2\pi = -5.0$ kHz and $\varepsilon_{\mathrm{f}} = 1.25\ E_{\mathrm{rec}}$ in the center of the stable limit cycle region, indicated by the white cross in Fig. 2C, and added white noise with varying strengths. In Fig. 3, A and B, top, single-shot realizations of the noisy pump signal are shown for weak and strong noise, respectively. The corresponding dynamics of N_{P} is shown in Fig. 3, A and B, bottom. In Fig. 3E, we show how increasing the noise strength can "melt" the CTC as inferred by the decreasing relative crystalline fractions calculated from single-sided amplitude

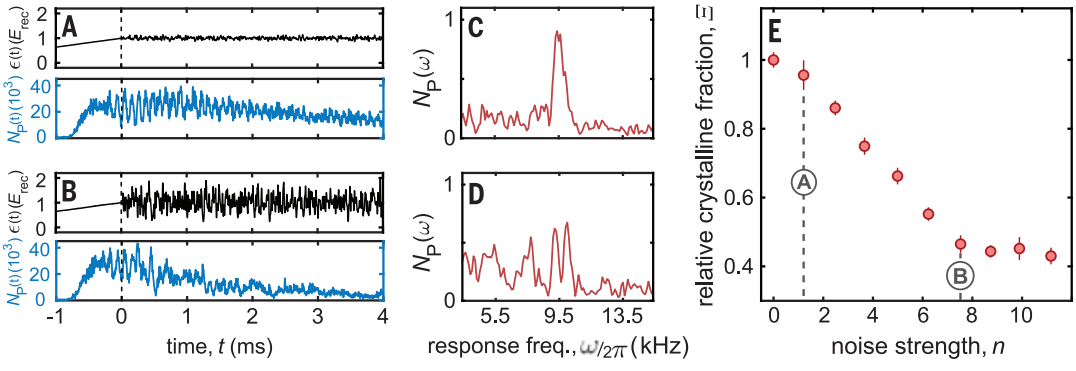


Fig. 3. Robustness against temporal perturbations. (**A** and **B**) Single experimental runs for noise strengths indicated in (E). (Top) Time traces of the pump strength ε. (Bottom) Corresponding dynamics of N_{P}. (**C** and **D**) Single-sided amplitude spectra of (A) and (B), respectively. (**E**) Relative crystalline fraction for varying noise strength n and fixed $\delta_{\mathrm{eff}}/2\pi = -5.0$ kHz and $\varepsilon_{\mathrm{f}} = 1.25\ E_{\mathrm{rec}}$.

spectra, similar to those shown in Fig. 3, C and D. The system takes time to react to the noise, so that a few oscillations can always be observed before decay sets in. This leads to an offset of 0.4 in the crystalline fraction, even for very strong noise. Nevertheless, we found that the limit cycle phase indeed exhibits robust oscillatory behavior over a wide range of the noise strength. This, together with the observation of spontaneous breaking of a continuous time translation symmetry, suggests that the observed limit cycle phase is a CTC.

We have experimentally demonstrated a CTC and provided a theoretical understanding. This class of dynamical many-body states expands the concepts of long-range order and spontaneous symmetry breaking into the time domain and is therefore of fundamental interest. This result, and the precision and control achieved with our atom-cavity platform, paves the way toward a broad and comprehensive study of dynamical many-body states of bosonic or fermionic quantum matter in the strongly correlated regime. For example, an increased atom-photon coupling could generate a new class of time crystals associated with symmetry-broken periodic entanglement. Furthermore, technological applications, such as toward time metrology, can be envisioned.

REFERENCES AND NOTES

1. F. Wilczek, *Phys. Rev. Lett.* **109**, 160401 (2012).
2. A. Shapere, F. Wilczek, *Phys. Rev. Lett.* **109**, 160402 (2012).
3. P. Nozières, *Europhys. Lett.* **103**, 57008 (2013).
4. P. Bruno, *Phys. Rev. Lett.* **111**, 070402 (2013).
5. H. Watanabe, M. Oshikawa, *Phys. Rev. Lett.* **114**, 251603 (2015).
6. D. V. Else, C. Monroe, C. Nayak, N. Y. Yao, *Annu. Rev. Condens. Matter Phys.* **11**, 467–499 (2020).
7. K. Sacha, *Time Crystals* (Springer, 2020).
8. V. Khemani, R. Moessner, S. L. Sondhi, arXiv:1910.10745 [cond-mat.str-el] (2019).
9. J. Zhang *et al*., *Nature* **543**, 217–220 (2017).
10. S. Choi *et al*., *Nature* **543**, 221–225 (2017).
11. J. Rovny, R. L. Blum, S. E. Barrett, *Phys. Rev. Lett.* **120**, 180603 (2018).
12. J. Smits, L. Liao, H. T. C. Stoof, P. van der Straten, *Phys. Rev. Lett.* **121**, 185301 (2018).
13. S. Autti, V. B. Eltsov, G. E. Volovik, *Phys. Rev. Lett.* **120**, 215301 (2018).
14. J. O'Sullivan *et al*., *New J. Phys.* **22**, 085001 (2020).
15. A. Kyprianidis *et al*., *Science* **372**, 1192–1196 (2021).
16. J. Randall *et al*., *Science* **374**, 1474–1478 (2021).
17. X. Mi *et al*., *Nature* **601**, 531–536 (2022).
18. Z. Gong, R. Hamazaki, M. Ueda, *Phys. Rev. Lett.* **120**, 040404 (2018).
19. H. Keßler *et al*., *Phys. Rev. Lett.* **127**, 043602 (2021).
20. P. Kongkhambut *et al*., *Phys. Rev. Lett.* **127**, 253601 (2021).
21. H. Taheri, A. B. Matsko, L. Maleki, K. Sacha, *Nat. Commun.* **13**, 848 (2022).
22. F. Iemini *et al*., *Phys. Rev. Lett.* **121**, 035301 (2018).
23. B. Buča, J. Tindall, D. Jaksch, *Nat. Commun.* **10**, 1730 (2019).
24. H. Keßler, J. G. Cosme, M. Hemmerling, L. Mathey, A. Hemmerich, *Phys. Rev. A* **99**, 053605 (2019).
25. H. Poincaré, *J. Math. Pures Appl.* **7**, 375 (1881).
26. S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Westview Press, 2015).
27. H. Keßler, J. Klinder, M. Wolke, A. Hemmerich, *New J. Phys.* **16**, 053008 (2014).
28. P. Domokos, H. Ritsch, *Phys. Rev. Lett.* **89**, 253003 (2002).
29. A. T. Black, H. W. Chan, V. Vuletić, *Phys. Rev. Lett.* **91**, 203001 (2003).
30. K. Baumann, C. Guerlin, F. Brennecke, T. Esslinger, *Nature* **464**, 1301–1306 (2010).
31. J. Klinder, H. Keßler, M. Wolke, L. Mathey, A. Hemmerich, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3290–3295 (2015).
32. P. Zupancic *et al*., *Phys. Rev. Lett.* **123**, 233601 (2019).
33. F. Piazza, H. Ritsch, *Phys. Rev. Lett.* **115**, 163601 (2015).
34. Supplementary text is available in the supplementary materials.
35. H. Keßler, J. G. Cosme, C. Georges, L. Mathey, A. Hemmerich, *New J. Phys.* **22**, 085002 (2020).
36. P. Kongkhambut *et al*., Data for "'Observation of a continuous time crystal". Zenodo (2022).

ACKNOWLEDGMENTS

H.K. thanks J. Klinder and C. Georges for helpful discussions and their support. J.G.C. thanks R. J. L. Tuquero for valuable insights and discussions. A.H. acknowledges useful discussions with C. Zimmermann and J. Marino. **Funding:** This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through grant DFG-KE2481/1–1. P.K., J.S., L.M., and A.H. acknowledge the DFG for funding through SFB-925–project 170620586, and the Cluster of Excellence "Advanced Imaging of Matter" (EXC 2056)–project 390715994. J.S. acknowledges support from the German Academic Scholarship Foundation.

Author contributions: P.K. and H.K. performed the experiments and data analysis. The simulations were performed by J.S. and J.G.C., supported by L.M. The project was designed and supervised by H.K. and A.H. All authors contributed to the discussion and interpretation of the results, as well as to writing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data presented in this paper and simulation scripts are deposited at Zenodo (*36*).

License information: Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo3382
Materials and Methods
Supplementary Text
Figs. S1 to S7
References (*37*, *38*)

Submitted 14 February 2022; accepted 25 May 2022
10.1126/science.abo3382

CRYSTALLIZATION

Synchronous assembly of chiral skeletal single-crystalline microvessels

Osamu Oki[1], Hiroshi Yamagishi[1]*, Yasuhiro Morisaki[2], Ryo Inoue[2], Kana Ogawa[2], Nanami Miki[2], Yasuo Norikane[1,3], Hiroyasu Sato[4], Yohei Yamamoto[1]*

Skeletal or concave polyhedral crystals appear in a variety of synthetic processes and natural environments. However, their morphology, size, and orientation are difficult to control because of their highly kinetic growth character. We report a methodology to achieve synchronous, uniaxial, and stepwise growth of micrometer-scale skeletal single crystals from planar-chiral double-decker molecules. Upon drop-casting of a heated ethanol solution onto a quartz substrate, the molecules spontaneously assemble into standing vessel-shaped single crystals uniaxially and synchronously over the wide area of the substrate, with small size polydispersity. The crystal edge is active even after consumption of the molecules and resumes stereoselective growth with successive feeding. The resultant morphology can be packed into polycyclic aromatic hydrocarbon–like microarchitectures and behaves as a microscopic container.

Concave polyhedral crystals, typically termed "skeletal crystals," are naturally and synthetically ubiquitous crystal forms that feature intricate morphologies with well-developed crystalline edges and facets instead of a body (*1*–*3*). The diffusion-limited growth and resultant intricate morphology are poorly controlled by means of conventional methodologies and are therefore of particular interest in the field of crystal engineering (*4*–*7*). Furthermore, as typically shown in snowflakes and bismuth, skeletal single crystals are not only complicated in their morphology but also feature geometric symmetry and homogeneous molecular arrangement throughout the entire crystalline grain, which distinguishes them from polycrystalline, semicrystalline, and amorphous solids (*8*–*11*).

Crystal engineering techniques for growing skeletal single crystals still lag behind their thermodynamic counterparts (*12*–*17*). Most efforts to sculpt skeletal structures have focused on amphiphilic surfactants added under thermodynamic conditions (*18*–*22*). These chemical additives are attached to specific crystal facets and regulate crystal growth, yielding spiky and concave grains. Although these methods are useful, the available morphologies are typically distinct from those developed under kinetic conditions. Controlling the growth kinetics of skeletal single crystals has been an even more difficult task because their synthetic conditions are far from equilibrium. We therefore explore whether fine modulation of the

[1]Department of Materials Science, Faculty of Pure and Applied Sciences, and Tsukuba Research Center for Energy Materials Science (TREMS), University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan. [2]Department of Applied Chemistry for Environment, School of Biological and Environmental Sciences, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan. [3]Research Institute for Advanced Electronics and Photonics, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki 305-8565, Japan. [4]Rigaku Corporation, 12-9-3 Matsubara, Akishima, Tokyo 196-8666, Japan.
***Corresponding author. Email: yamagishi.hiroshi.ff@u.tsukuba.ac.jp (H.Y.); yamamoto@ims.tsukuba.ac.jp (Y.Y.)**

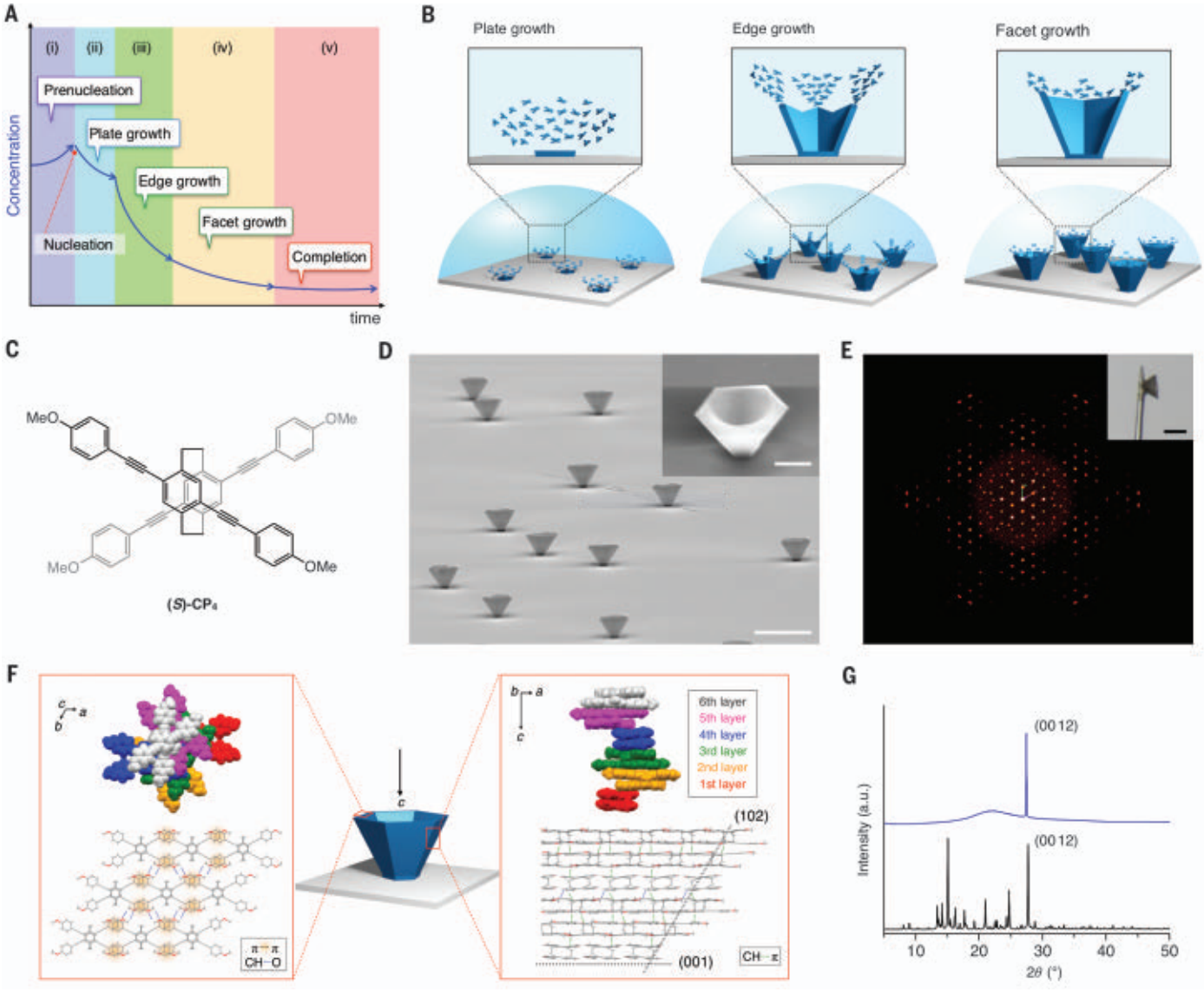


Fig. 1. Morphological and crystallographic investigation of skeletal vessel-shaped microcrystals. (**A**) Schematic diagram of the time-course profile of the concentration of **(*S*)-CP$_4$**. The concentration initially increases as the mother solvent evaporates (purple region) and peaks when nucleation and growth of the hexagonal plate begins (cyan region). The crystal growth regime changes into edge growth (green region) followed by facet growth (orange region), featuring a faster growth rate. After the molecules are mostly consumed, the crystal growth halts (red region). (**B**) Schematic representations of the synchronous and uniaxial skeletal crystal growth regimes of plate growth, edge growth, and facet growth, corresponding to (ii), (iii), and (iv) in Fig. 1A. (**C**) Molecular structure of **(*S*)-CP$_4$**. Me, methyl. (**D**) SEM images of the resultant microcrystals. The sample stage is tilted by 80° (main panel) and 50° (inset). Scale bars: 10 μm (main); 2 μm (inset). (**E**) Diffraction spots from a single vessel-shaped microcrystal projected onto the reciprocal lattice. The inset shows a photograph of the vessel-shaped microcrystal mounted on a thin needle for XRD measurement. Scale bar: 20 μm. (**F**) Schematic representations of the molecular packing arrangement of **(*S*)-CP$_4$** viewed from the *c*-axis (left) and *b*-axis (right) directions. The intermolecular interactions are visualized with orange circles (π–π) and blue (C–H···O) and green (C–H···π) dashed lines. The CPK models of **(*S*)-CP$_4$** molecules in each layer are colored red, orange, green, blue, magenta, and gray in this order. (**G**) PXRD pattern of microcrystals of **(*S*)-CP$_4$** grown on a quartz substrate (blue, top), and simulated XRD pattern from single-crystal XRD analysis (black, bottom). The broad peak originates from the amorphous quartz substrate. a.u., arbitrary units.

kinetic growth of skeletal crystals is feasible in the absence of chemical additives.

We report skeletal single crystals that grow in a synchronous, uniaxial, and stepwise manner within a few tens of seconds (Fig. 1, A and B). Skeletal crystals are grown from an *S*-planar-chiral aromatic molecule [2.2]paracyclophane appended with four (methoxyphenyl)ethynyl arms [**(*S*)-CP$_4$**], as well as its enantiomeric counterpart [**(*R*)-CP$_4$**] (Fig. 1C and fig. S1) (*23*, *24*). A heated solution of **(*S*)-CP$_4$** (0.10 mg in 100 μl of ethanol, 80°C) is gently cast on a clean amorphous quartz substrate under ambient condition and covered with a glass dish (fig. S2). Immediately after the casting, the EtOH solution is cooled by the substrate and the surrounding atmosphere and reaches supersaturation, resulting in the formation of micrometer-scale particles on the substrate.

Scanning electron microscopy (SEM) and fluorescence microscopy (FM) images reveal that the particles form a vessel-like concave morphology (Fig. 1D and fig. S3A). The base part of the vessel forms a hexagonal plate

with average side length and height of 0.8 and 0.3 μm, respectively (fig. S3, B to E). Planar lateral sidewalls grow outward with hexagonal symmetry on the hexagonal base plate with a dihedral angle of 60° relative to the substrate, forming a hexagonal pyramidal void inside the facets. The thickness of the sidewall (0.5 to 0.6 μm) is nearly constant from bottom to top (fig. S3, F to H).

A vessel-shaped particle is picked up with a microneedle and directly subjected to single-crystal x-ray structure analysis. The diffraction spots projected onto the reciprocal lattice show a hexagonally symmetric pattern (Fig. 1E). Concomitantly, the crystal structure of a single microvessel is successfully resolved. In the crystal, despite its apparent geometrical complexity, **(*S*)-CP$_4$** molecules are arranged in a highly symmetric manner with the polar space group $P6_5$. In the unit cell, six symmetrically equivalent **(*S*)-CP$_4$** molecules stack on one another with a counterclockwise rotation of 60° along a crystallographic sixfold screw axis (Fig. 1F, table S1, and fig. S4). Orthogonal to the screw axis, **(*S*)-CP$_4$** molecules form a unimolecular-thick double-decker aromatic sheet with all six aromatic rings virtually parallel to the crystallographic *ab* plane, with an average tilt angle of 6.94 ± 2.94°.

The **(*S*)-CP$_4$** molecules form complementary π–π and C–H···O interactions side by side within the double-decker sheet (Fig. 1F). The interlayers of the double-decker sheets stack on each other via multiple C–H···π and C–H···O interactions, although, counterintuitively, no obvious π–π interaction is found between the sheets (Fig. 1F and fig. S5). The plane-to-plane distances of the adjacent anisolyl rings are short (3.045 Å) but are largely displaced in the crystallographic *ab* plane and barely overlap. Structural analysis of **(*R*)-CP$_4$** is likewise accomplished, showing a mirror-image molecular arrangement with the space group $P6_1$ (table S2 and fig. S6).

The crystal structure, together with the simultaneously indexed crystal facets, reveals the molecular arrangement at the crystal surfaces and molecular interactions inside the microvessel (Fig. 1F). At the bottom facet of the vessel—namely, the crystallographic (001) plane—the aromatic rings of the **(*S*)-CP$_4$** molecules lie parallel to the quartz substrate. The lateral facets of the vessel—that is, the crystallographic {102} planes—are fully covered with the methoxy groups of **(*S*)-CP$_4$**, and this arrangement is beneficial to facilitating the solvophilic effect with EtOH. On the upper side of the vessel, concave facets are developed instead of the planar (00–1) facet so that **(*S*)-CP$_4$** molecules can conceal the nonpolar aromatic rings from the polar EtOH molecules.

Growth into vessel-shaped crystals occurs uniformly over a large area on the substrate surface with a narrow size distribution [average diameter of the hexagonal inscribed circle (d_{hic_av}): 4.3 μm, standard deviation (σ): 0.41 μm] (Fig. 2A and fig. S7). The powder x-ray diffraction (PXRD) profile of the microcrystals grown on the substrate exhibits a sharp, distinct diffraction peak at 27.5° (d = 3.2 Å), which is attributed to the (0012) plane (Fig. 1G and fig. S8). The observed single peak in the PXRD pattern unambiguously indicates the uniaxial growth of the vessel-shaped single crystals on the substrate.

The monodisperse size distribution of the resultant crystals suggests synchronous growth of the vessel crystals with a virtually identical rate. To address the growth kinetics, in situ visualization of the growing particles is conducted by means of FM (movie S1). At t = 0 s, immediately after the casting of the supersaturated EtOH solution, blue emission from **(*S*)-CP$_4$** dissolved in EtOH is homogeneously observed in the entire area. Nucleation of the hexagonal plate occurs at t = 2.0 s (Fig. 2C), followed by continuous growth until t = 12 s (Fig. 2D). At t = 12 s, **(*S*)-CP$_4$** dissolved in EtOH is almost consumed, and the growth is halted (Fig. 2E). Statistical analyses of d_{hic_av} of the crystals show that σ (<0.32 μm) and the polydispersity index (PDI; <0.02) of the particle size remain substantially small throughout the growth process (Fig. 2B).

We observe the morphologies of the growing particles by removing the mother solvent via a spin coater during the growth process. SEM images of the growing particles show spiky hexapodal edges standing on a thin hexagonal plate (fig. S9). Consequently, we conjecture how vessel-shaped growth is facilitated at the molecular scale [Figs. 1A and 3A, (ii)]. Immediately after casting, the temperature of the solution is still high, and crystal growth proceeds under pseudothermodynamic conditions, yielding a hexagonal plate with its face parallel to the substrate. Upon further cooling, the degree of supersaturation drastically increases, which alters the crystal growth regime to edge growth. Because the highly kinetic edge growth consumes the dissolved **(*S*)-CP$_4$** much faster than the cooling-induced increase in supersaturation, the growth regime changes into facet growth in a certain period of time, and the growth rate is slow enough to fill the gaps between the spikes but too fast to fill the inner void completely.

The transitions of the crystal growth regime are further supported by the in situ FM

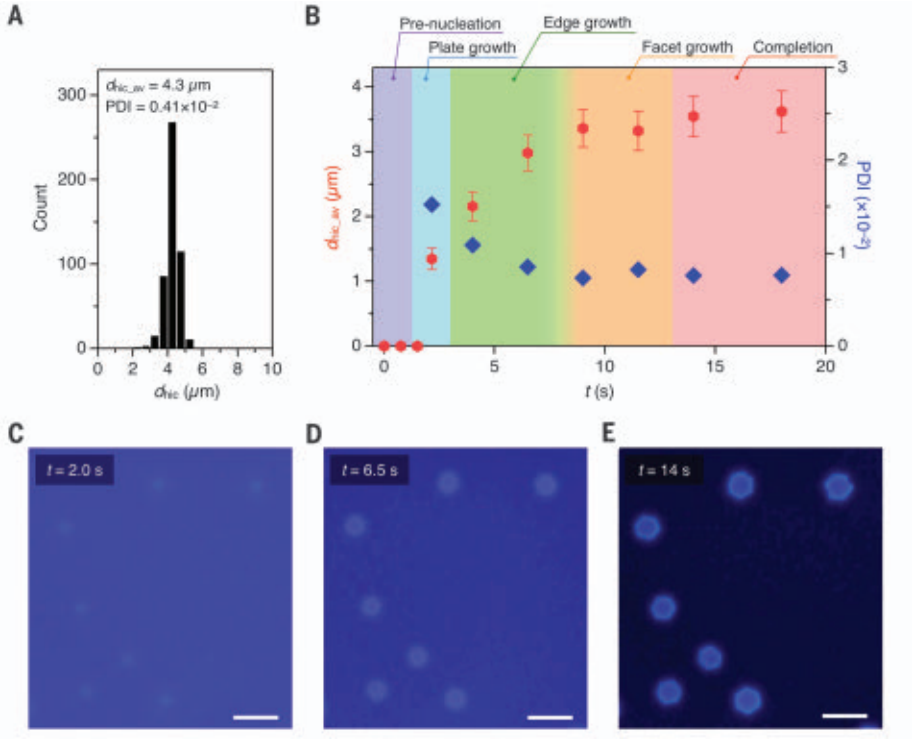


Fig. 2. Synchronous crystal growth. (**A**) Histogram of d_{hic} of the microcrystals measured from total selected areas I to V in fig. S7. (**B**) Plots of d_{hic_av} (red hexagons) and PDI (blue diamonds) of the microcrystals versus t. The standard deviation of d_{hic_av} was measured on the basis of 100 crystals found in each snapshot extracted from movie S1. Purple, cyan, green, orange, and red regions represent prenucleation, hexagonal plate growth, edge growth, facet growth, and completion of crystal growth, respectively. (**C** to **E**) Snapshots of FM images of microcrystals of **(*S*)-CP$_4$** at t = 2.0, 6.5, and 14 s after drop-casting of the EtOH solution. Scale bars: 10 μm. The excitation wavelength (λ_{ex}) = 340 to 390 nm.

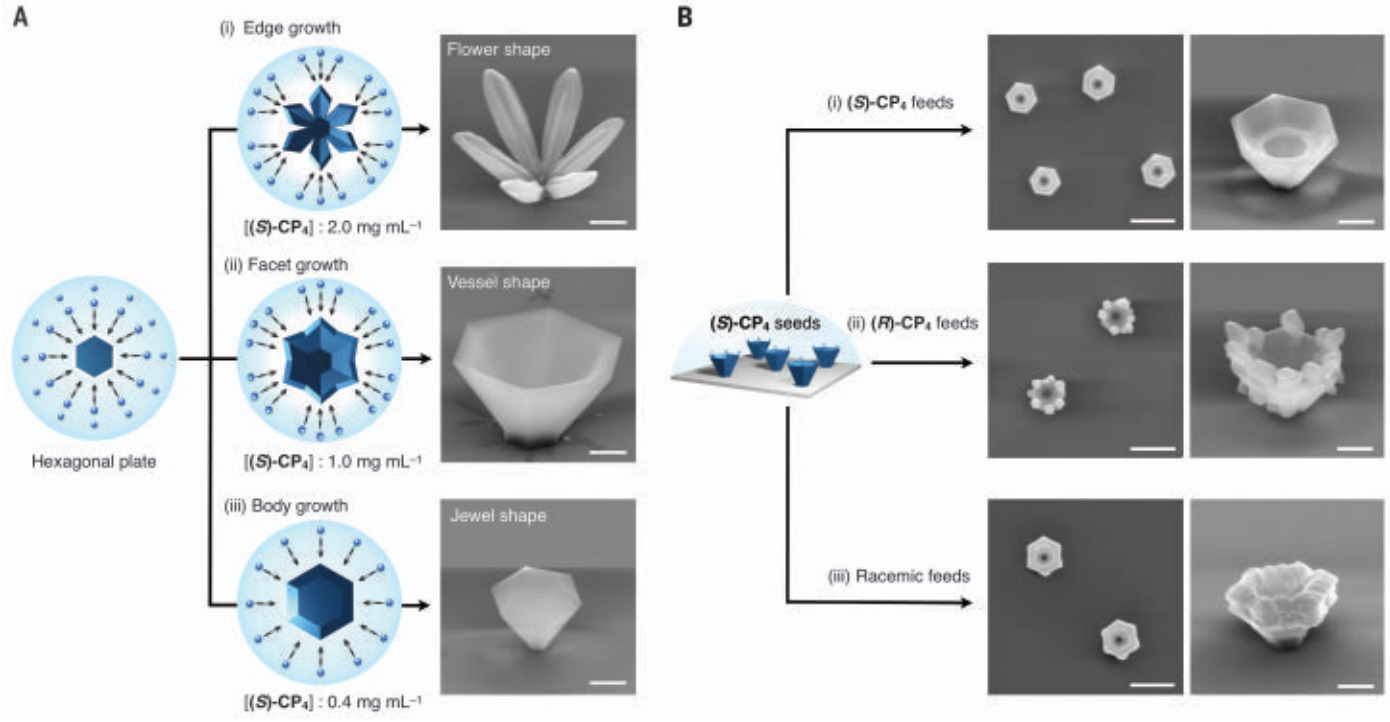


Fig. 3. Morphology control and hierarchical crystal growth. (**A**) Schematic representation of the (i) edge-, (ii) facet-, and (iii) body-growth crystallizations with respect to the different initial concentrations of **(*S*)-CP$_4$**, and corresponding SEM micrographs of the resultant microcrystals. Scale bars: 3 μm. (**B**) Scheme of the stepwise crystal growth and hierarchical microstructures after the addition of EtOH solutions of (i) **(*S*)-CP$_4$**, (ii) **(*R*)-CP$_4$**, and (iii) their racemic mixture to the vessel-shaped crystal seeds of **(*S*)-CP$_4$**, and corresponding SEM micrographs of the resultant microstructures. Scale bars: 20 μm (left) and 5 μm (right) for (i) and (ii); 10 μm (left) and 5 μm (right) for (iii).

observations (Fig. 2, C to E). In the snapshots, the blue emission in the background gradually fades and finally becomes dark as the crystals grow, representing a decrease in the concentration of **(*S*)-CP$_4$** in the focal plane. Accordingly, the concentrations of **(*S*)-CP$_4$** in the EtOH solution are evaluated during the crystal growth (fig. S10). The concentration initially increases because of the evaporation of EtOH and reaches a peak at t = 0.88 s, at which point nucleation occurs. Thereafter, as the hexagonal plates grow, the concentration decreases at a relatively slow rate. At t = 2.8 s, a transition to a faster decrease in concentration occurs, which is attributed to the change in the crystal growth regime from body growth to edge and facet growth (Fig. 1A).

This insight is highly informative because the transition from edge growth to facet growth takes place in an experimentally available concentration range. Accordingly, the initial concentration of **(*S*)-CP$_4$** in the feed EtOH solution is increased to 2.0 mg ml^{-1}, and the solution is cast onto a quartz substrate. After drying, the substate is covered by crystalline particles, featuring flower-like hexapodal branches oriented outward with C_6 symmetry [Fig. 3A, (i), and movie S2]. On the other hand, when the solution of **(*S*)-CP$_4$** is diluted to a concentration of 0.4 mg ml^{-1}, the resultant crystals feature a jewel-like convex polygonal morphology [Fig. 3A, (iii)], demonstrating comprehensive control of the growth regime from facet growth to edge and body growth by changing the initial concentration.

Conventional core/shell–type crystals have been synthesized by the successive feeding of the constituent molecules (*17*, *18*). Inspired by this technique, we conduct successive addition of fresh solution to explore whether the vessel crystals likewise grow incrementally. Fifteen seconds after casting of the first feed solution of **(*S*)-CP$_4$** (7.0 μl, 1.0 mg ml^{-1}) onto a substrate, the secondary feed solution is gently added at identical volume, concentration, and temperature (80°C). Although the exterior morphology of the resultant crystals still features a hexagonal vessel shape with smooth facets, stepped facets form in the interior, and d_{hic_av} increases from 7.0 to 11 μm, demonstrating an active nature of the crystalline edges at the top [Fig. 3B, (i)]. This successive crystal growth can be repeated at least five times to yield five-layered vessel-shaped microcrystals (fig. S11). Meanwhile, the single crystallinity of the microvessels is kept intact without lattice mismatch, as proven by the single-crystal structure analysis (figs. S12 and S13, and tables S3 and S4). The d_{hic_av} value of the microvessels observed by FM increases almost linearly as the feeding is repeated. Consistently, PXRD profiles exhibit a linear increase of peak intensity derived from the (0012) plane (fig. S14). These results suggest a synchronous growth rate, even in the sequential growth process. We also find that the inner void of the vessel can be partially filled by additionally feeding a diluted solution (0.4 mg ml^{-1}) to the vessel crystals (fig. S15).

We further find that the active edges of the vessels can differentiate the stereoisomer in the secondary growth process. When a solution of **(*R*)-CP$_4$** is applied as the secondary feed solution to vessel crystals of **(*S*)-CP$_4$**, irregular and granular crystals emerge at the vessel edges [Fig. 3B, (ii)]. The successive growth is not operative when a racemic mixture of **(*S*)-CP$_4$** and **(*R*)-CP$_4$** is applied, resulting in precipitation of ill-defined solids at the top edges of the vessel [Fig. 3B, (iii)].

The skeletal and dendritic crystals reported thus far are diverse in morphology (*4*–*7*). However, they are typically polydisperse in size and suspended in solvent. By contrast, the vessel-shaped crystals reported herein are uniform in size and are fixed on a substrate with their open windows facing up. We envisage that such a standing vessel can be used as a microscopic "vessel" that holds a tiny volume of liquid. As a proof of concept, a small piece of a solid 4,4′-dihexyloxy-3-methylazobenzene (AZO) crystal is placed inside a vessel. AZO liquifies immediately after the photoisomerization

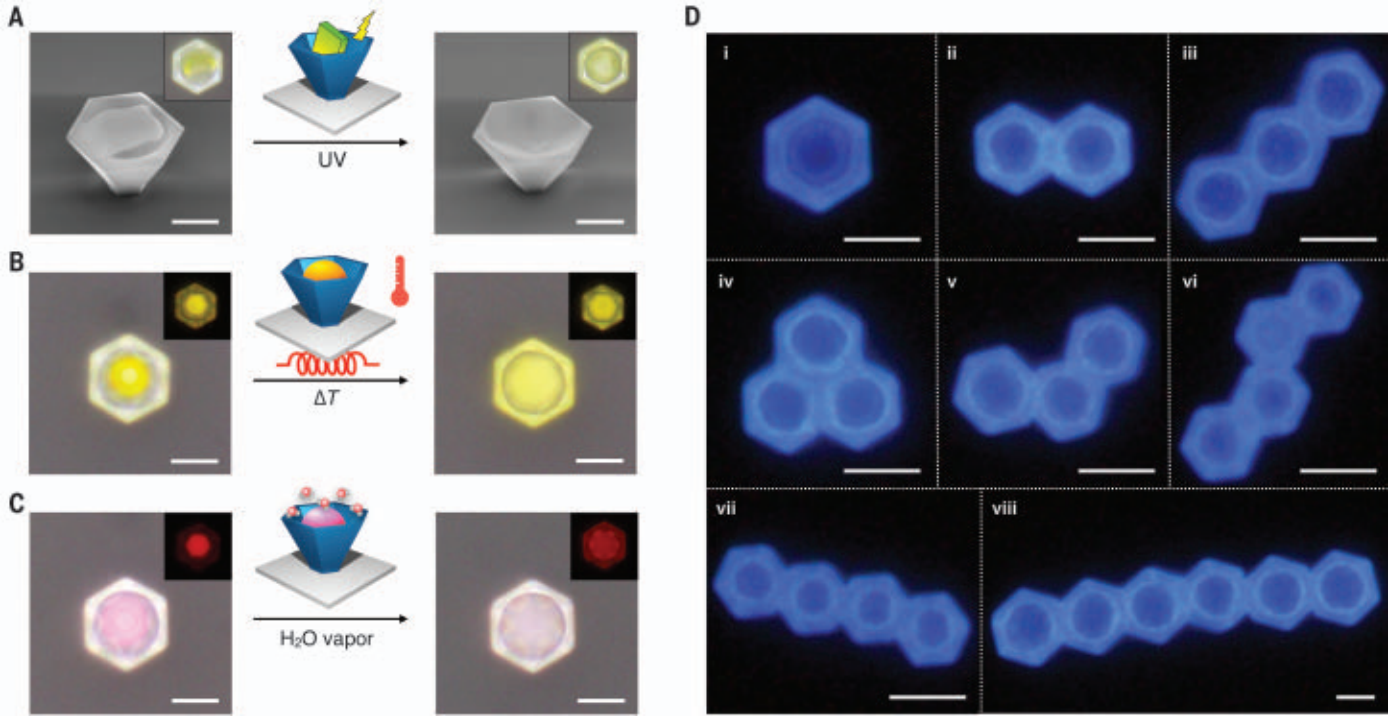


Fig. 4. Applications to microcontainers and PAH-like microarchitectures. (**A**) SEM and optical (inset) images of a microvessel in which a piece of an AZO microcrystal is placed before (left) and after (right) ultraviolet (UV) light irradiation at 340 to 390 nm. Scale bars: 4 μm. (**B**) Optical and FM (insets) images of a microvessel in which a PFBT microsphere is placed before (left) and after (right) thermal heating at 90°C for 1 min. Scale bars: 5 μm. For FM, λ_{ex} = 400 to 440 nm. (**C**) Optical and FM (insets) images of a microvessel in which hygroscopic PVA microspheres are placed before (left) and after (right) exposure to humid air. Scale bars: 5 μm. For FM, λ_{ex} = 460 to 495 nm. (**D**) FM images of fused microcrystals reminiscent of benzene (i), naphthalene (ii), anthracene (iii), phenalene (iv), phenanthrene (v), chrysene (vi), tetracene (vii), and hexacene (viii). Scale bars: 3 μm. λ_{ex} = 340 to 390 nm.

reaction upon exposure to ultraviolet light (*25*) (Fig. 4A and fig. S16). As shown in the optical and SEM images, AZO is held inside the vessel even after the transition to the liquid phase. The material scope is successfully expanded to two types of polymer particles that melt upon heating [poly(9,9-bis((*S*)-3,7-dimethyloctyl)-2,7-fluorene-*alt*-benzothiadiazole) (PFBT)] (*26*) and hydration [polyvinyl alcohol (PVA) doped with Acid Red 52] (Fig. 4, B and C, and figs. S17 and S18).

The microvessels grow synchronously and uniaxially, but their locations appear to be random in principle. Nonetheless, we occasionally find fused hexagonal vessels, which are reminiscent of the geometries of polycyclic aromatic hydrocarbons (PAHs). Dyads and triads are statistically feasible even if the locations are totally random [Fig. 4D, (i) to (vi)]. Much larger fused polyacene-like objects such as tetracene and hexacene are observed [Fig. 4D, (vii) to (viii), and fig. S19]. We assume that a faint linear scratch or chemical heterogeneity on the substrate facilitates nucleation of the crystals, leading to ordered microvessels (fig. S20) (*27*, *28*).

Lastly, we revisit the general issue of skeletal crystals—namely, how one can assemble skeletal single crystals with identical size and geometry. First, the crystalline nuclei should be fixed on the substrate throughout the crystallization. Given a certain area, the time-course profiles of temperature and concentration are kept homogeneous, leading to synchronous growth rate and regime. Uniaxiality is simultaneously achieved when a certain crystalline facet preferentially lies on the substrate. This is in sharp contrast to a suspension system, in which the crystals drift around in a solution with certain concentration and temperature gradients (fig. S21). Second, the molecules should crystallize faster than the evaporation- and cooling-induced resupersaturation. Otherwise, lengthy nucleation, which deteriorates synchronous growth, may occur during the crystal growth process (fig. S22).

To validate this strategy, we apply the same crystallization procedure to a single-decker aromatic molecule (**M1**) that features four (methoxyphenyl)ethynyl arms on a single benzene ring (*29*). Upon drop-casting, **M1** forms crystals in a synchronous and uniaxial manner to generate platelet crystals with a unimodal size distribution, with its crystalline (100) facet facing down to the substrate (figs. S23 and S24, table S5, and movie S3), demonstrating the generality of the first part of the aforementioned strategy. Second, the suppression of lengthy nucleation is also important, as revealed by the crystallization of **(*S*)-CP$_x$** (x = 3, 2, 1), which are **(*S*)-CP$_4$** analogs that have x (methoxyphenyl)ethynyl arm(s) and 4 – x phenylethynyl arm(s) (figs. S25 to S33) (*30*). Upon drop-casting, **(*S*)-CP$_3$** forms uniaxially oriented skeletal crystals with C_6-symmetric prominent edges (figs. S34 and S35 and table S6). Notably, the nucleation of **(*S*)-CP$_3$** continues much longer (20 s) than that of **(*S*)-CP$_4$** (2.0 s), resulting in polydisperse size distribution (movie S4).

We also find that the uniaxiality can be switched by chemically modifying the surface of the substrate. On a methylated quartz substrate, **(*S*)-CP$_4$** crystallizes into a merohedral twin of the vessel crystals, with its crystallographic c axis parallel to the substrate, while maintaining the size uniformity (figs. S36 to S38 and tables S7 and S8).

In conclusion, we find synchronous and uniaxial growth of skeletal crystals generated from a planar-chiral paracyclophane molecule. The crystal growth proceeds synchronously over a wide area on the substrate surface. The microparticles are single crystals and feature monodisperse size distribution and uniform vessel-like morphology with well-developed facets, whereas their interior remains concave.

The methodology is applicable to the crystallization of aromatic molecules with intricate features that can be produced within a few tens of seconds.

REFERENCES AND NOTES

1. V. Gornitz, in *Mineralogy* (Encyclopedia of Earth Science Series, Springer, 1981).
2. Y. Oaki, H. Imai, *Cryst. Growth Des.* **3**, 711–716 (2003).
3. Z. Yang *et al*., *Adv. Funct. Mater.* **30**, 1908108 (2020).
4. M. Cao *et al*., *Angew. Chem. Int. Ed.* **44**, 4197–4201 (2005).
5. J. Desarnaud, H. Derluyn, J. Carmeliet, D. Bonn, N. Shahidzadeh, *J. Phys. Chem. Lett.* **9**, 2961–2966 (2018).
6. T. Haxhimali, A. Karma, F. Gonzales, M. Rappaz, *Nat. Mater.* **5**, 660–664 (2006).
7. J. P. Zhu *et al*., *Chem. Commun.* **2005**, 5802–5804 (2005).
8. W. L. Noorduin, A. Grinthal, L. Mahadevan, J. Aizenberg, *Science* **340**, 832–837 (2013).
9. H. C. Hendrikse *et al*., *Adv. Mater.* **32**, e2003999 (2020).
10. C. Yu, Q. Xie, Y. Bao, G. Shan, P. Pan, *Crystals* **7**, 147 (2017).
11. J. Ge, J. Lei, R. N. Zare, *Nat. Nanotechnol.* **7**, 428–432 (2012).
12. W. Jiang, M. S. Pacella, H. Vali, J. J. Gray, M. D. McKee, *Sci. Adv.* **4**, eaas9819 (2018).
13. M. C. di, *Nat. Commun.* **11**, 308 (2020).
14. K. Iwai *et al*., *Angew. Chem. Int. Ed.* **59**, 12674–12679 (2020).
15. Q. Wen *et al*., *J. Am. Chem. Soc.* **142**, 14210–14221 (2020).
16. M. P. Zhuo, G. P. He, X. D. Wang, L. S. Liao, *Nat. Commun.* **12**, 2252 (2021).
17. C. R. R. Adolf, S. Ferlay, N. Kyritsakas, M. W. Hosseini, *J. Am. Chem. Soc.* **137**, 15390–15393 (2015).
18. T. Zhang *et al*., *J. Am. Chem. Soc.* **128**, 10960–10968 (2006).
19. J. D. Smith, M. M. Scanlan, A. N. Chen, H. M. Ashberry, S. E. Skrabalak, *ACS Nano* **14**, 15953–15961 (2020).
20. S. H. Yoo, H. S. Lee, *Acc. Chem. Res.* **50**, 832–841 (2017).
21. Z.-Q. Lin *et al*., *ACS Nano* **6**, 5309–5319 (2012).
22. L. Kang, H. Fu, X. Cao, Q. Shi, J. Yao, *J. Am. Chem. Soc.* **133**, 1895–1901 (2011).
23. M. Gon, Y. Morisaki, R. Sawada, Y. Chujo, *Chem. Eur. J.* **22**, 2291–2298 (2016).
24. Y. Morisaki, Y. Chujo, *Bull. Chem. Soc. Jpn.* **92**, 265–274 (2019).
25. Y. Norikane *et al*., *Org. Lett.* **16**, 5012–5015 (2014).
26. O. Oki *et al*., *J. Am. Chem. Soc.* **143**, 8772–8779 (2021).
27. J. Aizenberg, A. Black, G. Whitesides, *Nature* **398**, 495–498 (1999).
28. D. Okada *et al*., *Mater. Horiz.* **7**, 1801–1808 (2020).
29. P. Ehlers, A. Hakobyan, A. Neubauer, S. Lochbrunner, P. Langer, *Adv. Synth. Catal.* **355**, 1849–1858 (2013).
30. Y. Sasai, H. Tsuchida, T. Kakuta, T. Ogoshi, Y. Morisaki, *Mater. Chem. Front.* **2**, 791–795 (2018).

ACKNOWLEDGMENTS

Funding: This work was supported by CREST (JPMJCR20T4) and ACT-X (JPMJAX201J) from the Japan Science and Technology Agency (JST); Grants-in-Aid for Scientific Research on Innovative Areas "π-System Figuration" (JP17H05142 and JP17H05165), Scientific Research (A) (JP16H02081), Scientific Research (B) (JP19H02792), Challenging Research (Exploratory) (JP21K18860), and Young Scientist (JP19J20398) from the Japan Society for the Promotion of Science (JSPS); and the Ogasawara Foundation. **Author contributions:** O.O. and Y.Y. conceived the idea; O.O., H.Y., and Y.Y. designed the experiments; Y.M., R.I., K.O., N.M., and Y.N. synthesized molecules; O.O. conducted the experiments; H.S. performed single-crystal x-ray structure analysis; and O.O., H.Y., and Y.Y. prepared the manuscript. **Competing interests:** Y.N. is an inventor on patent applications JP 5765751, US 9126899, KR 10-1895105, GB 2786984, FR 2786984, DE 2786984, and CN ZL 2012 8 0068860.8 held by AIST that cover an azobenzene derivative (AZO). **Data and materials availability:** Crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre (CCDC 2184874 to 2184881). All other data are available in the main text or the supplementary materials. **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm9596
Materials and Methods
Figs. S1 to S38
Tables S1 to S8
References (*31*, *32*)
Movies S1 to S4
Data S1 to S8

Submitted 24 October 2021; accepted 5 July 2022
10.1126/science.abm9596

THE UNIVERSITY OF
CHICAGO

SENIOR INSTRUCTIONAL PROFESSOR AND DIRECTOR OF THE PHOENIX STEM PROGRAM

The Biological and Physical Sciences Collegiate Divisions at the University of Chicago are accepting applications for the position of **Senior Instructional Professor** (rank to be determined) and **Director** of the Phoenix STEM Program. This a full time, renewable appointment, with an initial term of two years. The start date for this position will be between **October 1, 2022** and **January 1, 2023**.

The Senior Instructional Professor/Director will be responsible for the STEM program design and administration. Responsibilities will include hiring and directing a team of instructors who will lead collaborative learning classes associated with STEM courses; create and oversee a team of academic advisors for Phoenix STEM scholars; organize and run a pre-matriculation program for incoming Phoenix scholars; interface with divisional masters, departments and summer school to integrate the Phoenix STEM program within the STEM majors, including working with the academic units to identify and mitigate roadblocks to success for Phoenix scholars; coordinate career, fellowship and research opportunities for Phoenix STEM scholars with relevant College organizations. Work with the UChicago Center for Teaching and Learning to implement faculty/instructor training on issues around the URM and women students experience in STEM Majors. Collaborate with data scientists to collect data on URM STEM student progress and experiences in the College, and design and implement approaches for measuring the success of the program; design supplemental activities to enhance interactions between Phoenix scholars and faculty; and write annual reports on the activities, effectiveness, and impact of the program. The Senior Instructional Professor/Director will also offer courses in a STEM field.

Qualifications:
The position requires a Ph.D. in STEM discipline with 3+ years in academic research or administration; 3+ years of teaching experience at the College level in STEM field and training in STEM pedagogy; and training and experience in diversity and inclusion in academia.

The successful candidate will also have excellent organizational skills; excellent management skills and the ability to lead and motivate a team; excellent writing skills; and experience in programmatic assessment (surveys, focus groups, quantitative data).

To apply for this position candidates must submit their application through the University of Chicago's Interfolio jobs board at **http://apply.interfolio.com/108463** and upload a cover letter, current curriculum vitae, contact information for three references.

Applications will be reviewed until the position is filled.

Equal Employment Opportunity Statement
*We seek a diverse pool of applicants who wish to join an academic community that places the highest value on rigorous inquiry and encourages diverse perspectives, experiences, groups of individuals, and ideas to inform and stimulate intellectual challenge, engagement, and exchange. The University's Statements on Diversity are at **https://provost.uchicago.edu/statements-diversity**. The University of Chicago is an Affirmative Action/Equal Opportunity/Disabled/Veterans Employer and does not discriminate on the basis of race, color, religion, sex, sexual orientation, gender identity, national or ethnic origin, age, status as an individual with a disability, protected veteran status, genetic information, or other protected classes under the law. For additional information please see the University's Notice of Nondiscrimination. Job seekers in need of a reasonable accommodation to complete the application process should call 773-834-3988 or email **equalopportunity@uchicago.edu** with their request.*

NIMS

ICYS Research Fellow at ICYS, NIMS, Japan

The International Center for Young Scientists (ICYS) of the National Institute for Materials Science (NIMS) invites applications for ICYS Research Fellow positions. ICYS will offer you the freedom to conduct independent and self-directed research in various areas of materials science with full access to NIMS advanced research facilities.

The common language at ICYS is English. Clerical and technical support in English will be given by the ICYS staff. An annual salary starts from approximately 5.87 million yen, which may be increased depending on the performance of the Research Fellow*. In addition, a research grant of 2 million yen per year will be provided to each Research Fellow. The initial contract term is two years, which may be extended depending on one's experience and performance. Also, advantage is given when applying to NIMS permanent researcher position (about 50% of the applicants are accepted).

All applicants must have obtained a PhD degree within the last ten years. Applicants should submit an **application form** including a research proposal during the ICYS term, **CV, a list of DOI of journal publications, PDF files of three significant publications**, and **PhD Certificate** to the ICYS Recruitment Desk by **September 29, 2022 JST**. The format for the application documents can be downloaded from our website. The selection will be made on the basis of originality and quality of the research proposal as well as the research achievements. Please visit our website for more details.

*Approximately 23% of annual salary will be deducted as social insurance premium, residence tax and income tax.

ICYS Recruitment Desk
National Institute for Materials Science
www.nims.go.jp/icys/recruitment/

SIAT

Shenzhen Institute of Advanced Technology
Chinese Academy of Sciences

Established in partnership between the Chinese Academy of Sciences and the Shenzhen Municipal Government, the Shenzhen Institute of Advanced Technology (SIAT) is a newly-created university with an objective to become the world's preeminent institute for emerging science and engineering programs. SIAT is equipped with state-of-art teaching and research facilities and is dedicated to cultivating international, visionary, and interdisciplinary talents while delivering research support to pursue innovation-driven development.

SIAT is located in Shenzhen, also known as the "Silicon Valley of China," a modern, clean, and green city, well-known for its stunning architecture, vibrant economy, and its status as a leading global technology hub. SIAT is seeking applications for faculty positions of all ranks in the following academic programs: Computer Science and Engineering, Bioinformatics, Robotics, Life Sciences, Material Science and Engineering, Biomedical Engineering, Pharmaceutical Sciences, Synthetic Biology, Neurosciences, etc. SIAT seeks individuals with a strong record of scholarship who possess the ability to develop and lead high-quality teaching and research programs. SIAT offers a comprehensive benefits package and is committed to faculty success throughout the academic career trajectory, providing support for ambitious and world-class research projects and innovative, interactive teaching methods.

Further information:
https://www.siat-sz.edu.cn/

Science Careers online @sciencecareers.org

ScienceCareers online @sciencecareers.org

IT'S NOT JUST A JOB. IT'S A CALLING.

Find your next job at ScienceCareers.org

The relevance of science is at an all-time high these days. For anyone who's looking to get ahead in —or just plain get into— science, there's no better, more trusted resource or authority on the subject than *Science* Careers. Here you'll find opportunities and savvy advice across all disciplines and levels. There's no shortage of global problems today that science can't solve. Be part of the solution.

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

Science Careers helps you advance your career. Learn how!

- Register for a free online account on **ScienceCareers.org**.
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.

Visit **ScienceCareers.org** today — all resources are free

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

EGL CHARITABLE FOUNDATION

The EGL Charitable Foundation invites you to apply to the

Gruss Lipper Postdoctoral Fellowship Program

Eligibility:

- Israeli citizenship
- Candidates must have completed PhD and/or MD/PhD degrees in Biomedical Science at an accredited Israeli University/Medical School, or be in their final year of studies
- Candidates must have been awarded a postdoctoral position by a U.S. host research institution

Details available at: www.eglcf.org
Application deadline: October 9, 2022

Who's the top employer for 2021?

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article and employer profiles at sciencecareers.org/topemployers

Science 2021 TOP EMPLOYER

By **Huanping Huang**

Try, try again

Earlier this summer, my alma mater finally held my Ph.D. commencement, for degrees earned during the 2019–20 academic year. I was thrilled to return to campus to celebrate this milestone, which had been postponed because of the pandemic. While there, I noticed a poster board titled "Unlock your potential," and one of the sticky notes on the board caught my attention: "Keep trying until you get it right." As a student, I would have easily overlooked this note. But now, it resonated with me so strongly that I stopped to reflect on how I got to this point.

Growing up in rural China to a family of farmers, I dreamed of becoming a scientist. Books and newspapers became my good friends on this path, and I made it to college in Beijing. I struggled but managed to graduate, though without a sense of academic fulfillment in my field of engineering. Still, my goal to become a scientist remained alive, and I decided to pursue a master's degree in a new field, meteorology. I was inspired to explore climate change and how agriculture could adapt, which is vital to farmers like my family. It was invigorating that my work made an impact. When I finished, I wanted to try something even more ambitious: a doctoral degree.

"I learned to enjoy the process of trying and the unexpected opportunities it brought."

My aspiration met hard reality when I failed my target school's Ph.D. entrance exam. Nearly all my friends and professors advised me to put my Ph.D. ambitions aside and accept a coveted position I had been offered at a meteorological bureau, observing and reporting weather conditions. But I doubted whether this job was right for me. I enjoyed research and I worried the job, with its emphasis on routine work and little room for creativity, would dampen my passion to explore. So, I decided to pursue a riskier path: applying abroad for doctoral programs.

While studying to pass the English language tests required for those programs, I worked as a research assistant to earn a living and support my applications. My living conditions—all I could afford was a windowless basement room that fit just a twin bed and one chair—occasionally tempted me to regret my decisions. As I took the English tests multiple times but never scored high enough, my self-doubt grew. Had I chosen the wrong path? Despite the setbacks, the fear of a bigger failure—being jobless the following year—motivated me to keep trying. Eventually, I passed the tests and was thrilled to be accepted into a doctoral program in the United States.

Once there, I encountered even more challenges—language and social barriers, cultural differences, demanding academic requirements, and more. Yet I held the same attitude that lifted me up from the basement: Keep trying. In particular, communicating proved to be very difficult. In Beijing, I was always eager to talk to people. But in the United States, I often wouldn't understand why something was funny even as people around me laughed. The frustration made me hesitant to attend social events, but my heart called on me to not stop trying. Fortunately, my classmates continued to invite me and would break the ice by introducing my roots as a farmer, which unexpectedly became an effective conversation starter. Over time, I spoke fluent English and made new friends. I moved on to a postdoc position and then a faculty job search. Similar to the earlier steps in my journey, the process was by no means smooth, as I only got one interview in the first 2 years, but the lessons I learned paved the way to success in the third year.

On the eve of the Ph.D. commencement, I realized that by trying again and again, I have unlocked my potential. Trying does not always lead to ideal results, which can be frustrating. However, by following my heart, I developed a clearer view of how I needed to adapt. By focusing more on what I achieved and less on what I missed, I learned to enjoy the process of trying and the unexpected opportunities it brought, regardless of how different they were from my initial expectations. Without the attempts and adjustments, I would not have found my right place in the world and grown into the person I wanted to be. ■

Huanping Huang is a postdoctoral fellow at Lawrence Berkeley National Laboratory and a newly hired assistant professor at Louisiana State University, Baton Rouge. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

What's Your Next Career Move?

From networking to mentoring to evaluating your skills, find answers to your career questions on *Science* Careers

To view the complete collection, visit
ScienceCareers.org/booklets

Science Careers
FROM THE JOURNAL SCIENCE AAAS

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**