

Models that play what-if with
an artificial society p. 144

A path toward quantum
supremacy p. 191

Norovirus targets
tuft cells p. 204

Science

\$15
13 APRIL 2018
sciencemag.org

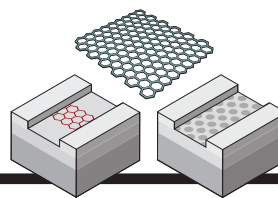
AAAS



WAITING FOR SPRING

How day length regulates
plant dormancy p. 212

CONTENTS



154

Nanoporous graphene

13 APRIL 2018 • VOLUME 360 • ISSUE 6385



NEWS

IN BRIEF

134 News at a glance

IN DEPTH

137 LATIN AMERICA'S LOST HISTORIES REVEALED

Clues to forgotten migrations emerge from today's genomes *By L. Wade*

► PODCAST

138 ALPHA CENTAURI'S SIREN CALL HAS FRUSTRATED PLANET HUNTERS

The nearest sunlike stars have failed to yield exoplanets so far, but searches for Earth-like ones are ramping up *By D. Clery*

139 CHEMISTS SEEK ANTIADDICTION DRUGS TO BATTLE HIJACKED BRAIN

Candidate medications could help addicts overcome cravings that lead to relapse and death *By R. F. Service*

140 ANCIENT SITES SAVAGED IN YEMEN, IRAQ

Firsthand accounts reveal worse damage than expected in war-torn regions *By A. Lawler*

142 STUDY QUESTIONS ANIMAL EFFICACY DATA BEHIND TRIALS

Information provided to ethical review panels may often be insufficient to judge a drug's therapeutic potential *By E. Yasinski*

143 HUMAN MUTATION RATE A LEGACY FROM OUR PAST

By assessing mutation rates among species, researchers are understanding why they vary *By E. Pennisi*

FEATURES

144 FREE AGENTS

Monumentally complex models are gaming out disaster scenarios with millions of simulated people

By M. M. Waldrop

► VIDEO

INSIGHTS

PERSPECTIVES

148 HOW CLEANER AIR CHANGES THE CLIMATE

Air quality improvements affect regional climate in complex ways *By B. H. Samset*

150 IMPROVED MEMORY DEVICES FOR SYNTHETIC CELLS

CRISPR enables efficient recording of signaling events in cells onto DNA

By J. M. L. Ho and M. R. Bennett

► RESEARCH ARTICLE P. 169

152 REDEMPTION FOR SELF-REACTIVE ANTIBODIES

Antibody self-reactivity is repaired through antibody gene mutation in B cells *By E. E. Kara and M. C. Nussenzweig*

► REPORT P. 223

153 CROWDSOURCED GENEALOGIES AND GENOMES

Genealogical study provides insight into history and life span and heralds crowdsourced genetic research

By A. A. Lussier and A. Keinan

► RESEARCH ARTICLE P. 171

154 A RECIPE FOR NANOPOROUS GRAPHENE

Nanoporous graphene created from molecular precursors shows promise for electronic applications *By A. Sinitskii*

► REPORT P. 199

156 STEPHEN HAWKING (1942–2018)

The world's best-known scientist richly deserved his fame *By J. Preskill*

157 JOHN SULSTON (1942–2018)

A visionary biologist with a deep social conscience *By J. Kimble*

POLICY FORUM

158 BYSTANDER RISK, SOCIAL VALUE, AND ETHICS OF HUMAN RESEARCH

Contentious risks demand a new approach *By S. K. Shah et al.*

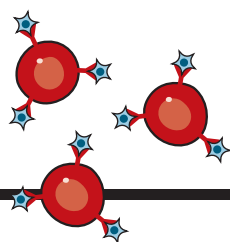
BOOKS ET AL.

160 ADAPTING TO LIFE IN THE BIG CITY

To thrive in rapidly changing urban areas, plants and animals are evolving at astonishing rates *By A. Mooers*

161 THE FUTURE OF ARTISANAL FISHING

Declining fish populations and policies that favor large operations threaten small fisheries *By D. Pauly*



152

Evolving antibody specificity



219

RNA processing machinery

LETTERS

162 EDITOR'S NOTE

By J. Berg

162 Social media for social change in science

By S. Z. Yammine et al.

163 Journal editors should not divide scientists

By M. Zaringhalam et al.

163 Online Buzz: SciComm speaks

164 Efforts large and small speed science reform

By A. J. Jefferson and M. A. Kenney

RESEARCH

IN BRIEF

166 From *Science* and other journals

RESEARCH ARTICLES

169 SYNTHETIC BIOLOGY

Rewritable multi-event analog recording in bacterial and mammalian cells

W. Tang and D. R. Liu

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: [dx.doi.org/10.1126/science.aap8992](https://doi.org/10.1126/science.aap8992)

► PERSPECTIVE P. 150

170 CELL BIOLOGY

MitoCPR—A surveillance pathway that protects mitochondria in response to protein import stress

H. Weidberg and A. Amon

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: [dx.doi.org/10.1126/science.aan4146](https://doi.org/10.1126/science.aan4146)

171 BIG DATA

Quantitative analysis of population-scale family trees with millions of relatives

J. Kaplanis et al.

► PERSPECTIVE P. 153

176 SINGLE-CELL GENOMICS

Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding

A. B. Rosenberg et al.

182 TOPOLOGICAL MATTER

Observation of topological superconductivity on the surface of an iron-based superconductor

P. Zhang et al.

REPORTS

186 ORGANIC CHEMISTRY

Predicting reaction performance in C–N cross-coupling using machine learning

D. T. Ahneman et al.

191 METROLOGY

Measurement of the fine-structure constant as a test of the Standard Model

R. H. Parker et al.

195 QUANTUM INFORMATION

A blueprint for demonstrating quantum supremacy with superconducting qubits

C. Neill et al.

199 NANOMATERIALS

Bottom-up synthesis of multifunctional nanoporous graphene

C. Moreno et al.

► PERSPECTIVE P. 154

204 NOROVIRUS

Tropism for tuft cells determines immune promotion of norovirus pathogenesis

C. B. Wilen et al.

209 CARBON CYCLE

Microbial oxidation of lithospheric organic carbon in rapidly eroding tropical mountain soils

J. D. Hemingway et al.

212 PLANT SCIENCE

Photoperiodic control of seasonal growth is mediated by ABA acting on cell-cell communication

S. Tylewicz et al.

215 STRUCTURAL BIOLOGY

Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum

K. Braunger et al.

219 STRUCTURAL BIOLOGY

Structure of the nuclear exosome captured on a maturing preribosome

J. M. Schuller et al.

223 IMMUNOLOGY

Germinal center antibody mutation trajectories are determined by rapid self/foreign discrimination

D. L. Burnett et al.

► PERSPECTIVE P. 152

DEPARTMENTS

133 EDITORIAL

Obfuscating with transparency

By Jeremy Berg

234 WORKING LIFE

My path to contentment

By Edmond Sanganyado

ON THE COVER



Trees such as this aspen (*Populus tremuloides*) protect their meristem and leaf primordia from low temperatures during winter by establishing dormancy in apical buds. Reduction

in day length, heralding the advent of winter, induces dormancy. The molecular mechanism underlying photoperiodic control of tree dormancy has been revealed to involve plant hormone-mediated blockage of plasmodesmata, channels that connect neighboring cells. See page 212.

Photo: Jeff Foott/Getty Images

Science Staff	130
New Products	227
Science Careers	228

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 © 2018 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): \$1808; Foreign postage extra: Mexico, Caribbean (surface mail) \$55; other countries (air assist delivery): \$89. 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; 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 is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that \$35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Editor-in-Chief Jeremy Berg

Executive Editor Monica M. Bradford **News Editor** Tim Appenzeller

Deputy Editors Lisa D. Chong, Andrew M. Sugden(UK), Valda J. Vinson, Jake S. Yeston

Research and Insights

DEPUTY EDITOR, EMERITUS Barbara R. Jasny **SR. EDITORS** Gemma Alderton(UK), Caroline Ash(UK), Julia Fahrenkamp-Uppenbrink(UK), Pamela J. Hines, Stella M. Hurtleby(UK), Paula A. Kiberstis, Marc S. Lavine(Canada), Steve Mao, Ian S. Osborne(UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Jelena Stajic, Peter Stern(UK), Phillip D. Szuroni, Sacha Vignieri, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Brent Grocholski, Priscilla N. Kelly, Seth Thomas Scanlon(UK), Keith T. Smith(UK) **ASSOCIATE BOOK REVIEW EDITOR** Valerie B. Thompson **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Amber Esplin, Chris Filiatreau, Cynthia Howe, Catherine Wolner **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Scott Miller, Jerry Richardson, Anita Wynn **PUBLICATIONS ASSISTANTS** Ope Martins, Nida Masiulis, Dona Mathieu, Hilary Stewart(UK), Alana Warnke, Alice Whaley(UK), Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ADMINISTRATIVE SUPPORT** Janet Clements(UK), Lizzanne Newton(UK)

News

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Richard Stone **DEPUTY NEWS EDITORS** Elizabeth Culotta, Martin Enserink(Europe), David Grimm, Eric Hand, David Malakoff, Leslie Roberts **SR. CORRESPONDENTS** Daniel Cley(UK), Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Catherine Matatic **NEWS WRITERS** Adrian Cho, Jon Cohen, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad(Cambridge, UK), Paul Voosen, Meredith Wadman **INTERNS** Roni Dengler, Katie Langin, Matt Warren **CONTRIBUTING CORRESPONDENTS** John Bohannon, Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt(Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile(Shanghai), Charles Piller, Tania Rabesandratana(London), Emily Underwood, Gretchen Vogel(Berlin), Lizzie Wade(Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein(Editor) **COPY EDITORS** Dorie Chevlen, Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Executive Publisher Rush D. Holt

Publisher Bill Moran **Chief Digital Media Officer** Josh Freeman

DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT Sarah Whalen **DIRECTOR, PRODUCT AND CUSTOM PUBLISHING** Will Schweitzer **MANAGER, PRODUCT DEVELOPMENT** Hannah Heckner **BUSINESS SYSTEMS AND FINANCIAL ANALYSIS DIRECTOR** Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYST** Eric Knott **SENIOR SYSTEMS ANALYST** Nicole Mehmedovich **SENIOR BUSINESS ANALYST** Cory Lipman **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **BUSINESS ANALYSTS** Meron Kebede, Sandy Kim, Jourdan Stewart **FINANCIAL ANALYST** Julian Iriarte **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SALES COORDINATOR** Shirley Young **DIRECTOR, COPYRIGHT, LICENSING, SPECIAL PROJECTS** Emilie David **DIGITAL PRODUCT ASSOCIATE** Michael Hardesty **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **RIGHTS, CONTRACTS, AND LICENSING ASSOCIATE** Lili Catlett **RIGHTS & PERMISSIONS ASSISTANT** Alexander Lee

MARKETING MANAGER, PUBLISHING Shawana Arnold **MARKETING ASSOCIATE** Steven Goodman **SENIOR ART ASSOCIATES** Paula Fry **ART ASSOCIATE** Kim Huynh

INTERIM DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGERS** Marco Castellani, Chris Murawski **SENIOR OPERATIONS ANALYST** Lana Guz **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge

WEB TECHNOLOGIES TECHNICAL DIRECTOR David Levy **TECHNICAL MANAGER** Chris Coleman **PORTFOLIO MANAGER** Trista Smith **PROJECT MANAGER** Tara Kelly, Dean Robbins **DEVELOPERS** Elissa Heller, Ryan Jensen, Brandon Morrison

DIGITAL MEDIA DIRECTOR OF ANALYTICS Enrique Gonzales **SR. MULTIMEDIA PRODUCER** Sarah Crespi **MANAGING DIGITAL PRODUCER** Kara Estelle-Powers **PRODUCER** Liana Birke **VIDEO PRODUCERS** Chris Burns, Nguyễn Khôi Nguyễn **DIGITAL SOCIAL MEDIA PRODUCER** Brice Russ

DIGITAL/PRINT STRATEGY MANAGER Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **DIGITAL PRODUCTION MANAGER** Lisa Stanford **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Antoinette Hodal, Lori Murphy, Anthony Rosen **CONTENT SPECIALISTS** Jacob Hedrick, Kimberley Oster

DESIGN DIRECTOR Beth Rakouskas **DESIGN MANAGING EDITOR** Marcy Atarod **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS MANAGING EDITOR** Alberto Cuadra **GRAPHICS EDITOR** Nirja Desai **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel, Katharine Sutfill **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **INTERACTIVE GRAPHICS EDITOR** Jia You **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **PHOTO EDITOR** Emily Petersen **IMAGE RIGHTS AND FINANCIAL MANAGER** Jessica Adams **INTERN** Mike Shanahan

SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders: 202-326-6430 **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst: 202-326-6463 **ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT** Justin Sawyers: 202-326-7061 science_advertising@aaas.org **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **SR. PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle **SR. TRAFFIC ASSOCIATE** Christine Hall **DIRECTOR OF BUSINESS DEVELOPMENT AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu: +86-131 6136 3212, xchu@aaas.org **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu + 81532-81-5142 asandhu@aaas.org **EAST COAST/E. CANADA** Laurie Faraday: 508-747-9395, FAX 617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FAX 415-520-6940 **MIDWEST** Jeffrey Dembski: 847-498-4520 x3005, Steven Loerch: 847-498-4520 x3006 **UK EUROPE/ASIA** Roger Goncalves: TEL/FAX +41 243 1358 **JAPAN** Kaoru Sasaki (Tokyo): + 81 (3) 6459 4174 ksasaki@aaas.org

GLOBAL SALES DIRECTOR ADVERTISING AND CUSTOM PUBLISHING Tracy Holmes: +44 (0) 1223 326525 **CLASSIFIED** advertise@sciencecareers.org **SALES MANAGER, US, CANADA AND LATIN AMERICA** SCIENCE CAREERS Claudia Paulsen-Young: 202-326-6577 **EUROPE/ROW SALES** Sarah Lelarge **SALES ADMIN ASSISTANT** Kelly Grace +44 (0)1223 326528 **JAPAN** Miyuki Tani(Osaka): +81 (6) 6202 6272 mtani@aaas.org **CHINA/TAIWAN** Xiaoying Chu: +86-131 6136 3212, xchu@aaas.org **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING ASSOCIATE** Aimee Aponte

AAAS BOARD OF DIRECTORS, CHAIR Susan Hockfield **PRESIDENT** Margaret A. Hamburg **PRESIDENT-ELECT** Steven Chu **TREASURER** Carolyn N. Ainslie **CHIEF EXECUTIVE OFFICER** Rush D. Holt **BOARD** Cynthia M. Beall, May R. Berenbaum, Rosina M. Bierbaum, Kaye Husbands Fealing, Stephen P.A. Fodor, S. James Gates, Jr., Michael S. Gazzaniga, Laura H. Greene, Robert B. Millard, Mercedes Pascual, William D. Provine

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Central Support** 866-434-2227 www.aaas.org/membercentral

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.

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

BOARD OF REVIEWING EDITORS (Statistics board members indicated with \$)

Adriano Aguzzi, U. Hospital Zürich
Takuzo Aida, U. of Tokyo
Leslie Aiello, Wenner-Gren Foundation
Judith Allen, U. of Manchester
Sebastian Amigorena, Institut Curie
Meinrat O. Andrae, Max Planck Inst. Mainz
Paola Ariotti, Harvard U.
Johan Auwerx, EPFL
David Awschalom, U. of Chicago
Clare Baker, U. of Cambridge
Nenad Ban, ETH Zürich
Franz Bauer, Pontificia Universidad Católica de Chile
Ray H. Baughman, U. of Texas at Dallas
Carlo Beenakker, Leiden U.
Kamran Behnia, ESPCI
Yasmine Belkaid, NIAID, NIH
Philip Benfey, Duke U.
Gabriele Bergers, VIB
Bradley Bernstein, Massachusetts General Hospital
Peer Bork, EMBL
Chris Bowler, Ecole Normale Supérieure
Ian Boyd, U. of St. Andrews
Emily Brodsky, U. of California, Santa Cruz
Ron Brookmeyer, U. of California, Los Angeles (\$) **Christian Büchel, UKE Hamburg**
Dennis Burton, The Scripps Res. Inst.
Carter Tribley Butts, U. of California, Irvine
Gyorgy Buzsaki, New York U. School of Medicine
Blanche Capel, Duke U.
Mats Carlsson, U. of Oslo
Ib Chorkendorff, Denmark TU
James J. Collins, MIT
Robert Cook-Deegan, Arizona State U.
Lisa Coussens, Oregon Health & Science U.
Alan Cowman, Walter & Eliza Hall Inst.
Roberta Croce, VU Amsterdam
Janet Currie, Princeton U.
Jeff L. Dangl, U. of North Carolina
Tom Daniel, U. of Washington
Chiara Daraio, Caltech
Nicolas Daughas, U. of Chicago
Frans de Waal, Emory U.
Stanislas Dehaene, Collège de France
Robert Desimone, MIT
Claude Desplan, New York U.
Sandra Díaz, Universidad Nacional de Córdoba
Dennis Discher, U. of Penn.
Gerald W. Dorn II, Washington U. in St. Louis
Jennifer A. Doudna, U. of California, Berkeley
Bruce Dunn, U. of California, Los Angeles
William Dunphy, Caltech
Christopher Dye, WHO
Todd Ehlers, U. of Tübingen
Jennifer Elisseeff, Johns Hopkins U.
Tim Elston, U. of North Carolina at Chapel Hill
Barry Everitt, U. of Cambridge
Vanessa Ezenwa, U. of Georgia
Ernst Fehr, U. of Zürich
Michael Feuer, The George Washington U.
Toren Finkel, NHLBI, NIH
Kate Fitzgerald, U. of Massachusetts
Peter Fratzl, Max Planck Inst. Potsdam
Elaine Fuchs, Rockefeller U.
Eileen Furlong, EMBL
Jay Gallagher, U. of Wisconsin
Daniel Geschwind, U. of California, Los Angeles
Karl-Heinz Glassmeier, TU Braunschweig
Ramon Gonzalez, Rice U.
Elizabeth Grove, U. of Chicago
Nicolas Gruber, ETH Zürich
Kip Guy, U. of Kentucky College of Pharmacy
Taekjip Ha, Johns Hopkins U.
Christian Haass, Ludwig Maximilians U.
Sharon Hammes-Schiffer, U. of Illinois at Urbana-Champaign
Wolf-Dietrich Hardt, ETH Zürich
Michael Hasselmo, Boston U.
Martin Heimann, Max Planck Inst. Jena
Ykä Helariutta, U. of Cambridge
Janet G. Hering, Eawag
Kai-Uwe Hinrichs, U. of Bremen
David Hodell, U. of Cambridge
Lora Hooper, UT Southwestern Medical Ctr. at Dallas
Fred Hughson, Princeton U.
Randall Hulet, Rice U.
Auke Ijspeert, EPFL
Akiko Iwasaki, Yale U.
Stephen Jackson, USGS and U. of Arizona
Seema Jayachandran, Northwestern U.
Kai Johnson, EPFL
Peter Jonas, Inst. of Science & Technology Austria
Matt Kaeblerlein, U. of Washington
William Kaelin Jr., Dana-Farber Cancer Inst.
Daniel Kammen, U. of California, Berkeley
Abby Kavner, U. of California, Los Angeles
Masashi Kawasaki, U. of Tokyo
V. Narry Kim, Seoul Nat. U.
Robert Kingston, Harvard Medical School
Etienne Koechlin, Ecole Normale Supérieure
Alexander Kolodkin, Johns Hopkins U.
Thomas Langer, U. of Cologne
Mitchell A. Lazar, U. of Penn.
David Lazer, Harvard U.
Thomas Lecuit, IBDM
Stanley Lemon, U. of North Carolina at Chapel Hill
Ottoline Leyser, U. of Cambridge
Wendell Lim, U. of California, San Francisco
Marcia C. Linn, U. of California, Berkeley
Jianguo Liu, Michigan State U.
Luis Liz-Marzán, CIC biomaGUNE
Jonathan Losos, Harvard U.
Ke Lu, Chinese Acad. of Sciences
Christian Lüscher, U. of Geneva
Laura Machesky, Cancer Research UK Beatson Inst.
Anne Magurran, U. of St. Andrews
Oscar Marin, King's College London
Charles Marshall, U. of California, Berkeley
Christopher Marx, U. of Idaho
C. Robertson McClung, Dartmouth College
Drahgo Medellin, U. of Mexico
Grham Medley, London School of Hygiene & Tropical Med.
Jane Memmott, U. of Bristol
Tom Misteli, NCI, NIH
Yasushi Miyashita, U. of Tokyo
Mary Ann Moran, U. of Georgia
Richard Morris, U. of Edinburgh
Alison Motsinger-Reif, NC State U. (\$) **Daniel Neumark, U. of California, Berkeley**
Kitty Nijmeijer, TU Eindhoven
Helga Nowotny, Austrian Council
Rachel O'Reilly, U. of Warwick
Harry Orr, U. of Minnesota
Pilar Ossorio, U. of Wisconsin
Andrew Oswald, U. of Warwick
Isabella Pagano, Istituto Nazionale di Astrofisica
Margaret Palmer, U. of Maryland
Steve Palumbi, Stanford U.
Jane Parker, Max Planck Inst. Cologne
Giovanni Parmigiani, Dana-Farber Cancer Inst. (\$) **John H. J. Petrini, Memorial Sloan Kettering**
Samuel Pfaff, Salk Inst. for Biological Studies
Kathrin Plath, U. of California, Los Angeles
Martin Plenio, Ulm U.
Albert Polman, FOM Institute for AMOLF
Elvira Poloczanska, Alfred-Wegener-Inst.
Philippe Poulin, CNRS
Jonathan Pritchard, Stanford U.
David Randall, Colorado State U.
Sarah Reisman, Caltech
Félix A. Rey, Institut Pasteur
Trevor Robbins, U. of Cambridge
Amy Rosenzweig, Northwestern U.
Mike Ryan, U. of Texas at Austin
Mitinori Saitou, Kyoto U.
Shimon Sakaguchi, Osaka U.
Miquel Salmeron, Lawrence Berkeley Nat. Lab
Nitin Samarth, Penn. State U.
Jürgen Sandkühler, Medical U. of Vienna
Alexander Schier, Harvard U.
Wolfram Schlenker, Columbia U.
Susannah Scott, U. of California, Santa Barbara
Vladimir Shalaev, Purdue U.
Beth Shapiro, U. of California, Santa Cruz
Jay Shendure, U. of Washington
Brian Shoichet, U. of California, San Francisco
Robert Siliciano, Johns Hopkins U. School of Medicine
Uri Simonsohn, U. of Penn.
Alison Smith, John Innes Centre
Richard Smith, U. of North Carolina at Chapel Hill (\$) **Mark Smyth, QIMR Berghofer**
Pam Soltis, U. of Florida
John Speakman, U. of Aberdeen
Tara Spire-Jones, U. of Edinburgh
Allan C. Spradling, Carnegie Institution for Science
Eric Steig, U. of Washington
Paula Stephan, Georgia State U.
V. S. Subrahmanian, U. of Maryland
Ira Tabas, Columbia U.
Sarah Teichmann, U. of Cambridge
Shubha Tole, Tata Inst. of Fundamental Research
Wim van der Putten, Netherlands Inst. of Ecology
Bert Vogelstein, Johns Hopkins U.
David Wallach, Weizmann Inst. of Science
Jane-Ling Wang, U. of Maryland
David Waxman, Fudan U.
Jonathan Weissman, U. of California, San Francisco
Chris Wikle, U. of Missouri (\$) **Terrie Williams, U. of California, Santa Cruz**
Ian A. Wilson, The Scripps Res. Inst. (\$) **Timothy D. Wilson, U. of Virginia**
Yu Xie, Princeton U.
Jan Zaanen, Leiden U.
Kenneth Zaret, U. of Penn. School of Medicine
Jonathan Zehr, U. of California, Santa Cruz
Maria Zuber, MIT

Obfuscating with transparency

Transparency is critical when it comes to decision-making that broadly affects the public, particularly when it comes to policies purported to be grounded in scientific evidence. The scientific community has been increasingly focused on improving the transparency of research through initiatives that represent good-faith efforts to enhance the robustness of scientific findings and to increase access to and utility of data that underlie research. Yet, concerns about transparency associated with scientific results continue to emerge in political discussions. Most recently in the United States, a new proposal by the Environmental Protection Agency (EPA) would eliminate the use of publications in its policy discussions for which all underlying data are not publicly available. Here, a push for transparency appears actually to be a mechanism for suppressing important scientific evidence in policy-making, thereby threatening the public's well-being.

Under the new policy, studies that do not fully meet transparency criteria would be excluded from use in EPA policy development. This proposal follows unsuccessful attempts to enact the Honest and Open New EPA Science Treatment (HONEST) Act and its predecessor, the Secret Science Reform Act. These approaches undervalue many scientific publications and limit the impact of valuable information in developing policies in the areas that the EPA regulates.

Increasingly, many publications, including those from the *Science* family of journals, are linked to underlying data in accessible forms in repositories where they are readily available to interested parties, particularly those who seek to reproduce results or extend the analysis. The details of data deposition depend on the types of data involved, including those related to research on human subjects, as well as on the availability and maturity of relevant repositories, among other factors. Nonetheless, many publications are not explicitly linked to their underlying data for a variety of reasons, including, for example, restrictions related to the use of human subjects data put in place prior to data collection. Under the

proposed transparency rules, publications based on such data would not be considered in policy discussions.

As a core skill, scientists are trained in judging research publications even without access to all the underlying data. Many factors are considered in analyzing research papers, including judging the articulation and logic of the research design, clarity of the description of the methods used for data collection and analysis, and appropriate citation of previous results. This does not necessarily require that scientists scrutinize the raw data. Most publications address potential sources of error and uncertainty, which

can be used to judge reliability of the results. Publications also increasingly disclose conflicts of interest that might influence the authors' approaches to data collection, interpretation, or conclusions. Of course, scientific progress depends not just on individual publications but on the accumulation of evidence from multiple sources. Scientists integrate results across multiple publications. If several publications address overlapping or similar questions, scientists judge the individual publications and then combine the results to generate interpretations that are consistent with the reliable observations across the entire set.

In developing effective policies, earnest evaluations of facts and fair-minded assessments of the associated un-

certainities are foundational. Policy discussions require an assessment of the likelihood that a particular observation is true and examinations of the short- and long-term consequences of potential actions or inactions, including a wide range of different sorts of costs. Those with training in making these judgments with access to as much relevant information as possible are crucial for this process. Of course, policy development requires considerations other than those related to science. Such discussions should follow clear assessment after access to all of the available evidence. The scientific enterprise should stand up against efforts that distort initiatives aimed to improve scientific practice, just to pursue other agendas.

—Jeremy Berg



Editor-in-Chief,
Science Journals.
jberg@aaas.org



“These approaches...limit the impact of valuable information in developing policies...”

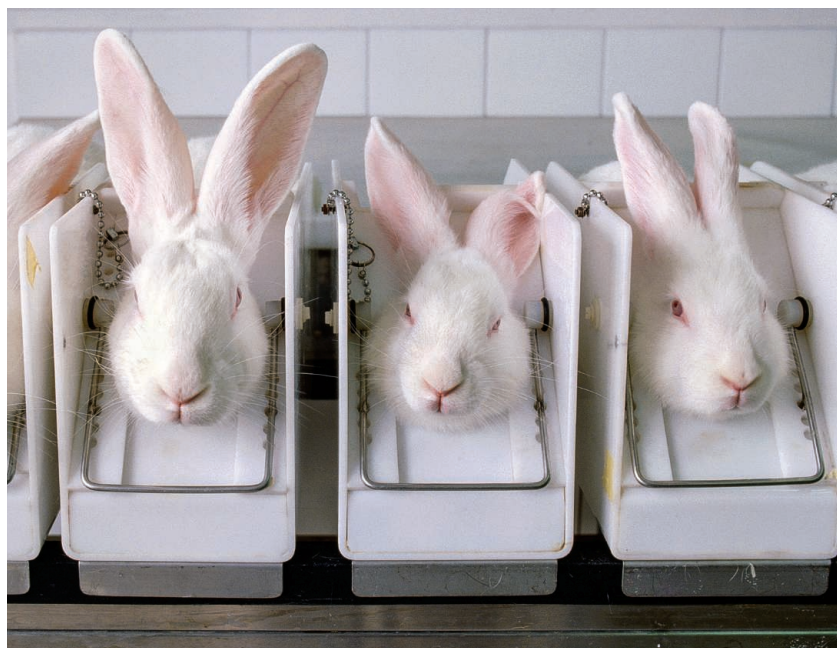
Number of Spanish scientists and citizens on an online petition prepared for the nation's parliament decrying scarce science funding, truncated careers for young researchers, and "the progressive abandonment of Spanish science."

IN BRIEF

Edited by Jeffrey Brainard

ANIMAL RESEARCH

USDA restores lab animal counts



Newly released inspection reports have resumed showing animal inventories.

Weeks after a rebuke from Congress over a lack of transparency, the U.S. Department of Agriculture (USDA) has restored details in its most recent animal welfare inspection reports. The agency sparked an outcry early last year when it scrubbed from its public database tens of thousands of reports about the treatment of lab animals housed at research institutions and companies. When USDA revived the database last August and began posting new reports, it omitted some information, including inventories that list the number and species of animals housed at a facility. In a report accompanying USDA's 2018 spending bill, lawmakers wrote that the redactions violate previous congressional directives and make it hard to track the agency's findings and activities. Newly posted inspection reports, dated March, appear to be the first since August 2017 to show animal inventories. A USDA spokesperson said the inventories haven't been included because the agency has been reviewing them for accuracy, but that it intends to include them in the future.

China tightens grip on data

SCIENCE POLICY | In a move few scientists saw coming, China's powerful State Council has decreed that all scientific data generated in that nation must be submitted to government-sanctioned data centers before appearing in publications. The rules, posted last week, apply to all groups and individuals generating research data in China. The directive calls for open access and data sharing but exempts from that provision data involving state and business secrets, national security, and individual privacy. The U.S. National Science Foundation (NSF) has qualms about the new regulation. "NSF bases its funding and its international collaboration on the principle of the freedom for scientists to publish all the data they generate with U.S. funding, regardless of where the data are collected," Nancy Sung, head of NSF's Beijing office, told *Science*. "We would be concerned about any potential impact to this principle."

Controversial linguist to return

WORKPLACE | Florian Jaeger, a linguist at the University of Rochester (U of R) in New York whose sexual behavior and comments involving students and colleagues sparked three investigations, an active lawsuit, and the resignation of U of R President Joel Seligman, will begin teaching again this fall. Jaeger will teach one upper-level course and supervise his graduate students and research lab, the university said on 2 April. "We acknowledge that there may be a negative reaction from some to Professor Jaeger's return to teaching," university spokesperson Sara Miller said in a statement that emphasized the university's belief that "people can learn and improve." Two months ago, the U of R Faculty Senate censured Jaeger for his interactions with students and colleagues but stopped short of calling for his ouster. In January, a university-hired team led by a former federal prosecutor concluded that Jaeger had not sexually harassed students or colleagues, or violated any university policies in place at the time of his alleged infractions.



CONSERVATION

Forecasts of bird migrations could prevent deaths

Scientists have developed a forecasting system for bird migrations that could help save millions of birds by reducing collisions with wind turbines and buildings. Billions of birds fly vast distances each spring to spend the summer in North America; most travel at night and can be attracted to brightly lit buildings. The warning system, developed by the Cornell Lab of Ornithology, could help alert managers all over the country when to turn off unnecessary lights or shut down wind farms. The lab has for years issued regional forecasts on its BirdCast website, based on

weather reports and local sightings of birds. In a 1 April preprint on bioRxiv, the researchers describe how they scaled up and automated these forecasts. By comparing 23 years of radar data on migrations with air temperature and other atmospheric factors, they constructed a computer model that can reliably predict bird movements up to 3 days in advance, and they suspect that longer forecasts are possible. "It's impressive how precise the prediction is, and how accurate," says Wouter Vansteelandt, a freelance bird migration researcher in Bennekom, the Netherlands, who was not involved in the research.

A relic from an early traveler

ANTHROPOLOGY | For more than a decade, researchers have scoured the Arabian Desert for evidence that some of the earliest *Homo sapiens* passed through. Now, they may have it. An ostensibly modern human finger bone uncovered in Saudi Arabia in 2016 (viewed from four sides, below) has been dated to about 88,000 years old. If the species identification holds up, this would be the oldest directly dated *H. sapiens* fossil found outside Africa or the neighboring eastern Mediterranean. The bone, found near the former banks of a freshwater lake in the Nefud Desert, supports the idea that early modern humans

spread into Eurasia in multiple waves, much earlier than the 50,000 to 60,000 years ago that some scientists have suggested. Though some scientists aren't fully convinced the finger is human, its date appears unimpeachable, says John Shea, an anthropologist at the State University of New York in Stony Brook who studies human origins.

India's twist on plagiarism

RESEARCH INTEGRITY | The Indian government has adopted its first regulations on academic plagiarism, which some researchers say are too lenient and others fear go too far and will be difficult to implement. The University Grants Commission of India (UGC India), which oversees higher education, decided last week that a small amount of plagiarism—10% of a thesis, book, or research paper—is acceptable, but that more extensive copying will result in increasingly severe punishments. Students can copy up to 60% of an original source before they are kicked out of a program, while faculty members who plagiarize to that extent will lose 2 years of pay increases and be banned from supervising students for 3 years. "Is this a joke?" asks one prominent Indian scientist who wants UGC India to reconsider its decision. But India's Society for Scientific Values in

New Delhi, which monitors academic misconduct, told the commission last fall that it's not uncommon for authors to use content from other sources in the methods section of their manuscripts, and that forcing researchers to paraphrase would "only lead to more confusion." The regulations also require universities to create a two-step investigation and appeal process. India still lacks any formal definition of scientific misconduct or mechanism for dealing with it.

Mars orbiter starts sniffing

PLANETARY SCIENCE | The long wait is over. Although the European Space Agency's ExoMars Trace Gas Orbiter (TGO) arrived at the Red Planet in October 2016, the spacecraft has only now settled into an orbit suitable for science. Initially, the probe was in a highly elliptical orbit, swooping from an altitude of 98,000 kilometers to just 200 kilometers above the surface. Over 16 months, these dips into the thin upper atmosphere produced a drag on the TGO's solar arrays that gradually slowed it down. Now in a circular, 400-kilometer orbit, the TGO will in a few weeks begin sniffing out trace gases that make up less than 1% of the atmosphere, such as methane. Mapping the distribution of methane could reveal geologic or even microbial sources.



The specimen is said to be the oldest *Homo sapiens* fossil discovered outside of Africa.

PHOTOS: (TOP TO BOTTOM) MINT IMAGES LIMITED/ALAMY STOCK PHOTO; IAN CARTWRIGHT

THREE Qs

Will AI get the blues?

As artificial intelligence (AI) allows machines to perform more tasks that humans do, will it experience similar psychological quirks, such as depression? Zachary Mainen, a neuroscientist at the Champalimaud Centre for the Unknown, a neuroscience and cancer research institute in Lisbon, thinks so. He spoke last month at a symposium at New York University in New York City where neuroscientists and AI experts discussed overlaps in the way humans and machines think. He talked with *Science* afterward. This interview has been edited for brevity and clarity; a longer version is available at <https://scim.ag/sadAI>.

Q: Why do you think AI might get depressed?

A: I'm drawing on the field of computational psychiatry, which assumes we can learn about a patient who's depressed or hallucinating from studying AI algorithms like reinforcement learning. If you reverse the arrow, why couldn't an AI be subject to the sort of things that go wrong with patients?

Q: Why might studying the effect of the neurotransmitter serotonin on the human brain shed light on machine emotions?

A: If serotonin is helping solve a general problem for intelligent systems, then machines might implement a similar function, and if serotonin goes wrong in humans, the equivalent in a machine could also go wrong. Similar issues face a person or an AI whenever the environment changes radically. Serotonin seems to help the brain in such new situations to rewire itself and get rid of old habits. So humans—or machines—with low serotonin or its equivalent may fail to rewire adequately, getting stuck in the rut that we call depression.

Q: Could AI have emotions, too?

A: Yes, I think robots would likely have something like emotions. Picture a robot sent to a distant planet to collect specimens with no help from Earth. Say it has a hardware malfunction—something in its arm is broken. It may get frustrated at first. Eventually, if it lacks the flexibility to change its algorithms or adopt new goals, it may even get depressed—[and] stop trying.

S [SCIENCEMAG.ORG/NEWS](https://scim.ag/news)
Read more news from *Science* online.

NIH's billion-dollar opioid plan

PUBLIC HEALTH | The National Institutes of Health (NIH) unveiled last week a \$1.1 billion blueprint for how it will tackle the opioid crisis that now results in more than 115 fatal overdoses a day. Called Helping to End Addiction Long-Term, the initiative includes a clinical trials network for testing nonaddictive pain medicines. Another component will seek better ways to prevent opioid misuse and treat the 2 million Americans who are addicted to such drugs. The funding includes NIH's existing annual budget of \$600 million for opioid research and \$500 million in new money that NIH received for fiscal 2018.

Seed theft draws 10-year term

BIOTECHNOLOGY | A plant breeder in Kansas was sentenced last week to 10 years in prison for conspiring to steal bioengineered rice seeds from his employer and pass them to researchers in China. Weiqiang Zhang was accused of plotting to pilfer trade secrets from Ventria Bioscience, a firm headquartered in Fort Collins, Colorado, that developed bioengineered rice that produces human proteins used in drugs and other therapeutic products. Zhang, a U.S. permanent resident who was working for Ventria as a rice breeder, kept hundreds of proprietary seeds in his home, giving some to a delegation of Chinese crop scientists in August 2013. When the visitors were leaving for home, U.S. customs officers discovered Ventria seeds in their luggage. A jury found Zhang guilty last year. A co-defendant,

U.S. Department of Agriculture geneticist Wengui Yan in Stuttgart, Arkansas, earlier pleaded guilty to one count of making false statements. Department of Justice officials are touting Zhang's sentence as a victory in the fight against intellectual property theft from China.

Climate panel diversifies

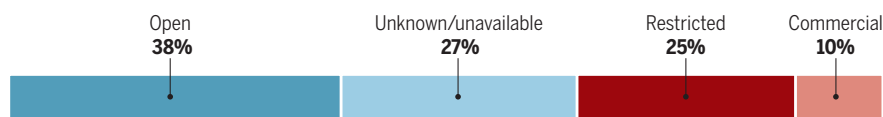
CLIMATE SCIENCE | More female scientists and researchers from developing nations will serve as authors on the *Sixth Assessment Report* of the United Nations Intergovernmental Panel on Climate Change (IPCC), its definitive review of global warming's physical basis, impacts, and potential solutions. Some 33% of the authors will be women, up from 21% in the last report; 44% come from developing countries, compared with 37% previously, IPCC announced last week. The three legs of the report will be finalized in 2021, with a synthesis published in 2022, in time for a review of global progress toward cutting greenhouse gas emissions enough to meet the long-term goals of the Paris agreement.

NASA chief scientist named

SCIENCE POLICY | Robert Lightfoot, NASA's outgoing acting administrator, announced this week that Jim Green, the agency's head of planetary science since 2006, will become NASA's chief scientist effective 1 May. Green succeeds Gale Allen, who has held the office in an acting capacity since 2016. Green will advise NASA's leadership and represent the agency as it pursues new missions to the moon.

REMOTE SENSING

Countries fail to share satellite climate data



Availability of data from 458 government-operated satellites

From 1957 to 2016, space-faring nations launched 458 government-operated, Earth-observing satellites, which gather data for weather forecasts and climate studies. But data from just 38% of the satellites are shared without restrictions, Mariel Borowitz, a space policy researcher at the Georgia Institute of Technology in Atlanta, notes in her new book *Open Space: The Global Effort for Open Access to Environmental Satellite Data*. Whereas Europe and the United States have set the standard for open data, she says, Russia and Japan tend to restrict their availability, for example, by requiring agreements and conditions that can be cumbersome. And sometimes countries attempt to sell satellite data, as in the case of Canada's RadarSat series. Nations less experienced in launching satellites often build them as technology demonstrations, with little thought to data dissemination. Still, Borowitz notes, data sharing is on the rise. "It's getting significantly better."

HUMAN EVOLUTION

Latin America's lost histories revealed

Clues to forgotten
migrations emerge from
today's genomes

By **Lizzie Wade**, in Austin

If you walked the cobblestone streets and bustling markets of 16th and 17th century Mexico City, you would see people born all over the world: Spanish settlers on their way to mass at the cathedral built atop Aztec ruins. Indigenous people from around the Americas, including soldiers who had joined the Spanish cause. Africans, both enslaved and free, some of whom had been among the first conquistadors. Asians, who traveled to Mexico on Spanish galleons, some by choice and some in bondage. All these populations met and mingled for the first time in colonial Latin America.

Historical documents describe this cultural mixture, but now international teams of researchers are enriching our view by analyzing the genomes of people today. Aided by sophisticated statistics and worldwide genetic databases, they can tease apart ancestry and population mixing with more nuance than ever before. The results, reported at a meeting here this week and in a preprint, tell stories of Latin America that have been largely forgotten or were never recorded in historical documents. From the immigration of enslaved Filipinos to that of formerly Jewish families forbidden to travel to the colonies, hidden histories are emerging.

"It's helping us to recognize the ways that really fine-scale historical experiences and practices have left this deeply significant im-

print on our genomes," says Deborah Bolnick, an anthropological geneticist at the University of Texas here.

Juan Esteban Rodríguez, a graduate student in population genetics at the National Laboratory of Genomics for Biodiversity (LANGEBIO) in Irapuato, Mexico, initially planned to study a recent thread in the global tapestry that is Mexican ancestry. Starting in the 19th century, many Chinese immigrants moved to Mexico to construct railroads in the country's northern states. Growing up near the U.S. border, Rodríguez knew this history well, and he wanted to see whether he could identify the Chinese immigrants' genetic contribution to the modern Mexican population.

But when he searched a database of 500 Mexican genomes—initially assembled for biomedical studies—and sought genetic variants more common in Asian populations, he found a surprise. Some people from northern Mexico did have significant Asian ancestry, but they weren't the only ones. Rodríguez discovered that about one-third of the people sampled in Guerrero, the Pacific coastal state that lies nearly 2000 kilometers south of the U.S. border, also had up to 10% Asian ancestry, significantly more than most Mexicans. And when he compared their genomes to those of people in Asia today, he found that they were most closely related to populations from the Philippines and Indonesia.

Rodríguez and his adviser, Andrés Moreno-Estrada, a population geneticist at

British ships often harassed Spanish galleons, which ferried long-forgotten peoples to Latin America, including enslaved Filipinos and former Jews.

LANGEBIO, turned to the historical record to figure out who these people's ancestors might be. They learned from historians who study ship manifests and other trade documents that during the 16th and 17th centuries, Spanish galleons sailed between Manila and the port of Acapulco in Guerrero, carrying goods and people, including enslaved Asians. Although historians knew of this transpacific slave trade, the origins of its victims were lost. Once they landed in Mexico, they were all recorded as "*chinos*"—Chinese, says Moreno-Estrada, who will present the work this weekend at the American Association of Physical Anthropologists (AAPA) annual meeting here. "We're uncovering these hidden stories of slavery and people who lost their identities when they disembarked in a whole new country."

Other researchers study the legacy of another marginalized group in colonial Mexico: Africans. Tens of thousands of enslaved and free Africans lived in Mexico during the 16th and 17th centuries, outnumbering Europeans, and today almost all Mexicans carry about 4% African ancestry. The percentage is much higher in some communities, says geneticist María Ávila-Arcos of the International Laboratory for Human Genome Research in Juriquilla, Mexico. She

found that in Afro-descendent communities in Guerrero and Oaxaca, many of which remain isolated, people had about 26% African ancestry, most of it from West Africa.

Other data also suggest a strong African presence in colonial Mexico. Bioarchaeologist Corey Ragsdale of Southern Illinois University in Edwardsville and his colleagues examined skeletons for dental and cranial traits that tend to be more common among Africans. They estimated that 20% to 40% of the people buried in cemeteries in Mexico City between the 16th and 18th centuries had some African ancestry, as they will present this weekend at the AAPA meeting. “It could be that Africans played as much of a role in developing population structure, and in fact developing the [Spanish] empire, as Europeans did,” Ragsdale says.

Ávila-Arcos hopes to use genetic data to trace the ancestors of those in her study back to specific West African groups or regions. She’s also found significant Asian ancestry in some of her volunteers, likely an echo of communities once formed by enslaved Africans and Asians on the Pacific coast.

Some Europeans carried hidden histories with them to colonial Latin America. A preprint recently posted on the bioRxiv server used genetic data from more than 6500 people born in Brazil, Chile, Colombia, Mexico, and Peru to tease apart how specific Native American groups and multiple populations from the Iberian peninsula contributed to modern genomes. “It’s undoubtedly the most comprehensive genetic analysis of Latin American populations to date,” Ávila-Arcos says. (The authors declined to comment because the paper has been submitted to a peer-reviewed journal.) One striking finding was that genetic variants common in the eastern Mediterranean and North Africa, and especially in Sephardic Jews, show up all over Latin America, in nearly a quarter of the individuals sampled.

The authors, led by geneticists Andrés Ruiz-Linares of Fudan University in Shanghai, China, and Garrett Hellenthal of University College London, trace a significant portion of this ancestry to *conversos*, or Jews who converted to Christianity in 1492, when Spain expelled those who refused to do so. *Conversos* were prohibited from migrating to the Spanish colonies, though a few are known to have made the trip anyway. But widespread Sephardic ancestry in Latin America implies that migration was much more common than records suggest.

For Ragsdale, the work serves as a reminder that even migrations scientists think are well understood can contain surprises. “The way we think about colonization is simplified,” Ragsdale says. “We’re missing a lot of subtleties here.” ■



ASTRONOMY

Alpha Centauri’s siren call has frustrated planet hunters

The nearest sunlike stars have failed to yield exoplanets so far, but searches for Earth-like ones are ramping up

By **Daniel Clery**, in Liverpool, U.K.

Alpha Centauri, a three-star system just 4 light-years away that is the sun’s nearest neighbor, ought to be a great place to look for Earth-like planets. But last week, at a meeting of the European Astronomical Society (EAS) here, astronomers lamented the way the system has thwarted discovery efforts so far—and announced new efforts to probe it. “It’s very likely that there are planets,” says Pierre Kervella of the Paris Observatory in Meudon, France, but the nature and positions of the stars complicate the search. “It’s a little frustrating for planet searchers.”

The system’s two sunlike stars, Alpha Centauri A and B, orbit each other closely while Proxima Centauri, a tempestuous red dwarf, hangs onto the system tenuously in a much more distant orbit. In 2016, astronomers discovered an Earth-mass planet around Proxima Centauri (*Science*, 26 August 2016, p. 857), but the planet, blasted by radiation and fierce stellar winds, seems unlikely to be habitable. Astrobiologists think the other two stars are more likely to host temperate, Earth-like planets.

Maksym Lisogorskyi, an astronomer at the University of Hertfordshire in Hatfield, U.K., tried to find them with an instrument on the European Southern Observatory’s (ESO’s) 3.6-meter telescope in Chile. He

and his colleagues looked for Doppler shifts in the spectral lines of the stars’ light that would be caused if a planet tugged them back and forth. But Lisogorskyi told the meeting that the stars’ surfaces are turbulent, and prone to flares that also jiggle the spectral lines, masking the subtle signals from any Earth-size planets. “The lines do all kinds of things,” he says. Although Alpha Centauri has been a primary target for the planet-finding instrument since it was inaugurated in 2005, it has seen nothing so far.

Also hampering observations are the current positions of the two stars. As viewed from Earth, they are very close together, making them harder to study individually, Lily Zhao of Yale University told the meeting. More precise observations should become possible as their 80-year orbit carries them farther apart. In the meantime, Zhao and her colleagues have succeeded in ruling out the presence of giant planets around either star, based on a decade’s worth of data from three instruments on different telescopes. “There are no Jupiters in the system, but there may be plenty of Earth-sized planets still to discover,” she said.

In a binary system like Alpha Centauri the lack of giant planets in Jupiter-like orbits is no surprise, because the gravity of each star would tend to kick any such planets orbiting the other star out of the system, Kervella says. But he says that temperate planets in the

The Very Large Telescope in Chile will target Alpha Centauri, which glows brightly in the southern sky.

habitable zone, closer in, would be immune to these perturbations. A chance to get a close look is coming soon: Kervella's team mapped out the system's trajectory and found that in a decade, Alpha Centauri A will pass in front of a more distant star and act as a gravitational lens, distorting the light of the star behind it. How the light from the distant star flickers and mutates over time will provide a wealth of information about any inner planets. By that time, ESO's 39-meter Extremely Large Telescope is expected to be operating and capable of observing the distortion in detail. "We will see all the planets, big and small," says astronomer Hans-Ulrich Käufl of ESO in Garching, Germany.

The privately funded Breakthrough Initiatives wants an even closer look. In 2016, the organization announced its Starshot program, a \$100 million effort to equip a microchip-size spacecraft with a camera and light-sails. A blast of photons from a giant ground-based laser would accelerate the craft to 20% of the speed of light, allowing it to make the 4-light-year trip in 20 years. During a flyby that might last only seconds, it would snap close-ups of the Alpha Centauri planets—assuming they exist.

Finding targets for the Starshot is one aim of a Breakthrough-funded effort that ESO announced last year: adapting an existing instrument on the Very Large Telescope in Chile to directly image possible planets. Called VISIR, the instrument will be equipped with a coronagraph—a mask to block out the light of the star so that the much fainter planets can be seen. VISIR observes in the midinfrared, an advantage for imaging a temperate planet because the disparity in brightness between the dim planet and its brilliant parent star is smaller in this part of the spectrum. The ESO team is testing the hardware and hopes to start observing in mid-2019 with 100 hours of dedicated telescope time.

Others at the EAS meeting think the fastest and cheapest way to detect an Earth-like planet around either of the sunlike stars is with a space telescope. A privately backed organization called Project Blue is seeking \$70 million to build and launch a 50-centimeter telescope that would stare at Alpha Centauri. Last year, the project raised \$150,000 through crowdfunding to design the spacecraft. Franck Marchis, an astronomer at the SETI Institute in Mountain View, California, a partner with Project Blue, says such a telescope, outfitted with a coronagraph, would be able to obtain an image. "It's doable. The technology is there," Marchis said. "The goal is to image a pale blue dot." ■

DRUG DEVELOPMENT

Chemists seek antiaddiction drugs to battle hijacked brain

Candidate medications could help addicts overcome cravings that lead to relapse and death

By Robert F. Service

Powerful chemical countermeasures could one day enter the battle against opioid addiction, which killed more than 42,000 people in the United States in 2016. Doctors and first responders already use medications to combat the effects of opioids, including the high and the slowed breathing of an overdose. But the new candidate drugs target the neural circuitry of addiction itself.

A compound known as OV329 is the latest addition. In animal studies it quiets the brain's reward system, sharply reducing cravings and halting addicted animals' tendency to self-administer cocaine and other habit-forming drugs. Other drugs in the pipeline also target the reward system, albeit through a different mechanism. All raise hopes that doctors could soon have a new way to treat addiction, and not just to drugs and alcohol. The medicines could potentially also be used to fight food and gambling addictions.

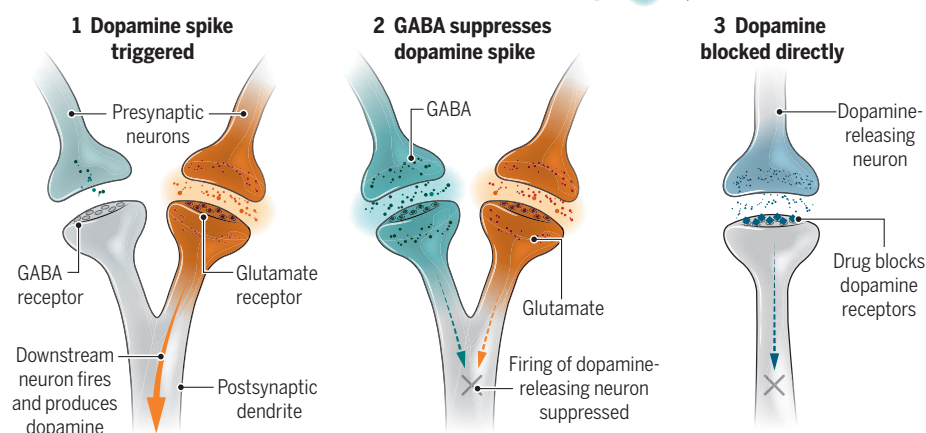
"It's a great unmet medical need," says Richard Silverman, a chemist at Northwestern University in Evanston, Illinois,

who developed OV329. OV329 has now been picked up by Ovid Therapeutics in New York City, which is continuing animal studies and hopes to launch human trials of the would-be drug. "It's a very interesting compound and clearly very promising," says Andrea Hohmann, a neuroscientist at Indiana University in Bloomington who is not involved in the work.

Addiction occurs when drugs or other pleasurable stimuli hijack the brain's normal reward system, which has evolved to reinforce beneficial behaviors, such as eating food and having sex. Such behaviors produce spikes in the release of the neurotransmitter dopamine in brain regions that are associated with motivation. Opioids and other drugs activate other neural receptors and create a euphoric high, which in turn triggers addiction-forming dopamine spikes. When addicts try to quit, encountering familiar scenes associated with drug-taking can trigger dopamine spikes, leading to cravings that make it hard to stay clean. "Their brain is constantly reminding them of how good it felt," says Brett Abrahams, Ovid's director of preclinical biology. "That's what we're fighting against."

Off the hook

Addictive drugs hijack the brain's reward system by triggering dopamine spikes (1), which lead to cravings. Candidate treatments block the spikes by boosting the inhibitory transmitter GABA (2) or blocking dopamine receptors (3).



Most existing medications for addiction try to counter the effects of specific drugs. For example, buprenorphine, used in fighting opioid addiction, binds to opioid receptors without providing as much euphoria as the opioid. At the same time, it satisfies some of the addict's cravings by triggering the dopamine reward system. But addicts on buprenorphine continue to want the real thing, and they often raise the level of opioids they take in an attempt to get high, causing an overdose.

OV329 and other compounds come at the problem from another direction. OV329 blocks an enzyme, called GABA-AT, that breaks down GABA, an "inhibitory" neurotransmitter that helps suppress the firing of some neurons. The resulting higher GABA levels dampen the neural firing of dopamine-releasing neurons—and thereby block the brain's reward system.

OV329 isn't the first drug to act this way. A GABA-AT-blocking drug called vigabatrin is already on the market to treat epilepsy by calming overactive neurons. It has also been studied as a possible antiaddiction medication, but the results have been mixed

a blockbuster drug to treat epilepsy and muscle pain, has gone one step further. In a paper published on 30 January in the *Journal of the American Chemical Society*, he and his colleagues report slightly tweaking the structure of CPP-115 to create OV329. OV329 binds more tightly to GABA-AT than its predecessor did—and that makes it 10 times more potent. When the researchers gave OV329 to cocaine- or nicotine-addicted rats, dopamine spikes were neutralized, blocking the addictive reward response and halting the animals' self-administration of the drugs. "It's extremely exciting," says Stephen Dewey, a neuroscientist at New York University's School of Medicine in New York City who led trials on vigabatrin and is now collaborating with Silverman.

Other compounds in the new wave of would-be addiction medications target the dopamine reward pathway more directly, by blocking a subset of dopamine receptors known as D3 receptors (D3Rs) that are abundant in brain regions associated with motivation and reward. Imaging studies have shown that people with cocaine addictions have even higher levels of D3Rs in those brain regions. And in recent years several studies have shown that D3R-blocking drugs sharply reduce an animal's propensity to self-administer drugs such as cocaine, methamphetamine, and opioids. Unfortunately, several such D3R blockers have drawbacks, among them that the compounds don't persist long enough in the body.

But that problem may be on the way to being solved. Last year, for example, researchers led by Amy Hauck Newman, a medicinal chemist at the National Institute on Drug Abuse in Baltimore,

Maryland, reported in *Neuropharmacology* that two of the group's newer D3R blockers, CAB2-015 and BAK4-5.4 appeared highly stable and potent. When rats hooked on the common opioid pain medication oxycodone were given the D3R blockers, the animals sharply reduced their drug taking.

Newman says she and her colleagues have still newer compounds in the works that also appear highly effective in rats, and studies with nonhuman primates are underway. Although it's still early days for these compounds, blocking D3Rs "is a really clever strategy and a fruitful way to go," Dewey says. ■

ARCHAEOLOGY

Ancient sites savaged in Yemen, Iraq

Firsthand accounts reveal worse damage than expected in war-torn regions

By Andrew Lawler, in Munich, Germany

A new front has opened in the destruction of archaeological heritage in the Middle East. Across northern Iraq and Syria, the Islamic State (IS) group devastated antiquities during its reign of terror starting in 2014, pulverizing classical statues such as those of Palmyra in Syria and bulldozing a 3000-year-old ziggurat at Iraq's Nimrud. The IS group has now been routed by Iraqi and Syrian forces, curbing the destruction but giving archaeologists a firsthand look at an aftermath that is grimmer than many had expected. Meanwhile, the assault on antiquity has extended to Yemen, 2000 kilometers to the south, another archaeological treasure house riven by conflict.

"Our immortal history has been wasted by wars," lamented Mohanad Ahmad al-Sayani, chair of Yemen's General Organization of Antiquities and Museums in Sana'a.

In Yemen, the cultural losses have gone largely unnoticed by the wider world but are keenly felt by archaeologists. Although the country has been far less studied than Mesopotamia, it played a critical role in the rise of empires and economies in the region starting around 1000 B.C.E., researchers said at a meeting here last week of the International Congress on the Archaeology of the Ancient Near East.

By 1200 B.C.E., the kingdom of Saba in what is now central Yemen controlled the export of frankincense, derived from a tree that grew only along the country's southern coast. The prized resin was burned for a millennium and a half in temples from Persia to Rome. The vast wealth of Saba—home to the biblical Queen of Sheba—funded impressive temples, cities, and engineering marvels. Among them was the Marib Dam, built on Wadi Adhanah in the eighth century B.C.E. to help expand agriculture in this arid region; some claim it is the world's oldest dam.

Today, Yemen is racked by civil war and Islamic extremists who, in a campaign against heresy, have destroyed ancient



Dopamine-blocking medications could target not only addictive drugs such as cocaine, but also compulsive behaviors like gambling or eating.

in people. And because the drug isn't very good at binding to its GABA-AT target, patients must take large doses, which in turn can lead to retinal damage.

In 2003, Silverman and his colleagues came up with a compound, known as CPP-115, that was 186 times more effective than vigabatrin at blocking GABA-AT. A company called Catalyst Pharmaceuticals is testing the drug to treat spasms in infants, and it has already cleared initial safety trials in people, raising hopes that it may also be useful in fighting addiction.

Now, Silverman, a veteran drug developer whose work 30 years ago led to Lyrica,

mosques in the port city of Aden, and a multidomed shrine in the Hadhramaut region (see map, right).

Bombs dropped by the Saudi-led coalition have wreaked the most damage, Al-Sayani said. The Marib Dam, in an unpopulated area far from the capital, was struck in 2015, leaving a deep gash in the well-preserved northern sluice gate. The regional museum of Dhamar in the southwest, which contained thousands of artifacts from the Himyarite Kingdom, was completely destroyed. The Himyarites conquered Saba in 280 C.E., took over the frankincense monopoly, and became key players in the expanding Indian Ocean trade between the Roman Empire and India until Ethiopian forces overthrew them in 525 C.E.

Al-Sayani showed images from a dozen flattened or severely damaged sites, including medieval castles such as Aden's Sira Fortress, and the centuries-old al-Qassimi neighborhood in Sana'a. More than 60 sites have been destroyed or severely damaged since the conflict began in 2015, Al-Sayani said, chiefly from Saudi bombings. Although some were strategic targets, he charged that the Saudi attacks were a conscious campaign to wreck Yemen's heritage and demoralize its citizens. "After 3 years of assessing the damage, I believe the bombing is being done with a purpose, since many of these sites are not suitable or useful for military use," he says.

The destruction seems deliberate, agrees archaeologist Sarah Japp of Berlin's German Archaeological Institute. "The Saudis were given information on important cultural heritage sites, including exact coordinates," by UNESCO, said Japp, who was based in Sana'a before the war. UNESCO intended to protect the sites, but she fears that the data may instead have been used for targeting. "There is no reason to say all of these [bombings] are just accidents." The Saudi embassy in Berlin and officials in Riyadh did not respond to repeated requests for comment.

Meanwhile, 2000 kilometers to the north in Syria and Iraq, the damage wrought by years of IS group control is only now coming into focus. "It is nothing short of a catastrophe," said Michel al-Maqdissi, former head of excavations in Syria's antiquities department in Damascus, who now works at the Louvre in Paris and maintains contacts in Syria.

Some of the worst reports come from Mari, a 60-hectare site on the banks of the Euphrates River that 4000 years ago was one of the world's largest cities. Just north of Sumer and the Akkadian Empire, Mari served as a key trading center for Mesopotamian goods and Anatolian metals and stone, and once boasted the best preserved early palace in the Middle East.

But no longer. Archaeologist Pascal



Some Yemenis suggest that the 2800-year-old Marib Dam, one of the country's best known ancient sites (shown before it was bombed), was deliberately targeted.

Butterlin of Pantheon-Sorbonne University in Paris, who worked at Mari for years and has gathered information from Syrian sources, displayed an image of the palace from the ground that shows near total destruction of Mari's central area. The site's ancient statues were removed to museums long ago, so the reasons behind the destruction remain murky, although the IS group's desire to profit from antiquities is well-known. A nearby large mound called Tell Medkouk was bulldozed completely to unearth objects for looting. From satellite data on the center of Mari, Butterlin estimates that looters dug some 1500 pits, many of

them more than 5 meters deep and 6 meters wide. The vehicle tracks "make it look like they had traffic jams there," he said. He suspects that thousands of looted cuneiform tablets, small figurines, and bronze objects won't show up on the art market for years, as sellers wait for international outrage to cool.

The situation is even worse at Dura-Europos, which until recently was a remarkably well-preserved city upstream of Mari. From the first century B.C.E., this city lay on the frontier of the Roman and Persian empires, which took turns controlling it, and once held both one of the world's oldest Jewish synagogues and oldest Christian

churches. "The scale of the disaster there is profound," said Chekmous Ali, a Syrian archaeologist now at the University of Strasbourg in France. "There are innumerable pits—some 9500—and the necropolis is gone."

Across the border in Iraq, the old city of Mosul once boasted a host of Islamic and Christian monuments, many destroyed or damaged during the IS group's 3 years of control. But the worst devastation came last summer, when more than 30,000 bombs and missiles hit historic buildings during the battle for the city, said Karel Nováček of Palacký University Olomouc in the Czech Republic. "The old city was

annihilated," he said at the meeting. He charges that the destruction continues, as Iraqi construction crews clear the wreckage without trying to preserve what's left or tally the damage.

"The heritage management is nonexistent," he said. "We need careful removal of the rubble, but that is not happening." His team is assembling what data they can from old reports and photographs that could provide some basis for reconstructing historic sites. He plans to lead an on-the-ground assessment in June, in hopes of providing Iraqis a chance to mend what they can of their battered cultural heritage. ■





ETHICS

Study questions animal efficacy data behind trials

Information provided to ethical review panels may often be insufficient to judge a drug's therapeutic potential

By Emma Yasinski

Before biomedical researchers can test a new therapy in humans, a review panel is typically asked to consider not just the risks, but also the potential benefits. After all, it makes little sense to expose people to a new drug or vaccine if there is little or no chance that it will do some good. This “therapeutic potential” is almost always based on preclinical studies in animals. But a new study suggests that review panels may have a hard time evaluating those studies because they don't receive nearly enough information.

That information usually comes to the institutional review board (IRB) in the form of an “investigator brochure,” a packet with relevant findings on the potential therapy. But when researchers examined more than 100 brochures provided to IRBs in Germany, they found that the vast majority included animal efficacy studies that were unpublished or potentially vulnerable to bias. And the brochures often seemed to leave out less flattering studies, the team reported on 5 April in *PLOS Biology*.

The findings point to a potentially pervasive global problem, says lead author Daniel Strech, a bioethicist at Hannover Medical School in Germany. Almost half of the examined trials, for example, were sponsored by a major pharmaceutical company that likely used the same paperwork in other countries.

Researchers not involved in the study are divided on the severity of the problem. “This is incredibly alarming,” says Shai Silberberg, director of research quality at the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland. The work “shows that decision-makers for ethics related to clinical trials don't get the information they really need.” But Gerald Batist, who has led many clinical studies at Jewish General Hospital's Segal Cancer Centre in Montreal, Canada, says it's well-known that animal models are poor predictors of human studies; that makes it important to get more compounds into clinical trials and see what really works, he says, rather than focus on the efficacy information provided to IRBs beforehand.

Clinical trials in the United States and Europe generally need approval from both an IRB and government regulatory agencies such as the U.S. Food and Drug Administration (FDA). Those agencies tend to focus on safety and leave weighing the potential benefits to IRBs, says Jonathan Kimmelman, a bioethicist at McGill University in Montreal and a co-author on the new study. But evidence is growing that potential benefits aren't always well-supported. A 2014 study published in *Nature* argued that poorly designed animal studies were used to justify human trials of several drugs against the neurodegenerative disease amyotrophic lateral sclerosis that ultimately failed.

Many trial sponsors seem to omit negative animal studies when they seek approval for a clinical trial.

Strech, Kimmelman, and their colleagues decided to study the issue more systematically. Investigator brochures aren't public information, but the chairs of three IRBs at German medical institutions agreed to share the documents with the researchers after they agreed not to identify individual trial sponsors, researchers, or candidate drugs. (Three other IRB chairs refused.) Overall, the team obtained 109 brochures from phase I and II trials approved from 2010 to 2016 that together cited 708 efficacy studies in animals.

Only 11% of those were reported in a peer-reviewed paper, the team found; most were confidential company studies, meaning the IRBs could not see the full studies and whether they had been peer reviewed was unclear. Fewer than 5% of the studies included information on whether the authors had taken steps to minimize bias, such as randomizing the animals to treatment or placebo groups, or blinding researchers to which animal was in which group. And 82% of the brochures reported only studies that showed a drug worked, suggesting sponsors left out studies finding no effect, Strech says. He is “surprised” that IRBs and regulatory agencies accept such incomplete packets. “Why is nobody complaining about this?”

It's not clear why so much information is missing, Strech says. Trial sponsors may choose not to publish studies because they're worried about tipping off competitors, and they may withhold information because they think IRBs, which usually include nonscientists, aren't equipped to judge the studies. IRBs may assume that no sponsor would invest in a clinical trial without convincing evidence, and give the companies the benefit of the doubt rather than thoroughly analyze the animal studies.

In a January 2017 *Nature* commentary, Kimmelman and his McGill colleague Carole Federico, also an author on the new study, suggested several measures to strengthen oversight. IRBs could appoint ad hoc members with specialist expertise to help evaluate individual proposals, they said. And regulatory agencies could create a special mechanism to evaluate so-called first-in-human studies, where the potential risks are highest; ban trial sponsors from cherry-picking their preclinical studies; and give better instructions on weighing efficacy evidence to the ethics panels. “If the FDA has clear guidance on what they expect from the IRB,” Silberberg says, “then it'll happen.” ■

Emma Yasinski is a science journalist in Jupiter, Florida.

GENETICS

Human mutation rate a legacy from our past

By assessing mutation rates among species, researchers are understanding why they vary

By **Elizabeth Pennisi**, in Tempe, Arizona

Kelley Harris wishes humans were more like paramecia. Every newborn's DNA carries more than 60 new mutations, some of which lead to birth defects and disease, including cancers. "If we evolved parameciumlike replication and DNA repair processes, that would never happen," says Harris, an evolutionary biologist at the University of Washington in Seattle. Researchers have learned that these single-cell protists go thousands of generations without a single DNA error—and they are figuring out why human genomes seem so broken in comparison.

The answer, researchers reported at the Evolution of Mutation Rate workshop here late last month, is a legacy of our origins. Despite the billions on Earth today, humans numbered just thousands in the early years of our species. In large populations, natural selection efficiently weeds out deleterious genes, but in smaller groups like those early humans, harmful genes that arise—including those that foster mutations—can survive.

Support comes from data on a range of organisms, which show an inverse relationship between mutation rate and ancient population size. This understanding offers insights into how cancers develop and also has implications for efforts to use DNA to date branches on the tree of life. "Clarifying why mutation rates vary is crucial for understanding all areas of biology," says evolutionary biologist Michael Lynch of Arizona State University (ASU) here.

Mutations occur, for example, when cells

copy their DNA incorrectly or fail to repair damage from chemicals or radiation. Some mistakes are good, providing variation that enables organisms to adapt. But some of these genetic mistakes cause the mutation rate to rise, thus fostering more mutations.

For a long time, biologists assumed mutation rates were identical among all species, and so predictable that they could be used as "molecular clocks." By counting differences between the genomes of two species or populations, evolutionary geneticists could date when they diverged. But now that geneticists can compare whole genomes of parents and their offspring, they can count the actual number of new mutations per generation.

That has enabled researchers to measure mutation rates in about 40 species, including newly reported numbers for orangutans, gorillas, and green African monkeys. The primates have mutation rates similar to humans, as ASU co-organizer Susanne Pfeifer reported in the December 2017 issue of *Evolution*. But, as Lynch and others reported at the meeting, bacteria, paramecia, yeasts, and nematodes—all of which have much larger populations than humans—have mutation rates orders of magnitude lower.

The variation suggests that in some species, genes that cause high mutation rates—for instance, by interfering with DNA repair—go unchecked. In 2016, Lynch detailed a possible reason, which he calls the drift barrier hypothesis. It invokes genetic drift, or chance genetic changes—"noise in the evolutionary process that is greater than the directional force" of selection, as he puts it. Genetic drift plays a bigger role

in smaller populations. In large populations, harmful mutations are often counteracted by later beneficial mutations. But in a smaller population with fewer individuals reproducing, the original mutation can be preserved and continue to do damage.

Today, 7.6 billion people inhabit Earth, but population geneticists focus on the effective population size, which is the number of people it took to produce the genetic variation seen today. In humans, that's about 10,000—not so different from that of other primates. Humans tend to form even smaller groups and mate within them. In such small groups, Harris says, "we can't optimize our biology because natural selection is imperfect."

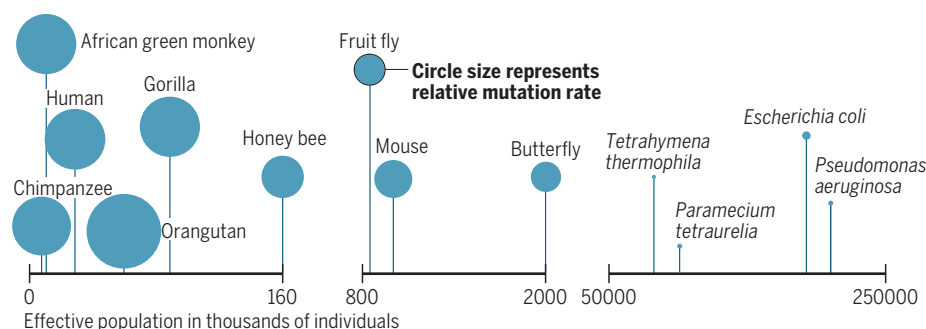
Harris detected those imperfections even among populations of people—further evidence, she notes, to support the drift barrier hypothesis. Rather than look at the overall number of DNA changes, Harris focused on the frequency of changes in each kind of DNA base in the populations she studied. That "mutation spectrum" varies widely between different groups of people, she reported. In 2017, she and her colleagues estimated that between 15,000 and 2000 years ago, Europeans had an unusually high number of some conversions of the base cytosine to thymine. She has since found differences in the mutation spectrum between Japanese and other East Asian populations. "The way the genome tends to break is not the same in Europeans" as in people elsewhere, she says.

Now, Harris and researchers at the University of Copenhagen have extended the analysis to ancient DNA. Among Europeans, the excess cytosine to thymine mutations existed in early farmers but not in hunter-gatherers, she reported. She speculates that these farmers' wheat diet may have led to nutrient deficiencies that predisposed them to a mutation in a gene that in turn favored the cytosine-to-thymine changes, suggesting environment can lead to changes in mutation rate. Drift likely played a role in helping the mutation-promoting gene stick around.

Eventually she hopes to pinpoint the pathways and the genes responsible. That's increasingly necessary, says Charles Baer, an evolutionary biologist at the University of Florida in Gainesville. It's become clear that "mutation rates can evolve pretty quickly and in all sorts of ways. If you really want to understand mutation rate, you have to put a fine magnifying glass to it." ■

The highs and lows of mutation rates

The rate at which new mutations appear in a genome (sizes of circles) is inversely proportional to the so-called effective population size of the species. Microbes (right) have the largest populations and lowest mutation rates.





FREE AGENTS

Monumentally complex models are gaming out disaster scenarios with millions of simulated people *By M. Mitchell Waldrop*

At 11:15 on a Monday morning in May, an ordinary looking delivery van rolls into the intersection of 16th and K streets NW in downtown Washington, D.C., just a few blocks north of the White House. Inside, suicide bombers trip a switch.

Instantly, most of a city block vanishes in a nuclear fireball two-thirds the size of the one that engulfed Hiroshima, Japan. Powered by 5 kilograms of highly enriched uranium that terrorists had hijacked weeks earlier, the blast smashes buildings for at least a kilometer

in every direction and leaves hundreds of thousands of people dead or dying in the ruins. An electromagnetic pulse fries cellphones within 5 kilometers, and the power grid across much of the city goes dark. Winds shear the bomb's mushroom cloud into a plume of radioactive fallout that drifts eastward into the Maryland suburbs. Roads quickly become jammed with people on the move—some trying to flee the area, but many more looking for missing family members or seeking medical help.



A plume of radioactive fallout (yellow) stretches east across Washington, D.C., a few hours after a nuclear bomb goes off near the White House in this snapshot of an agent-based model. Bar heights show the number of people at a location, while color indicates their health. Red represents sickness or death.

cluster busy for a day and a half—forcing the agents to be relatively simple-minded. “There’s a fundamental trade-off between the complexity of individual agents and the size of the simulation,” says Jonathan Pfautz, who funds agent-based modeling of social behavior as a program manager at the Defense Advanced Research Projects Agency in Arlington, Virginia.

But computers keep getting bigger and more powerful, as do the data sets used to populate and calibrate the models. In fields as diverse as economics, transportation, public health, and urban planning, more and more decision-makers are taking agent-based models seriously. “They’re the most flexible and detailed models out there,” says Ira Longini, who models epidemics at the University of Florida in Gainesville, “which makes them by far the most effective in understanding and directing policy.”

THE ROOTS of agent-based modeling go back at least to the 1940s, when computer pioneers such as Alan Turing experimented with locally interacting bits of software to model complex behavior in physics and biology. But the current wave of development didn’t get underway until the mid-1990s.

One early success was Sugarscape, developed by economists Robert Axtell of George Mason University in Fairfax, Virginia, and Joshua Epstein of New York University (NYU) in New York City. Because their goal was to simulate social phenomena on ordinary desktop computers, they pared agent-based modeling down to its essence: a set of simple agents that moved around a grid in search of “sugar”—a foodlike resource that was abundant in some places and scarce in others. Though simple, the model gave rise to surprisingly complex group behaviors such as migration, combat, and neighborhood segregation.

Another milestone of the 1990s was the Transportation Analysis and Simulation System (Transims), an agent-based traffic model developed by Barrett and others at the Los Alamos National Laboratory in New Mexico. Unlike traditional traffic models, which used equations to describe moving vehicles en masse as a kind of fluid, Transims modeled each vehicle and driver as an agent moving through a city’s road network. The simulation included a realistic mix of cars, trucks, and buses, driven by people with a realistic mix of ages, abilities, and destinations. When

It’s all make-believe, of course—but with deadly serious purpose. Known as National Planning Scenario 1 (NPS1), that nuclear attack story line originated in the 1950s as a kind of war game, a safe way for national security officials and emergency managers to test their response plans before having to face the real thing.

Sixty years later, officials are still reckoning with the consequences of a nuclear catastrophe in regular NPS1 exercises. Only now, instead of following fixed story lines and predictions assembled ahead of time, they are using computers to play what-if with an entire artificial society: an advanced type of computer simulation called an agent-based model.

Today’s version of the NPS1 model includes a digital simulation of every building in the area affected by the bomb, as well as every road, power line, hospital, and even cell tower. The model includes weather data to simulate the fallout plume. And the scenario is peopled with some 730,000 agents—a synthetic population statistically identical to the real population of the affected area in factors such as age, sex, and occupation. Each agent is an autonomous subroutine that responds in reasonably human ways to other agents and the evolving disaster by switching among multiple modes of

behavior—for example, panic, flight, and efforts to find family members.

The point of such models is to avoid describing human affairs from the top down with fixed equations, as is traditionally done in such fields as economics and epidemiology. Instead, outcomes such as a financial crash or the spread of a disease emerge from the bottom up, through the interactions of many individuals, leading to a real-world richness and spontaneity that is otherwise hard to simulate.

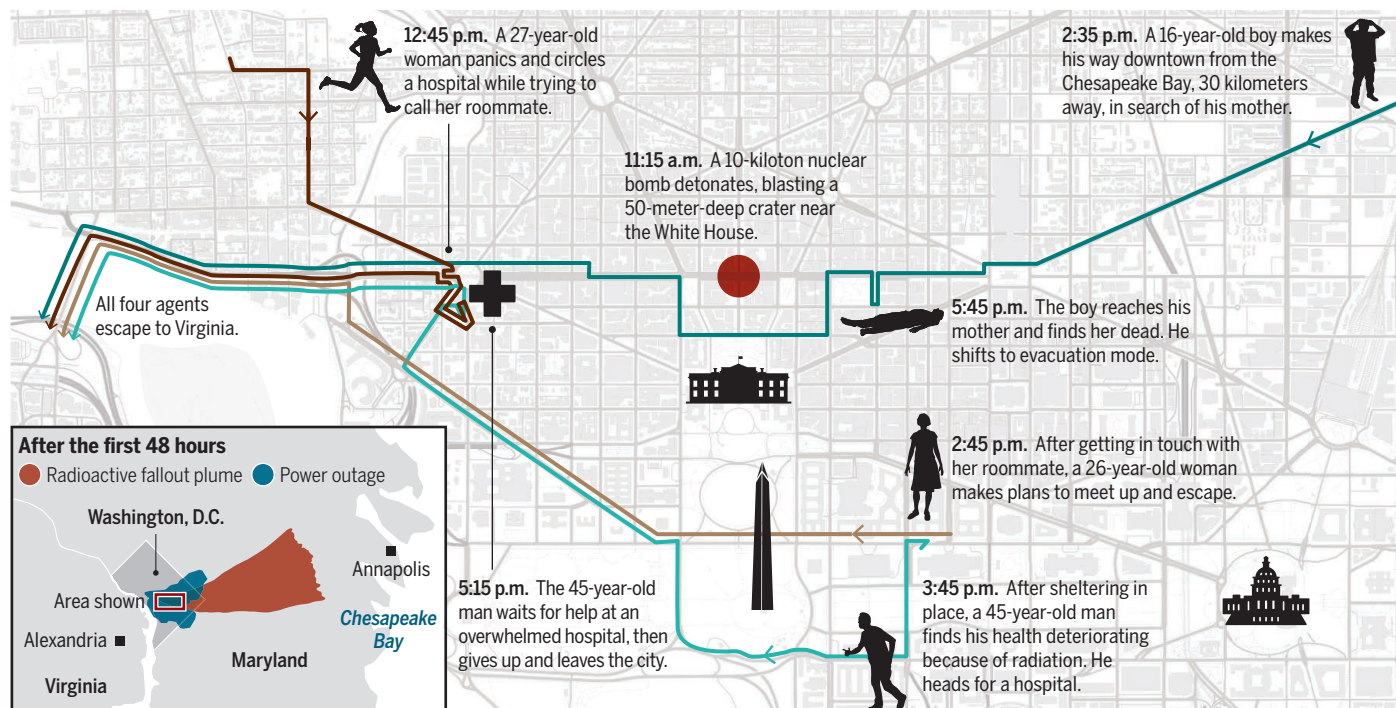
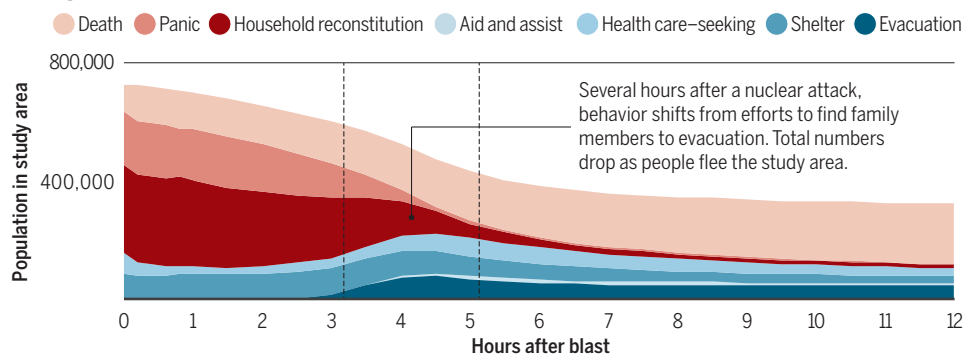
That kind of detail is exactly what emergency managers need, says Christopher Barrett, a computer scientist who directs the Biocomplexity Institute at Virginia Polytechnic Institute and State University (Virginia Tech) in Blacksburg, which developed the NPS1 model for the government. The NPS1 model can warn managers, for example, that a power failure at point X might well lead to a surprise traffic jam at point Y. If they decide to deploy mobile cell towers in the early hours of the crisis to restore communications, NPS1 can tell them whether more civilians will take to the roads, or fewer. “Agent-based models are how you get all these pieces sorted out and look at the interactions,” Barrett says.

The downside is that models like NPS1 tend to be big—each of the model’s initial runs kept a 500-microprocessor computing

A recipe for disaster

The U.S. government relies on an agent-based model to predict the effects of a nuclear attack in downtown Washington, D.C. The model contains many layers—infrastructure, transportation, weather—and hundreds of thousands of “agents” interact in this virtual landscape, changing their behavior in ways thought to mimic actual human behavior. The model helps planners identify trouble spots and assess potential damage. It also yields surprising patterns, such as some agents’ movements toward the blast in efforts to find family members.

Agent behaviors



applied to the road networks in actual cities, Transims did better than traditional models at predicting traffic jams and local pollution levels—one reason why Transims-inspired agent-based models are now a standard tool in transportation planning.

A similar shift was playing out for epidemiologists. For much of the past century, they have evaluated disease outbreaks with a comparatively simple set of equations that divide people into a few categories—such as susceptible, contagious, and immune—and that assume perfect mixing, meaning that everybody in the affected region is in contact with everyone else. Those equation-based models were run first on paper and then on computers, and they are still used widely. But epidemiologists are increasingly turning to agent-based models to include factors that the equations ignore, such as geography, transportation networks, family structure, and behavior change—all of which can strongly affect how disease spreads. Dur-

ing the 2014 Ebola outbreak in West Africa, for example, the Virginia Tech group used an agent-based model to help the U.S. military identify sites for field hospitals. Planners needed to know where the highest infection rates would be when the mobile units finally arrived, how far and how fast patients could travel over the region's notoriously bad roads, and a host of other issues not captured in the equations of traditional models.

In another example, Epstein's laboratory at NYU is working with the city's public health department to model potential outbreaks of Zika, a mosquito-borne virus that can lead to catastrophic birth defects. The group has devised a model that includes agents representing all 8.5 million New Yorkers, plus a smaller set of agents representing the entire population of individual mosquitoes, as estimated from traps. The model also incorporates data on how people typically move between home, work, school, and shopping; on sexual behavior (Zika can

be spread through unprotected sex); and on factors that affect mosquito populations, such as seasonal temperature swings, rainfall, and breeding sites such as caches of old tires. The result is a model that not only predicts how bad such an outbreak could get—something epidemiologists could determine from equations—but also suggests where the worst hot spots might be.

In economics, agent-based models can be a powerful tool for understanding global poverty, says Stéphane Hallegatte, an economist at the World Bank in Washington, D.C. If all you look at are standard metrics such as gross domestic product (GDP) and total income, he says, then in most countries you're seeing only rich people: The poor have so little money that they barely register.

To do better, Hallegatte and his colleagues are looking at individual families. His team built a model with agents representing 1.4 million households around the globe—roughly 10,000 per country—and looked at

how climate change and disasters might affect health, food security, and labor productivity. The model estimates how storms or drought might affect farmers' crop yields and market prices, or how an earthquake might cripple factory workers' incomes by destroying their cars, the roads, or even the factories.

The model suggests something obvious: Poor people are considerably more vulnerable to disaster and climate change than rich people. But Hallegatte's team saw a remarkable amount of variation. If the poor people in a particular country are mostly farmers, for example, they might actually benefit from climate change when global food prices rise. But if the country's poor people are mostly packed into cities, that price rise could hurt badly.

That kind of granularity has made it easier for the World Bank to tailor its recommendations to each country's needs, Hallegatte says—and much easier to explain the model's results in human terms rather than economic jargon. "Instead of telling a country that climate change will decrease their GDP by X%," he says, "you can say that 10 million people will fall into poverty. That's a number that's much easier to understand."

GIVEN HOW MUCH is at stake in those simulations, Barrett says, users always want to know why they should trust the results. How can they be sure that the model's output has anything to do with the real world—especially in cases such as nuclear disasters, which have no empirical data to go on?

Barrett says that question has several answers. First, users shouldn't expect the models to make specific predictions about, say, a stock market crash next Tuesday. Instead, most modelers accommodate the inevitable uncertainties by averaging over many runs of each scenario and displaying a likely range of outcomes, much like landfall forecasts for hurricanes. That still allows planners to use the model as a test bed to game out the consequences of taking action A, B, or C.

Second, Barrett says, the modelers should not just slap the model together and see whether the final results make sense. Instead, they should validate the model as they build it, looking at each piece as they slot it in—how people get to and from work, for example—and matching it to real-world data from transit agencies, the census, and other sources. "At every step, there is data that you're calibrating to," he says.

Modelers should also try to calibrate agents' behaviors by using studies of human psychology. Doing so can be tricky—humans are complicated—but in crisis situations, modeling behavior becomes easier because it tends to be primal. The NPS1 model, for example, gets by with built-in rules that cause

the agents to shift back and forth among just a few behaviors, such as "health care-seeking," "shelter-seeking," and "evacuating."

Even so, field studies point to crucial nuances, says Julie Dugdale, an artificial intelligence researcher at the University of Grenoble in France who studies human behavior under stress. "In earthquakes," she says, "we find that people will be more afraid of being without family or friends than of the crisis itself." People will go looking for their loved ones first thing and willingly put themselves in danger in the process. Likewise in fires, Dugdale says. Engineers tend to assume that when the alarm sounds, people will immediately file toward the exits in an orderly way. But just watch the next time your building has a fire drill, she says: "People don't evacuate without first talking to others"—and if need be, collecting friends and family.

"We find that people will be more afraid of being without family or friends than of the crisis itself."

Julie Dugdale, University of Grenoble

The evidence also suggests that blind, unthinking panic is rare. In an agent-based model published in 2011, sociologist Ben Aguirre and his colleagues at the University of Delaware in Newark tried to reproduce what happened in a 2003 Rhode Island nightclub fire. The crowds jammed together so tightly that no one could move, and 100 people died. Between the police, the local paper, and survivors' accounts, Aguirre's team had good data on the victims, their behavior, and their relationships to others. And when the researchers incorporated those relationships into the model, he says, the runs most consistent with the actual fire involved almost no panic at all. "We found that people were trying to get out with friends, co-workers, and loved ones," Aguirre says. "They were not trying to hurt each other. That was a happenstance."

The NPS1 model tries to incorporate such insights, sending its agents into "household reconstitution" mode (searching for friends and family) much more often than "panic" mode (running around with no coherent goal). And the results can sometimes be counterintuitive. For example, the model suggests that right after the strike, emergency managers should expect to see some people rushing toward ground zero, jamming the roads in a frantic effort to pick up children from school or find missing spouses. The model also points to a good

way to reduce chaos: to quickly restore partial cell service, so that people can verify that their loved ones are safe.

IF AGENT-BASED MODELERS have a top priority, it's to make the simulations easier to build, run, and use—not least because that would make them more accessible to real-world decision-makers.

Epstein, for example, envisions national centers where decision-makers could access what he calls a petabyte playbook: a library containing digital versions of every large city, with precomputed models of just about every potential hazard. "Then, if something actually happens, like a toxic plume," he says, "we could pick out the model that's the closest match and do near-real-time calculation for things like the optimal mix of shelter-in-place and evacuation."

At Virginia Tech, computer scientist Madhav Marathe is thinking along the same lines. When a Category-5 hurricane is bearing down, he says, someone like the mayor of San Juan can't be waiting around for a week-long analysis of the storm's possible impact on Puerto Rico's power grid. She needs information that's actionable, he says—"and that means models with a simple interface, running in the cloud, delivering very sophisticated analytics in a very short period of time."

Marathe calls it "agent-based modeling as a service." His lab has already spent the past 4 years developing and testing a web-based tool that lets public health officials build pandemic simulations and do what-if analyses on their own, without having to hire programmers. With just a few clicks, users can specify key variables such as the region of interest, from as small as a single city to the entire United States, and the type of disease, such as influenza, measles, Ebola, or something new. Then, using the tool's built-in maps and graphs, users can watch the simulation unfold and see the effect of their proposed treatment protocols.

Despite being specialized for epidemics, Marathe says, the tool's underlying geographic models and synthetic populations are general, and they can be applied to other kinds of disasters, such as chemical spills, hurricanes, and cascading failures in power networks. Ultimately, he says, "the hope is to build such models into services that are individualized—for you, your family, or your city." Or, as Barrett puts it, "If I send Jimmy to school today, what's the probability of him getting Zika?"

So it won't just be bureaucrats using those systems, Barrett adds. It will be you. "It will be as routine as Google Maps." ■

M. Mitchell Waldrop is a journalist based in Washington, D.C.

PERSPECTIVES

CLIMATE

How cleaner air changes the climate

Air quality improvements affect regional climate in complex ways

By Bjørn Hallvard Samset

Aerosols have a strong influence on the present climate, but this influence will likely be reduced over the coming decades as air pollution measures are implemented around the world. At a global level, aerosols have helped to reduce the warming effect from greenhouse gas emissions, and necessary reductions in air pollution may thus make it harder to achieve ambitious

global climate and environmental aims, such as the Paris Agreement's 2°C target. Furthermore, the local nature of air pollution means that the impacts of changes to aerosol emissions—on temperature, precipitation, extreme events, and health—are likely to differ widely from one place to another. Model and observational studies are beginning to assess these impacts, particularly the link between aerosols and precipitation, to elucidate the climate effects of cleaning up our air.

Human influence on the climate is a tug-of-war, with greenhouse gas-induced warming being held partly in check by cooling from aerosol emissions. In a Faustian bargain, humans have effectively dampened global climate change through air pollution. Increased greenhouse gas concentrations from fossil fuel use are heating the planet by trapping heat radiation. At the same time,

CICERO Center for International Climate Research, Oslo, Norway. Email: b.h.samset@cicero.oslo.no

Smog covers Lujiazui, Shanghai, China. Cleaning up air pollution affects regional air temperature and precipitation.



ent patterns of impact. A reduction in aerosol emissions—as has already occurred in the United States and Europe and is assumed to continue in most climate scenarios—can be expected to have disproportionately strong impacts near emission regions, where most of the world's population lives. The effects of global warming on society are therefore different if the warming is due to loss of aerosol cooling, rather than from greenhouse gas-induced warming. Simply put, it matters not only that we limit global warming to 2°C, but also how we do it.

Since 1990, there has been little change in the global volume of anthropogenic aerosol emissions. Regionally, however, there are large differences, with reductions in Europe and the United States balanced by increases in Africa and Asia (see the photo) (4). Recent simulations of the industrial era suggest that aerosols have prevented most surface warming from greenhouse gases in East Asia and, at the same time, changed

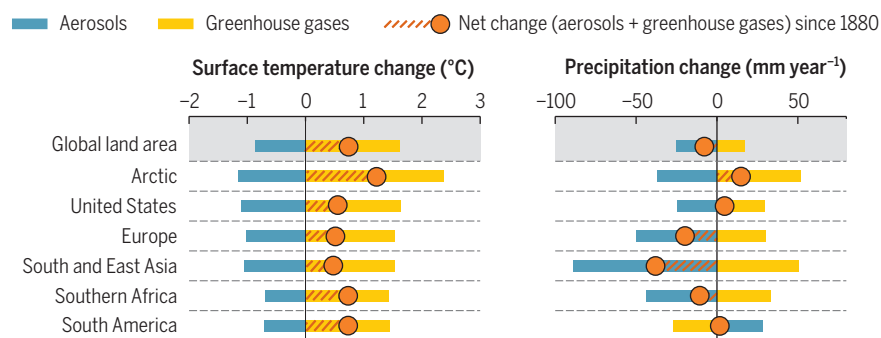
sulfate aerosols, have declined by 75% since 2007, whereas those from India increased by 50% over the same period (7).

Aerosols also affect region-specific climate and weather phenomena, such as the South Asian monsoon. Indian summer monsoon rainfall has steadily declined since the 1950s, and model simulations indicate that aerosol forcing is critical to explaining this trend (8). Aerosol-induced surface cooling is thought to lead to anomalous circulation patterns over much of the region, weakening moisture transport from the Indian Ocean and thereby reducing monsoon rainfall (9).

Furthermore, aerosols are mainly emitted over Northern Hemisphere land masses, resulting in a hemispheric asymmetry that may have driven a shift in the position of the Intertropical Convergence Zone over the past century (10). Overall, today's precipitation patterns in the Northern Hemisphere are likely markedly influenced by aerosols, both near and far from emission sources.

Tug-of-war between aerosol cooling and greenhouse gas warming

Surface temperature and precipitation have, since preindustrial times, been affected by both greenhouse gases and aerosols. Model simulations comparing the periods 1985 to 2005 and 1880 to 1900 show that across the global land area, aerosols have limited the impacts of greenhouse gas warming. The regional patterns are more complex for precipitation. Data from (14).



emissions of aerosols—particles that make up a substantial fraction of air pollution—have an overall cooling effect by reflecting incoming sunlight (1). The net effect of greenhouse gases and aerosols is the ~1°C of global warming observed since 1880 CE. The individual contributions of greenhouse gases and aerosols are, however, much more uncertain. Recent climate model simulations indicate that without anthropogenic aerosols, global mean surface warming would be at least 0.5°C higher, and that in their absence there would also be a much greater precipitation change (2, 3) (see the figure).

Many climate effects from aerosols are, however, regional rather than global. Whereas the major greenhouse gases, carbon dioxide and methane, get distributed globally, aerosols are removed from the atmosphere in a matter of days, leading to quite differ-

what would have been a precipitation increase into a marked drying (see the figure). Although there are large differences between models, these results are broadly consistent with observations (5).

Regional cooling has likely also strongly influenced the rates of occurrence of extreme events (3) and the hydrological cycle (6). Modeling cannot, however, give definitive answers regarding these effects, because the model resolution is too coarse and it remains difficult to accurately reproduce the relevant cloud processes. It therefore remains unclear how an Asian aerosol cleanup would affect local precipitation and extreme weather events such as storms and droughts. The topic is urgent because Asian emissions levels are changing rapidly. According to one recent study, Chinese emissions of SO₂, a main precursor of cooling

To add to the complexity, not all aerosols cool the climate. Carbonaceous aerosols, by-products of incomplete combustion, absorb sunlight and can therefore heat the atmosphere. The global warming effects of black carbon, the main absorbing aerosol type, are likely to be moderate, but black carbon can have substantial regional climate impacts (11). Absorbing aerosols change the temperature profile of the atmosphere and therefore also alter circulation, cloud formation, and precipitation. These processes may have contributed to the observed drying trend in Southern Africa since the 1950s (12). Also, the deposition of dark aerosols on white snow has likely contributed to the strong Arctic warming since the 1980s (13).

Currently, most anthropogenic aerosol emissions are related to fossil fuel use. The massive emission reductions necessitated

by the Paris Agreement will therefore also reduce aerosol-induced cooling. Health and air quality considerations provide further, strong motivations for rapid reductions in particle emissions. Legislation targeting air pollution, such as the U.S. Clean Air Act and the European Union's Ambient Air Quality Directive, has proven that such mitigation is possible. Despite limited regulation, aerosol concentrations are currently falling in parts of Asia, although the driving factors are incompletely understood. Health concerns may drive local and regional aerosol reductions faster than foreseen in the climate scenarios used, e.g., in the IPCC (Intergovernmental Panel on Climate Change) assessments. This, in turn, implies that reductions in greenhouse gas emission may need to be even more rapid than has been assumed, in order to meet the goals of the Paris Agreement. Policy measures may also target cooling sulfate aerosols and heating carbonaceous aerosols differently, making it even more challenging to predict the outcomes of specific mitigation strategies.

Aerosol emissions are an important component of human influence on the climate today. Fossil fuel use reductions and air quality measures make it likely that this influence will be greatly reduced over the coming decades, with consequences for the climate that may even dominate over those from greenhouse gas warming in some regions. However, understanding of the complex interactions between cooling and heating aerosols, atmospheric circulation, and precipitation patterns remains limited. The regional effects of cleaning our air, in all their complexity, must be taken into account when developing climate adaptation and mitigation strategies, if we are to be prepared for the changes to come. ■

REFERENCES

1. O. Boucher *et al.*, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T.F. Stocker *et al.*, Eds. (Cambridge Univ. Press, Cambridge/New York, 2013), chap. 7.
2. H.D. Matthews, K. Zickfeld, *Nat. Clim. Change* **2**, 338 (2012).
3. B.H. Samset *et al.*, *Geophys. Res. Lett.* **45**, 1020 (2018).
4. G. Myhre *et al.*, *Atmos. Chem. Phys.* **17**, 2709 (2017).
5. D.L. Hartmann *et al.*, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T.F. Stocker *et al.*, Eds. (Cambridge Univ. Press, Cambridge/New York, 2013), chap. 2.
6. L. Liu *et al.*, *J. Climate* **10.1175/JCLI-D-17-0439.1** (2018).
7. C. Li *et al.*, *Sci. Rep.* **7**, 14304 (2017).
8. R. Krishnan *et al.*, *Clim. Dyn.* **47**, 1007 (2016).
9. R.E. Bartlett *et al.*, *Clim. Dyn.* **50**, 1863 (2018).
10. H.E. Ridley *et al.*, *Nat. Geosci.* **8**, 195 (2015).
11. C.W. Stjern *et al.*, *J. Geophys. Res. Atmos.* **122**, 11462 (2017).
12. Ø. Hodnebrog, G. Myhre, P.M. Forster, J. Sillmann, B.H. Samset, *Nat. Commun.* **7**, 12136 (2016).
13. M. Sand *et al.*, *Nat. Clim. Change* **6**, 286 (2016).
14. A.M.L. Ekman, *J. Geophys. Res. Atmos.* **119**, 817 (2014).

10.1126/science.aat1723

SYNTHETIC BIOLOGY

Improved memory devices for synthetic cells

CRISPR enables efficient recording of signaling events in cells onto DNA

By Joanne M. L. Ho¹ and Matthew R. Bennett^{1,2}

Synthetic biologists have long sought to make cells more like computers. This is not because they think cells will be more efficient than silicon—current microelectronics make excellent computers and are less messy than cell cultures—but instead because synthetic cells can interface with biology to perform biochemical tasks. Synthetic cells might one day be capable of attacking tumors or releasing site-specific drugs inside the human body. But to carry out these tasks, synthetic biologists must be able to program cells much in the same way we

“Because the order and timing of biological events determine cell fate, memory devices that can reveal the order of cellular events are of great utility.”

program computers—by providing them with decision-making capabilities based on inputs. Indeed, prototypes of many of the genetic parts necessary for turning cells into biocomputers have been constructed, including transcriptional logic gates (1), timers (2, 3), counters (4), memory devices (5, 6), tunable sensors (7, 8), and even in vitro DNA systems that can perform complex calculations (9). On page 169 of this issue, Tang and Liu (10) expand the capabilities of cellular computers by engineering a new memory device that records events directly onto DNA.

To collect, process, and act on information, cells must be able to accurately record signals. Given the fundamental importance of memory devices to biocomputing, sev-

eral cell-based memory systems have been built in the past decade. The first synthetic memory device was the bistable toggle switch, which featured two repressible promoters arranged in a mutually inhibitory gene regulatory network. In this setup, the system can switch between either of two stable states by exposure to small molecules (11). Next, a memory device that provided a more permanent record etched in the genome was built with DNA recombinases (4). Orthogonal recombinases, enzymes that cut and paste specific pieces of DNA, have enabled implementation of logic-gated memory devices (6) even in complex environments such as the mammalian gut (12, 13). Such devices allow for real-time surveillance of complex microbiomes and may lead to the development of living diagnostics and therapeutics. However, owing to the limited number of orthogonal recombinases, their use precludes facile multiplexing of complex information systems.

An elegant solution to this problem was the development of SCRIBE (synthetic cellular recorders integrating biological events) (5), a memory device that translates exogenous signals into point mutations in a bacterial genome via reverse transcription and recombination. With the discovery and development of CRISPR for genome editing, researchers have recognized the utility of CRISPR-mediated DNA storage as a stable means of capturing large amounts of data. Recently, digitization of a movie into the genomes of living bacteria was accomplished, with no single cell encoding more than a small snippet of the movie (14). However, these memory devices face the problem of low conversion frequency, which necessitates the use of large cell populations. By using DNA base editors and high-copy number plasmids carrying the recorder DNA (see the figure), Tang and Liu have developed a sensitive recording device that needs few cells. These improvements are a considerable advance in the arena of cell-based memory systems.

Tang and Liu's memory device, called CRISPR-mediated analog multi-event recording apparatus (CAMERA), can record a variety of biological and chemical signals,

¹Department of Biosciences, Rice University, Houston, TX, USA. ²Department of Bioengineering, Rice University, Houston, TX, USA. Email: matthew.bennett@rice.edu

as well as signals from the environment. When exposed to a specific stimulus, cells express CRISPR-associated (Cas) nucleases and single guide RNAs (sgRNAs) that alter the sequence of a reporter gene. Upon specific binding of an sgRNA to its 20-base DNA target, the Cas nuclease edits the target sequence. Because recording plasmids are identical apart from a three-base coding mutation in the reporter gene, high-throughput sequencing is used to measure the plasmid ratio, which reflects the stimulus intensity. The first version (CAMERA 1) uses the *Streptococcus pyogenes* Cas9 (SpCas9) as the writing module. To maintain the accuracy of memory recording and readout, memory devices that record to the genome require sampling of large populations of cells for accurate information retrieval. To circumvent the need for large cell populations, Tang and Liu designed a high-copy number plasmid compensation system to store DNA modification states. By monitoring a large number of recorder plasmids within each cell, only 10 to 100 cells are needed for accurate recording in analog format. Using this strategy, the authors successfully recorded the intensity and order of presentation of multiple stimuli. The au-

thors further increased the versatility of the memory device by designing recorder plasmids with different antibiotic resistances. With this design, they could erase and rewrite data by simply adding antibiotics to reset the plasmid ratio. This rewritable memory device is multiplexable, durable, can implement Boolean logic in response to a large spectrum of inputs, and is particularly useful in cases where limited cell material is available—for example, sample collection to track cell lineages and map complex cell states in organismal development.

However, CRISPR nucleases make double-stranded DNA breaks (DSBs); owing to the mutagenic process of DSB repair by the nonhomologous end joining pathway, this approach can cause stochastic undesired mutations such as insertions and deletions (indels) and translocations. Thus, the authors developed a second version (CAMERA 2) that uses base editors rather than nucleases as the writing module. These base editors consist of a cytidine deaminase fused to a catalytically dead Cas9 to change C-G base pairs to T-A within target sequences without DNA cleavage (15). This feature lowers the frequency of stochastic miswriting events and increases recording accuracy.

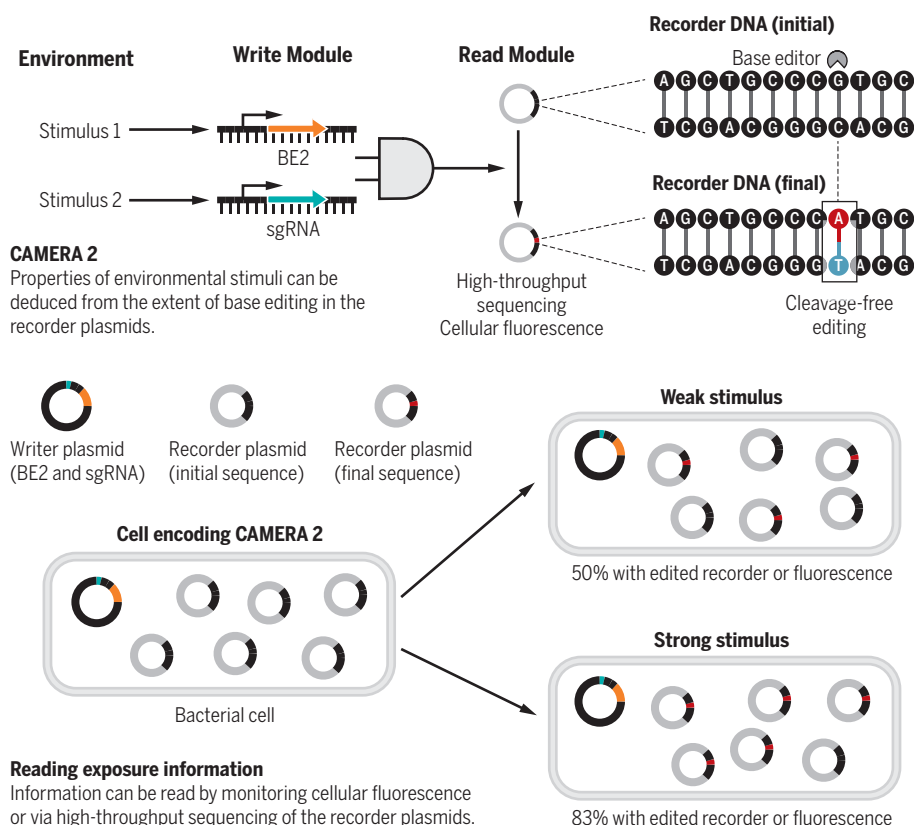
The editing events accumulate in a linear fashion, making base editors an excellent choice for recording in an analog format. Because of its continuous nature, analog recording provides a more accurate and nuanced record of environmental signals compared to digital recording. Additionally, the slow yet constant rate of DNA editing by the base editor means that CAMERA 2 can function as a molecular clock that records over hundreds of cell generations.

Because the order and timing of biological events determine cell fate, memory devices that can reveal the order of cellular events are of great utility. To this end, the authors designed two overlapping base editing targets such that the first editing event must occur before the second writing module can recognize the target. By layering two sgRNA circuits, the order of exposure to various inducers was recorded. Remarkably, reliable recording of exposure to a wide array of stimuli could be achieved with just 10 cells. Finally, the authors implemented CAMERA 2 in a human cell line and achieved robust multiplexed recording with minimal crosstalk between stacked sgRNAs.

As synthetic biologists build more tightly regulated induction systems, we can expect the development of increasingly sensitive and complex cellular memory devices. This work has clear clinical applications. Potentially, synthetic cell recorders could be developed as a probiotic that stably localizes to the gut, and a small clinical sample can provide a record of the environmental assaults experienced by an individual or the dosage of therapeutic drugs administered to a patient. Given the remarkable stability of DNA, these records can potentially persist for a million years (the theoretical upper limit for readability of DNA stored under ideal conditions). This powerful technology enables scientists to obtain real-time information regarding cell states during important processes including cell division, lineage differentiation, metabolic aging, tumorigenesis, and disease progression. ■

A DNA base editor to record cellular exposure to stimuli

CAMERA 2 uses base editor 2 (BE2) to record cellular exposure to specific stimuli, and properties of the stimuli can be deduced from the extent of base editing.



REFERENCES

1. T.S. Moon *et al.*, *Nature* **491**, 249 (2012).
2. J. Stricker *et al.*, *Nature* **456**, 516 (2008).
3. M.B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
4. A.E. Friedland *et al.*, *Science* **324**, 1199 (2009).
5. F. Farzadfar, T.K. Lu, *Science* **346**, 1256272 (2014).
6. P. Siuti *et al.*, *Nat. Biotechnol.* **31**, 448 (2013).
7. K.N.-M. Daefluer *et al.*, *Mol. Syst. Biol.* **13**, 923 (2017).
8. Y. Chen *et al.*, *Nat. Commun.* **9**, 64 (2018).
9. L. Qian, E. Winfree, *Science* **332**, 1196 (2011).
10. W. Tang, D.R. Liu, *Science* **360**, eaap8992 (2018).
11. T.S. Gardner *et al.*, *Nature* **403**, 339 (2000).
12. J.W. Kotula *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4838 (2014).
13. M. Mimee *et al.*, *Cell Syst.* **1**, 62 (2015).
14. S.L. Shipman *et al.*, *Nature* **547**, 345 (2017).
15. A.C. Komor *et al.*, *Nature* **533**, 420 (2016).

10.1126/science.aat3236

IMMUNOLOGY

Redemption for self-reactive antibodies

Antibody self-reactivity is repaired through antibody gene mutation in B cells

By **Ervin E. Kara**¹ and
Michel C. Nussenzweig^{1,2}

Immunity to pathogens and tolerance to self are cardinal features of immune systems. Immunological specificity is encoded by receptors expressed on the surface of lymphocytes that are generated through random assembly of variable, diversity, and joining (VDJ) gene segments during B and T lymphocyte development. In addition, B lymphocytes further diversify this initial repertoire through somatic hypermutation of antibody genes in germinal centers (transient structures that form in lymphoid organs in which high-affinity antibodies arise). On page 223 of this issue, Burnett *et al.* (1) devise a strategy to track the fate of self-reactive B cells in germinal centers elicited by a foreign antigen that structurally mimics a self-antigen (antigen mimicry). The authors provide evidence that hypermutation of antibody genes in germinal centers can repair self-reactive antibodies. These results have implications for how broadly neutralizing antibodies to HIV-1 may be formed.

Because antibody gene recombination events are random, more than 50% of immune receptors (antibodies) assembled during B cell development are self-reactive (2). Most developing B cells replace self-reactive receptors through persistent recombination or receptor editing (3). Any remaining self-reactive cells are either removed through cell death (4) or develop into short-lived inactivated (anergic) cells that express low levels of surface immunoglobulin M (IgM) antibodies (5). The result is that the vast majority of mature B cells (~95%) lack self- or poly-reactivity (binding to multiple antigens) (2).

Foreign antigenic challenge activates B cells to differentiate into germinal center

cells. B cells entering the germinal center join a dynamic environment that is divided into dark and light zones. The dark zone is the site of rapid rounds of cell division and antibody gene diversification from somatic hypermutation, which involves the mutator enzyme activation-induced cytidine deaminase (AID). Subsequently, dark zone B cells migrate to the light zone, where a small proportion are selected through interaction with antigen and cognate T cells to return to the dark zone for further rounds of clonal

at a relatively high frequency in the naïve compartment to compete with non-self-reactive cells for germinal center entry.

Burnett *et al.* show that despite these disadvantages, self-reactive B cells can enter and thrive in germinal centers. They demonstrate that in addition to diversifying antibody genes, somatic hypermutation in germinal centers can also repair self-reactive cells and make them non-self-reactive by means of a process the authors call receptor revision (see the figure). The observation that such cells

can survive in germinal centers despite their initial self-reactivity is consistent with the high threshold for selection against self- and poly-reactivity in the germinal center (9, 10).

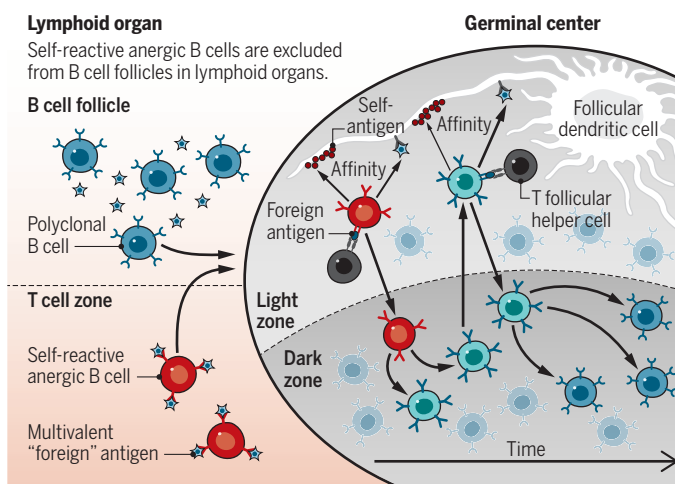
Burnett *et al.* carried out detailed antibody gene mutational analyses of single self-reactive B cells and found that clones carrying a mutation that simultaneously increased affinity to foreign antigens and reduced affinity to self were enriched in germinal centers. Additional compound mutations in antibody genes ultimately conferred a 5000-fold differential affinity toward foreign over self-antigens. This mutational pathway was specific to self-reactive cells. The authors detail how somatic mutation alters the structure of the self-reactive receptor to discriminate foreign from self and propose a model

in which self-reactive anergic B cells can be redeemed in germinal centers through the introduction of a randomly occurring but strongly selected “foundation” mutation that initially guides affinity away from self, followed by additional selection for mutations that increase affinity to foreign antigens.

The idea that self-reactive cells can contribute to immunity through germinal center redemption may be particularly important in responses to pathogens that cloak themselves in host antigens to avoid immunity. HIV-1 is one such pathogen. It covers its antigenic membrane proteins with host glycans that are believed to shield it from broadly effective antibody responses that neutralize most HIV-1 strains (host gly-

Revising self-reactive B cells

Stimulation with multivalent foreign antigen can recruit self-reactive anergic B cells into the germinal center. These cells then incur somatic mutations in antibody genes, and those that enhance affinity toward foreign antigen and reduce affinity to self-antigen are selected to propagate.



expansion and somatic hypermutation (6). High-affinity antibodies are products of repeated rounds of division, mutation, and selection. Overall, ~50% of germinal center cells die every 6 hours, a rate of cell death that maintains homeostasis in the context of rapid cell division (7).

The few self-reactive anergic cells that survive, which have the potential to elicit autoantibodies, are at a disadvantage for germinal center entry because they are short-lived and excluded from the B cell follicle (8). In addition, these cells are more difficult to activate than non-self-reactive B cells (5). Thus, germinal center entry by anergic cells requires receptor cross-linking by high-affinity multivalent antigen. Further, it is likely that anergic cells need to be present

¹Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA. ²Howard Hughes Medical Institute (HHMI), The Rockefeller University, New York, NY 10065, USA. Email: nussen@rockefeller.edu

cans are a self-antigen, and so antibodies to such antigens are selected against). When broadly neutralizing antibodies to HIV-1 do arise, they display a number of highly unusual features, including high levels of somatic mutation. These features allow the antibodies to interact with the self-antigen glycan shield and to reach past it to contact foreign peptidic determinants on the HIV-1 spike protein (11). Thus, many of these antibodies recognize a particular combination of self- and foreign antigens, and their development may involve redemption of self-reactive B cells through mutation, as described by Burnett *et al.*

Consistent with a requirement for some level of self-reactivity, broadly neutralizing antibodies to HIV-1 frequently demonstrate cross-reactivity to self-antigens (12). More generally, poly-reactivity was found in 75% of a large collection of human monoclonal antibodies to HIV (13). The notion that some of these antibodies arise from self-reactive precursors is supported by antibody gene knock-in experiments that have demonstrated that B cells that express predicted germline versions of broadly neutralizing antibodies to HIV-1 frequently show precursor cell deletion or absence of allelic exclusion that is indicative of self-reactivity (14, 15). However, when present at high precursor frequencies and challenged with high-affinity multivalent antigen, the knock-in cells can participate in immune responses (14, 15).

The findings of Burnett *et al.* further our understanding of the biology of B cell anergy and provide a framework for thinking about why such cells might be allowed to persist in immune systems despite their self-reactivity. ■

REFERENCES AND NOTES

1. D. L. Burnett *et al.*, *Science* **360**, 223 (2018).
2. H. Wardemann *et al.*, *Science* **301**, 1374 (2003).
3. D. Gay, T. Saunders, S. Camper, M. Weigert, *J. Exp. Med.* **177**, 999 (1993).
4. D. A. Nemazee, K. Burki, *Nature* **337**, 562 (1989).
5. C. C. Goodnow *et al.*, *Nature* **334**, 676 (1988).
6. G. D. Victora, M. C. Nussenzweig, *Annu. Rev. Immunol.* **30**, 429 (2012).
7. C. T. Mayer *et al.*, *Science* **358**, eaao2602 (2017).
8. J. G. Cyster, S. B. Hartley, C. C. Goodnow, *Nature* **371**, 389 (1994).
9. T. Tiller *et al.*, *Immunity* **26**, 205 (2007).
10. T. D. Chan *et al.*, *Immunity* **37**, 893 (2012).
11. A. Escolano, P. Dosenovic, M. C. Nussenzweig, *J. Exp. Med.* **214**, 3 (2017).
12. B. F. Haynes *et al.*, *Science* **308**, 1906 (2005).
13. H. Mouquet *et al.*, *Nature* **467**, 591 (2010).
14. R. Zhang *et al.*, *Sci. Transl. Med.* **8**, 336ra362 (2016).
15. R. K. Abbott *et al.*, *Immunity* **48**, 133 (2018).

ACKNOWLEDGMENTS

We thank members of the Nussenzweig laboratory for comments. E.E.K. is supported by a National Health and Medical Research Council C. J. Martin Fellowship; M.C.N. is a HHMI investigator.

10.1126/science.aat5758



GENETICS

Crowdsourced genealogies and genomes

Genealogical study provides insight into history and life span and heralds crowdsourced genetic research

By Alexandre A. Lussier¹ and Alon Keinan^{1,2}

Genealogies are likely the first, centuries-old “big data,” with their construction as old as human civilization. Recent renewed interest led to the largest genealogical websites (Ancestry.com, MyHeritage, and Geni) amassing 130 million users who generated billions of online genealogical profiles, offering ample research opportunities that would otherwise require extensive recruitment. On page 171 of this issue, Kaplanis *et al.* (1) showcase the research potential of this type of crowdsourced data, studying genealogies based on processing 86 million public Geni profiles.

An important research tool throughout human history, genealogical studies reach from anthropology to modern genetics and medicine. Of note are centuries of Icelandic genealogical enthusiasts combining family information with early-age scriptures. The founding of deCODE genetics sped up the process to create an online genealogy of 864,000 Icelanders, the Íslendingabók, by 2003. It has proven an incredible research tool, for example, in studying the relative roles of genetic heritability and shared environment in many complex diseases and other traits (2). Another unique genealogy has been constructed from historical records by the Mormon Church since 1921. Continued more

recently by hundreds of thousands of volunteers, they created about half a million new profiles daily, with many connected to health care records (3)—a true crowdsourcing effort.

Entering the age of genomics, genealogy enthusiasts greeted a new tool. Direct-to-consumer (DTC) genetic testing companies all provide a service for finding relatives, while obtaining powerful, crowdsourced genome-wide data for millions of individuals. AncestryDNA and 23andMe applied their data to study migrations, structure, and admixture of U.S. populations (4, 5). However, most crowdsourced genetic research is medically driven, focusing on the genetic basis of complex traits. For example, a recent study describing how genetic risk factors are shared across many traits included analyses of 23andMe customers for 17 of the traits (6).

Kaplanis *et al.* demonstrate the potential of large-scale genealogies, although with no genetic data, but at the hands of statistical and population geneticists. Extensively processing and validating genealogical data, they compiled 5.3 million genealogies, including one with 13 million individuals that often depicts at least 20 generations.

They analyze relatedness and distance at birth between married couples. Distance for most was less than 10 km before the Industrial Revolution (1750), followed by a gradual increase, which then accelerated to over 100 km after the start of the Second Industrial Revolution (1870). Average relatedness remained the same (equivalent to fourth cousins) prior to the Second Industrial Revolution, when it began decreasing in line with increasing distance. Kaplanis *et al.* postulate that recently decreased relatedness is due to

¹Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, USA. ²Cornell Center for Comparative and Population Genomics, Center for Vertebrate Genomics, and Center for Enervating Neuroimmune Disease, Cornell University, Ithaca, NY 14853, USA. Email: alon.keinan@cornell.edu

shifting cultural norms, rather than increased distance, because of the inconsistent relationship between relatedness and distance. This appears concurrent with popular writing from the time, which led 13 U.S. states to pass cousin marriage prohibitions by the 1880s (7) (although more distant relatedness is in question in Kaplanis *et al.*). A related study considered 160,000 couples in the Íslendingabók to show that increased couple relatedness (equivalent to third or fourth cousins) is associated with higher fertility that is not explained by socioeconomic influences on number of offspring, and hence is claimed to have a potential biological basis (8).

The main results of Kaplanis *et al.* involve life span. The resolution of the data set allows them to discern not only that average life span decreased during World War I and World War II, but also that the decrease was larger for individuals of military age. Despite these major events, life span appears to have increased at an almost constant rate of ~4 years per generation since ~1850. They conducted a meticulous study of factors affecting life span, attributing ~7% to gender, birth year, and geography combined. They estimated life span heritability at $16.1 \pm 0.4\%$, lower than most previous studies, although among them, the largest genealogy-based study until now provided a comparable estimate of $15 \pm 3\%$ in the Mormon genealogy (9). Kaplanis *et al.* estimate that an additional ~4% of life span is attributable to dominance (where having a single copy of a genetic variant constitutes the majority of the effect of having two) and none to interaction between different genetic variants.

Despite extensive analyses, Kaplanis *et al.* only scratch the surface of their resource, which is publicly available, stripped of personal information. It may be interesting to reanalyze life-span factors focused on very high longevity, and revisit other questions previously studied with smaller genealogies. The resource may benefit many disciplines, with unique promise in the combination with genetic data of the same individuals, an opportunity that led to large investments in DTC genetic services by the companies with the largest genealogical websites.

DTC genetic data are not publicly available, but Kaplanis *et al.* provide an academic version of their resource where individuals can consent to being identified. It can be used on websites to which participants upload their genetic data, as Kaplanis *et al.* implemented in DNA.Land, which, for example, collates family history of breast cancer and allows users to contribute their genomes to the National Breast Cancer Coalition (10). In a recent study, deCODE genetics highlighted yet again the power of large-scale genealogies with matched genetic data. They reconstructed

an ancestor's genome by mining descendants for inherited genetic fragments, which they tested via unique genealogical analyses (11).

One critical limitation of available crowdsourced data is that the "crowd" is mostly from 15% of the worldwide population that comprises Europe and North America. The overwhelming majority of DTC genetic testing customers are from these regions, as are 85% of the profiles in the Kaplanis *et al.* study. Partly due to local laws and consent, the potential unleashed by integrating worldwide diversity should provide an incentive to overcome these obstacles. Another shortcoming is the underutilization of the X chromosome by DTC genetic companies for both customer services and medical research (12). Its inclusion via newly developed analytical methods may improve these and, importantly, provide a key step toward closing the gender disparity in disease diagnosis and treatment (12).

The era of precision medicine heralds a greater potential for crowdsourcing, with distinct opportunities when familial, genetic, and medical data are integrated. Funding details of large-scale endeavors such as the U.S. National Institutes of Health *All of Us* program have put an effective price tag on the recruitment of each participant, their genetic data, and medical records. Recently founded companies, in turn, are attempting to resurrect the option for participants to lease their data to researchers. This may increase the potential for research based on crowdsourced, although not fully crowd-funded, data.

Beyond explosive growth of DTC genetic testing services (of ~16 million current customers, almost two-thirds joined since early 2017), whole-genome sequencing will likely become a cost-effective DTC choice within 2 to 3 years. This will enable tracing, and flagging of potentially harmful, *de novo* mutations in families and allow crowdsourced genetic research to more substantially advance disease risk prediction, diagnosis, and treatment.

Although many fields make use of crowdsourcing, none is better positioned, since all 7.5 billion of us have a genealogy, DNA, traits, and medical information to share. ■

REFERENCES

1. J. Kaplanis *et al.*, *Science* **360**, 171 (2018).
2. N. Zaitlen *et al.*, *PLoS Genet.* **9**, e1003520 (2013).
3. S. Knight *et al.*, *Hum. Hered.* **81**, 1 (2016).
4. E. Han *et al.*, *Nat. Commun.* **8**, 14238 (2017).
5. K. Bryc *et al.*, *Am. J. Hum. Genet.* **96**, 37 (2015).
6. J. K. Pickrell *et al.*, *Nat. Genet.* **48**, 709 (2016).
7. D. B. Paul, H. G. Spencer, *PLoS Biol.* **6**, e320 (2008).
8. A. Helgason *et al.*, *Science* **319**, 813 (2008).
9. R. A. Kerber *et al.*, *J. Gerontol. A Biol. Sci. Med. Sci.* **56**, B130 (2001).
10. J. Yuan *et al.*, *Nat. Genet.* **50**, 160 (2018).
11. A. Jagadeesan *et al.*, *Nat. Genet.* **50**, 199 (2018).
12. Editorial, *Nat. Med.* **23**, 1243 (2017).

10.1126/science.aat2634

NANOMATERIALS

A recipe for nanoporous graphene

Nanoporous graphene created from molecular precursors shows promise for electronic applications

By Alexander Sinitskii

Graphene is widely regarded as a promising material for electronic applications because the exceptionally high mobilities of its charge carriers enable extremely fast transistors (1). However, the lack of an energy band gap in graphene limits its use in logic applications; without a band gap, the devices remain highly conductive at any gate voltage and thus cannot be fully switched off. Researchers have therefore turned their attention to semiconducting forms of graphene that have the necessary band gap to enable transistors with high on-off ratios. On page 199 of this issue, Moreno *et al.* (2) report on the synthesis and device characterization of nanoporous graphene with semiconducting properties.

So far, researchers mostly have focused their attention on a different class of graphene-based materials that could have the desired energy band gap: one-dimensional graphene nanoribbons. Theoretical studies have shown that depending on their structural parameters, such as shape, width, and edge structure, graphene nanoribbons may possess not only a tunable electronic band gap (3), but also other intriguing physical properties, such as edge magnetism and highly localized electronic states. However, carving few-nanometer-wide strips is not the only nanostructuring approach that could create a band gap in graphene. The same can be achieved by patterning an array of closely spaced nanoscopic holes in graphene, thus forming a nanoporous graphene. Similar to nanoribbons, the properties of nanoporous graphene strongly depend on their structural parameters, such as the pore diameter and the periodicity of the structure (4).

Department of Chemistry and Nebraska Center for Materials and Nanoscience, University of Nebraska–Lincoln, Lincoln, NE 68588, USA. Email: sinitskii@unl.edu

Original attempts to fabricate graphene nanoribbons and nanoporous graphene for device studies relied on top-down approaches, in which the structures were directly etched from graphene (see the figure). Nanoribbons were fabricated by using electron-beam lithography (5), whereas nanoporous graphenes were patterned using self-assembled etch masks (6, 7). However, to open band gaps of ~ 1 eV (comparable to that in silicon, a conventional semiconductor material), the feature sizes in graphene nanoribbons and nanoporous graphenes should be less than 2 nm (3, 4). This is beyond the structural resolution of top-down approaches.

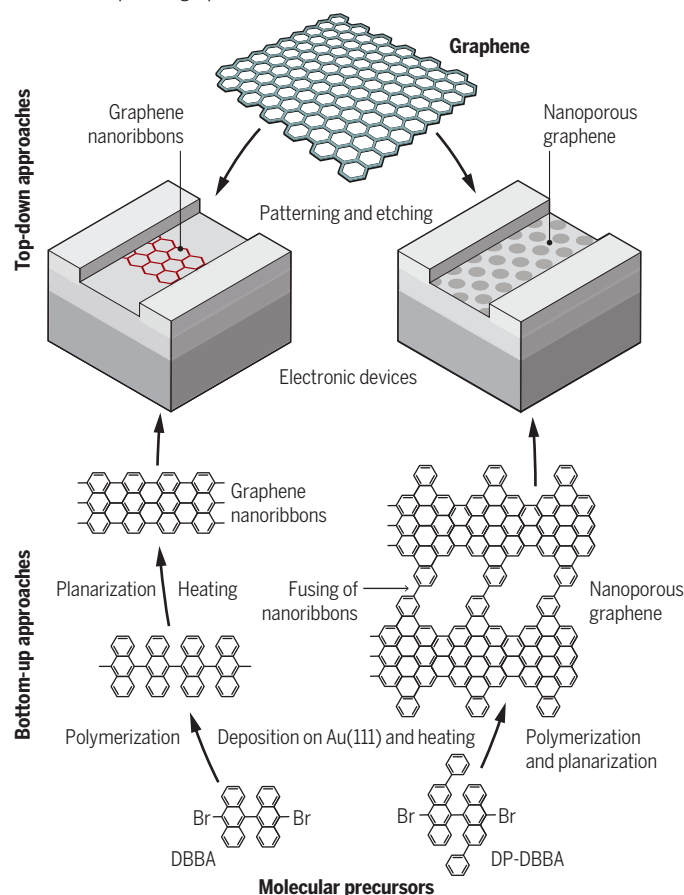
In 2010, Cai *et al.*, showed that very narrow graphene nanoribbons can be made with atomic precision by a bottom-up approach from smaller molecular building blocks, such as DBBA (10,10'-dibromo-9,9'-bianthracene) molecules (8) (see the figure). When DBBA is sublimed onto a single crystal Au(111) substrate under ultra-high vacuum (UHV) conditions at about 200°C, the molecules couple into linear polymer chains. Upon further annealing to about 400°C, the chains planarize, producing graphene nanoribbons. Cai *et al.* showed that other halogenated polycyclic aromatic hydrocarbon precursors can engage in similar on-surface reactions, producing a large variety of graphene nanoribbons with different structural parameters (8).

Moreno *et al.* now report that nanoporous graphenes can also be prepared by a bottom-up approach through on-surface coupling of specially designed halogenated molecular precursors on a single crystal Au(111) surface (see the figure). It was previously demonstrated that when straight graphene nanoribbons are produced on Au(111) at high coverage, they start fusing together to form wider but still straight ribbons (9). For example, the nanoribbon shown in the figure is $N = 7$ carbon atoms wide, and their parallel fusing can result in wider nanoribbons with $N = 14, 21$, etc. But what if the original nanoribbons are not straight?

Moreno *et al.* designed a new nanoribbon precursor that is closely related to the DBBA monomer (see the figure). When the resulting diphenyl-substituted DBBA (DP-

Two ways to make nanostructured graphene

Graphene nanoribbons and nanoporous graphene can be made by top-down or bottom-up approaches. Top-down approaches involve patterning and etching of graphene sheets, but the feature sizes of the resulting structures are too large for device applications that require a band gap. Bottom-up assembly from molecular precursors can overcome this limitation, as shown by Moreno *et al.* for nanoporous graphene.



DBBA) is sublimed on Au(111) substrate in UHV, similar to DBBA, the molecules first polymerize at about 200°C and then form graphene nanoribbons at about 400°C. However, because of the nonuniform width of the resulting ribbons, their fusion during the additional annealing at 450°C produces graphene nanostructures with nanoscopic holes (see the figure). Spectroscopic studies revealed that the resulting nanoporous graphene has a highly anisotropic electronic structure with a band gap of about 1 eV.

Some previous studies also used a bottom-up strategy to prepare nanoporous graphenes based on polyphenylene units through surface-assisted coupling of halogenated molecular building blocks (10, 11). Other authors have also fused chevron-shaped graphene nanoribbons to form nanoscale graphene pores (12). However, Moreno *et al.* go further by showing that their nanoporous graphenes can be transferred to a dielectric substrate for the fabrication of transistors with high on-off ratios. Device characterization of

graphene nanoribbons remains extremely challenging because of their relatively short length (typically < 50 nm), the necessity of accurate alignment relative to the device structure, and high contact resistances; in studies of electrical properties of nanoribbons the yield of working devices is often rather low (13). By contrast, Moreno *et al.*'s nanoporous graphenes form larger electrically conducting domains, from which devices for electrical property measurements could be produced at an impressive $\sim 75\%$ yield.

Moreno *et al.*'s study opens numerous avenues for research in different disciplines. It will likely stimulate chemists to develop new molecular precursors for nanoporous graphenes with various combinations of structural parameters (size, geometry and arrangement of pores). One possibility for improvement of the reported nanoporous graphene design is the absence of long-range order in the direction perpendicular to the nanoribbon chains. Future studies may yield such order in new kinds of nanoporous graphene with complete atomic precision. The complex, highly anisotropic structure of nanoporous graphenes may also be of great interest for spectroscopic studies. The author's demonstration of high-yield

fabrication of nanoporous graphene-based electronic devices should stimulate further nanoscale transport measurements. Finally, graphene nanostructures with nanoscopic pores may be of interest for applications such as separation, sensing, and potentially even DNA sequencing. ■

REFERENCES AND NOTES

1. Y. M. Lin *et al.*, *Science* **327**, 662 (2010).
2. C. Moreno *et al.*, *Science* **360**, 199 (2018).
3. L. Yang *et al.*, *Phys. Rev. Lett.* **99**, 186801 (2017).
4. T. G. Pedersen *et al.*, *Phys. Rev. Lett.* **100**, 136804 (2008).
5. M. Y. Han, B. Özyilmaz, Y. B. Zhang, P. Kim, *Phys. Rev. Lett.* **98**, 206805 (2007).
6. J. Bai *et al.*, *Nat. Nanotechnol.* **5**, 190 (2010).
7. A. Sinitskii, J. M. Tour, *J. Am. Chem. Soc.* **132**, 14730 (2010).
8. J. M. Cai *et al.*, *Nature* **466**, 470 (2010).
9. H. Huang *et al.*, *Sci. Rep.* **2**, 983 (2012).
10. M. Bieri *et al.*, *Chem. Commun.* **2009**, 6919 (2009).
11. R. Gutzler *et al.*, *Chem. Commun.* **2009**, 4456 (2009).
12. J. D. Teeter *et al.*, *Nanoscale* **9**, 18835 (2017).
13. J. P. Llinas *et al.*, *Nat. Commun.* **8**, 633 (2017).

ACKNOWLEDGMENTS

The work was supported by the National Science Foundation through CHE-1455330.

10.1126/science.aat5117

RETROSPECTIVE

Stephen Hawking (1942–2018)

The world's best-known scientist richly deserved his fame

By John Preskill

Stephen William Hawking died on 14 March (Albert Einstein's birthday) at the age of 76 after decades of battling the incurable disease amyotrophic lateral sclerosis (ALS). His early scientific work transformed our understanding of general relativity, Einstein's theory of gravitation. Later in life, Stephen became an immensely successful popularizer of science; his courage and high spirits in the face of his disability inspired millions. Stephen Hawking's achievements as a scientist, communicator, and public figure were commensurate with his great fame.

Stephen was born in Oxford on 8 January 1942 (which, as he enjoyed pointing out, was the 300th anniversary of Galileo's death) and entered the University of Oxford in 1959. Although his mathematical aptitude was quickly recognized, Stephen was not a diligent student, and his performance was lackluster. Nonetheless, he graduated in 1962 with a bachelor's degree in natural sciences. Soon after beginning doctoral studies at the University of Cambridge, where he joined the research group of physicist Dennis Sciama, Stephen was diagnosed with ALS. Although terminal, the disease progressed more slowly than anticipated, and with Sciama's encouragement, Stephen focused on research with newfound determination, completing his Ph.D. in 1965. He remained at Cambridge for the rest of his career, confounding expectations by surviving for 55 years after his initial diagnosis. From 1979 until 2009, Stephen was the Lucasian Professor of Mathematics at Cambridge, following in the footsteps of Isaac Newton, Paul Dirac, and other scientific luminaries. He remained scientifically active until his death.

Stephen's scientific career divides naturally into two phases, which could be called his classical gravity phase and his quantum gravity phase. During his classical phase, one of his early achievements was proving that time had a beginning—that the laws of physics as we now understand them must have broken

down very early in the history of the universe, at the Big Bang. Stephen also greatly advanced our understanding of black holes, where gravitational forces are so strong that time comes to an end; in particular, he discovered and elucidated a deep analogy between classical black holes and thermodynamics.

Stephen's pivot from classical to quantum gravity was precipitated by his greatest scientific achievement, which shook the world of physics in 1974. According to Einstein's theory, nothing, including light, can escape from inside a black hole, which explains why it's black. But Stephen found that black holes



are not really completely black. Instead, due to the subtle consequences of quantum physics, they emit what we now call Hawking radiation. He computed the temperature of a radiating black hole, and derived a beautiful formula for its entropy, validating and refining an earlier conjecture by theoretical physicist Jacob Bekenstein (which Stephen had hoped to refute). A major milestone in the history of science, the theory of Hawking radiation established a profound connection among gravitation, quantum physics, and information science, which still guides the ongoing search for a more complete theory of quantum gravity. Stephen's subsequent research focused on that quest, emphasizing the role of quantum physics in the origin and early history of the universe.

Although his scientific accomplishments alone would suffice to ensure an enduring

legacy, Stephen Hawking also became one of the world's most successful science communicators. Stephen firmly believed that the quest for a complete theory of the universe should be accessible to everyone, at least in broad principle, not just to a few specialists. That conviction drove him to write *A Brief History of Time*. Whereas other scientists have tried to write books for lay readers, Stephen earned tenure on the *New York Times* best seller list, thanks in part to that ingenious title. The book sold more than 10 million copies and was translated into dozens of languages. Its extraordinary success led to more books, including a series for children, which Stephen coauthored with his daughter Lucy.

Stephen achieved scientific greatness despite a severe physical disability, while displaying a zest for life and buoyant sense of humor that seemed miraculous under the circumstances. People rooted for Stephen, and he appreciated having millions of fans.

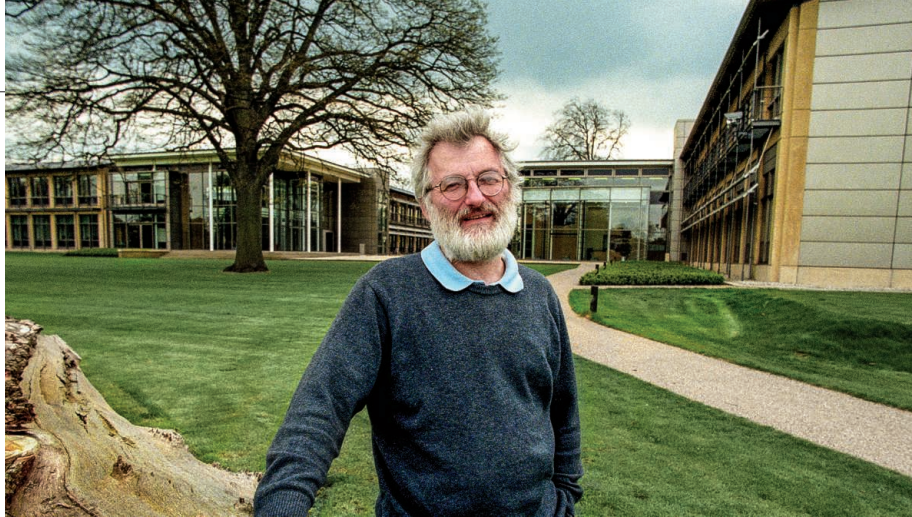
I first got to know Stephen at a 1982 workshop in Cambridge, but we became closer after he began making regular visits to the California Institute of Technology in 1991. Stephen was fun to be with; we could always make each other laugh, and he enjoyed being treated irreverently. In the middle of a scientific discussion, I could interject, "And what makes you so sure of that, Mr. Know-It-All?" knowing that Stephen would respond with his eyes twinkling: "Wanna bet?"

With our friend Kip Thorne, we made some of those bets "official," and we were all taken aback by how much attention they received. Stephen conceded our most famous bet (regarding whether black holes destroy information) in 2004, before an

audience in Dublin of 700 scientists and at least 50 reporters from print and electronic media. To pay his debt, he presented me with *Total Baseball: The Ultimate Baseball Encyclopedia*. You can't buy one of those in Ireland, so Stephen's assistant had arranged to have it shipped overnight. Not knowing what else to do, I held the book over my head as though I had just won the Wimbledon final, while what seemed like a million flashbulbs popped to record the moment.

We made bets for fun, but physicists passionately care about the scientific issues in question, founded on some of Stephen's most far-reaching contributions. Combining extraordinary depth of thought with an irrepressible sense of play—that's what I'll remember best about Stephen Hawking. ■

Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA.
Email: preskill@caltech.edu



RETROSPECTIVE

John Sulston (1942–2018)

A visionary biologist with a deep social conscience

By Judith Kimble^{1,2}

Sir John Sulston, a pivotal figure in the Human Genome Project, died on 6 March 2018. He was 75. His extraordinary ability to tackle and solve biological problems of immense scale and vision, coupled with his lifelong commitment to ethics, shaped the *Caenorhabditis elegans* nematode and human genome communities. Sulston shared the 2002 Nobel Prize in Physiology or Medicine for discoveries in organ development and programmed cell death. In addition to his earlier work, Sulston will be remembered for leading the British effort to sequence the human genome and defending free access to the data.

Born in Buckinghamshire in 1942, Sulston described his young self as a mechanically minded artisan who preferred science to sport. As an adult, he combined those artisan's gifts of design and creation with vision and hard work. He received his B.A. in natural sciences in 1963 from Pembroke College, Cambridge, UK, and his Ph.D. in Chemistry in 1966 from the University of Cambridge. After a brief postdoctoral fellowship at the Salk Institute in California, he returned to Cambridge and took a position at the Medical Research Council Laboratory of Molecular Biology (MRC LMB). In 1992, Sulston became director of the Sanger Centre. After stepping down in 2000, he continued to devote himself to pressing societal issues.

John Sulston first touched my life through a 1975 letter to my graduate adviser detailing his unpublished method to glean cell-lineage

data from living nematodes. John had discovered that, by looking through a microscope, he could see not only each cell division but also the fate of each daughter cell, including its movement and differentiation. We were working on an organ peripheral to John's interests, and he suggested we use his method to determine its cell lineage. That generosity decided my Ph.D. project. I met John in 1977 at a *C. elegans* workshop. Already a major player because of his pioneering lineage work, John opted to present a poster. Astonishingly, the "poster" was just a 35-mm slide of the postembryonic lineage taped to a window! He wanted viewers to discuss concepts rather than data details, a decision some considered quirky but others like me found refreshingly focused on the larger picture.

John's intense commitment to pushing the limits of scientific frontiers, along with his approachability and easy-going nature, convinced me to do postdoctoral research with him at the MRC LMB. His 1976 lineage publication had reported reproducible cell deaths, paving the way for *C. elegans* studies to dissect the regulation of programmed cell death. Soon after I arrived at LMB in 1978, John sequestered himself to decipher the embryonic lineage. He sat in a darkened room each day for about 12 hours, time for the cells of an early embryo to transform themselves into a wriggling worm. Normally charismatic and social, John tackled each day of solitude with renewed drive to track each division and daughter cell as it assumed its role in the developing embryo. After more than a year, John finished his work, connecting the embryonic and postembryonic lineages to generate the first complete developmental map of a metazoan. This feat laid the foundation for the now-burgeoning *C. elegans* field,

which has revealed secrets applicable to all animals, including human health and cancer.

With his first megaproject complete, John settled on his next visionary idea: generating a physical map of the *C. elegans* genome. The scope of this effort was immense, but John understood the genome's significance as a path to molecular understanding. His first step, in 1982, established methods to break the genome into bits and assemble a map from its pieces. Ever generous, he made map fragments publicly available soon after assembly and long before publication, catapulting a host of molecular studies.

With the advent of DNA sequencing, John's vision broadened to the Human Genome Project. As director of the Sanger Centre, he led a large and talented team to improve methods, produce enormous quantities of data, and computationally analyze sequences of the worm and then the human genome. Throughout this time, John worked at the bench, devising new methods to sequence seemingly impossible parts of the *C. elegans* genome, which was published in 1998 as the first complete metazoan genome sequence.

John's greatest challenge came when Celera Genomics set out to sequence the human genome and patent its contents. The idea that a private company might sequence the human genome for profit and prevent free access to the scientific community was anathema to John. Given the enormous implications of the human genome for human health, he considered free access nonnegotiable. For the first time in his life, John became the center of controversy, but his heroic efforts kept the human, mouse, and now many other genomes in the public domain, as described in his book, *The Common Thread*.

In 2001, after stepping down as Sanger director, John reluctantly accepted an invitation to be knighted after being convinced that the recognition benefited science. He next was awarded a Nobel Prize in 2002. He then threw himself into his work as chair of the Institute for Science, Ethics, and Innovation at the University of Manchester and chair of a Royal Society task force to assess the effects of increasing human population on human health and the environment.

John and his wife, Daphne—inseparable for more than 50 years—raised two children, Ingrid and Adrian. As his first postdoc, I was included in family cycling outings and visits to his home, which was cluttered yet comfortable in classical English style. Each November, John invited the lab to celebrate Guy Fawkes Night, with glowing lanterns hung around their garden and a roaring bonfire. John Sulston and his wonderfully generous and humble spirit will be sorely missed. ■

¹University of Wisconsin–Madison, Department of Biochemistry, Madison, WI, USA. ²Howard Hughes Medical Institute, Madison, WI, USA. Email: jekimble@wisc.edu



A lab worker exposes his arm to *Aedes aegypti* mosquitoes, which spread the Zika virus.

POLICY FORUM

RESEARCH ETHICS

Bystander risk, social value, and ethics of human research

Contentious risks demand a new approach

By S. K. Shah,¹ J. Kimmelman,² A. D. Lyerly,³ H. F. Lynch,⁴ F. G. Miller,⁵ R. Palacios,⁶ C. A. Pardo,⁷ C. Zorrilla⁸

Two critical, recurring questions can arise in many areas of research with human subjects but are poorly addressed in much existing research regulation and ethics oversight: How should research risks to “bystanders” be addressed? And how should research be evaluated when risks are substantial but not offset by direct benefit to participants, and the benefit to society (“social value”) is context-dependent? We encountered these issues while serving on a multidisciplinary, independent expert panel charged with addressing whether human challenge trials (HCTs) in which healthy volunteers would be deliberately infected with Zika virus could be ethically justified (1). Based on our experience on that panel, which concluded that there was insufficient value to justify a Zika HCT at the time of our report, we propose a new review mechanism to preemptively address issues of bystander risk and contingent social value.

BYSTANDER RISKS

A Zika HCT would pose risks to bystanders not enrolled in the study because subjects could transmit the virus through sexual activity, pregnancy, mosquito vectors, or other unknown ways (2). Zika HCTs are not the only type of research involving bystander risk. Dual-use research, such as research on avian influenza that could be used to make biological weapons, could harm bystanders. HIV cure trials that withdraw antiretrovirals to study new treatments can place subjects’ sexual partners at risk of HIV infection (3). Experimental mitochondrial replacement techniques can affect future children (4).

There are some, albeit incomplete, protections for certain types of bystanders. In dual-use research, bystander risk is addressed by institutional biosafety committees and, in the United States, by the National Safety Advisory Board for Biosecurity (NSABB). Many countries have regulations on research involving pregnancy and reproduction. For research that can harm identifiable communities by generating data that could cause stigma or contradict cultural beliefs, community consultation can help address group harm, although this may be difficult to implement for fragmented communities (5).

However, there is no clear or systematic mechanism for protecting bystanders from research risks arising through sexual or environmental transmission. These types of bystanders are generally not attended to in research regulations, international ethics

guidance, and institutional review board (IRB) deliberations because they are not “human subjects”—researchers do not directly intervene or interact with them, or collect their identifiable data (6). IRBs are instructed only to consider risks to subjects, and because “possible long-range effects” should not be considered, it is unclear whether existing regulations permit IRBs to address bystander risk (7). A U.S. Institute of Medicine report merely instructs investigators that they may have ethical obligations to anticipate and plan to address bystander risks that are “foreseeable and significant” (8).

It is important to protect all research bystanders because they may be unable to protect themselves; obtaining their consent might be impossible in some cases and problematic in others. For example, Zika HCT participants might choose sexual partners spontaneously or anonymously, precluding advance consent from all at-risk bystanders. Some studies might expose large numbers of bystanders to risk of environmental transmission, rendering consent infeasible. One approach might be to inform participants about risks to bystanders in consent forms. Yet researchers cannot simply pass their responsibilities on to subjects who lack accountability to inform and protect others.

Unfortunately, researchers or IRBs attempting to address bystander risks will find no consensus on or framework for determining when bystander risks are ethically and legally justifiable (6). For instance, an individual could be infected with Zika virus from an HCT participant, and then become pregnant. This risk could be mitigated but not eliminated. How high of a chance of bystander harm should be tolerated? How should direct conflicts between the interests of participants and bystanders be resolved?

IRBs are not well-suited to consider these issues. IRBs are explicitly charged with protecting participants’ welfare and liberty, but lack a mandate to protect bystanders. They may also be reluctant to address such unsettled ethical issues (9). Public trust is critical to biomedical research, and the public often reacts strongly when researchers or institutions impose risks on individuals—even if risks of similar magnitude are already naturally present (10). Taking steps to protect bystanders may help avoid public outcry in the event that bystanders are harmed.

SUBSTANTIAL RISK, CONTINGENT VALUE

In addition to bystander risk, Zika HCTs pose risk of potentially long-term harms to participants (11), and substantial uncertainty about the level of risk. That HCT researchers

¹University of Washington and Seattle Children’s Research Institute, Seattle, WA, USA. ²McGill University, Montreal, Canada. ³University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ⁴Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁵Weill Cornell Medical College, New York, NY USA. ⁶Instituto Butantan, São Paulo, Brazil. ⁷Johns Hopkins University, Baltimore, MD, USA. ⁸University of Puerto Rico School of Medicine, San Juan, Puerto Rico, USA. Email: shahsk@uw.edu

deliberately induce pathology in otherwise healthy people may be hard for the public to understand, even if ethically justifiable (12). When direct benefits cannot justify the risks participants face, the risks must be weighed against potential social value. Zika HCTs therefore needed an especially high social value to justify the risks, but measuring this value was challenging because it hinged on several factors external to the research.

The panel deliberated about whether results from Zika HCTs were likely to meaningfully contribute to Zika vaccine research. This required considering evidence about active or planned research trajectories for candidate vaccines from regulators and researchers. HCTs are not part of the standard requirements for regulatory approval to license and market new vaccines, so the extent to which regulators would be willing to rely on them was unclear. We assessed whether the results from proposed research would alter the research trajectory in some critical way and whether the epidemiology would continue to make field testing possible and necessary.

Addressing highly contingent social value is a persistent ethical challenge, with some clear (if not exhaustive) examples. When the Ebola epidemic was ongoing and different studies sought to test many interventions in the same population, the value of particular trials was relative to the alternatives and depended on outcomes of others in a way that was difficult to predict and stage (13). Bioterrorism countermeasure studies, such as anthrax vaccine studies, need to be relevant to current threats in order to have social value, but the probability of an attack is hard to assess. Some have argued that trials with highly uncertain risk, innovative designs, and broader policy implications (e.g., repeated transplantation of neural stem cells) require a new approach to ethics review (14).

Our current system is ill-equipped to handle this challenge. Many studies may not require exacting judgments about social value, but if the risks are contentious, precision becomes more important. IRBs review one protocol at a time and lack both the capacity and mandate to gather evidence to contextualize each individual protocol. Decisions made by individual IRBs may also be inconsistent with each other, causing unfairness and confusion. The type of programmatic assessment done by the Zika HCT ethics panel, in contrast, required forecasting the trajectory of the Zika vaccine research program and evolution of the epidemic. Such information may be necessary to determine whether a study contributes to a broader research program, and thus has adequate social value, rather than duplicating effort or generating results that will not be integrated into a path toward development or application.

A MORE COMPREHENSIVE APPROACH

The type of review provided by the Zika HCT ethics panel was an important response to an evolving research ethics landscape where the status quo was insufficient. IRBs tasked primarily with protecting individual subjects in individual protocols are not the right bodies to address programmatic issues we identified.

We therefore propose that agencies funding biomedical research establish databases of reviewers qualified to serve on ad hoc Comprehensive Ethics Review Committees (CERCs). CERCs should involve ethicists, policy experts, clinicians, patient representatives, and scientists, supplementing with subject matter expertise and/or community input as needed. They should conduct proactive review of research programs (e.g., all proposed projects involving a new study design in a particular disease area) and have at least two clear triggers for activation: (i) substantial risk to bystanders who cannot protect themselves, and (ii) research with contentious risk and highly contingent social value. CERCs should be activated by members of funding agencies and their reviewers, and have jurisdiction primarily over projects submitted to a particular agency, but also made available to IRBs or researchers.

To illustrate how CERCs could fill important gaps in the current research oversight system, consider how the Zika HCT ethics panel engaged with the bioethics literature, ensured its members had expertise in relevant ethical and scientific areas, consulted with other experts, and gathered additional evidence. This went far beyond what a typical IRB could reasonably accomplish, and would be duplicative and taxing if marshalled at every institution facing similar questions.

We assessed the acceptability of bystander risks by comparing their likelihood and magnitude to data on adverse events from similar trials that are generally viewed as ethically acceptable (e.g., phase I drug trials with healthy volunteers, malaria HCTs). We noted that key uncertainties about Zika virus transmission might resolve, and protections to avoid the most serious consequences for bystanders could then be developed. We therefore proposed high-priority research questions identifying the duration of infectivity and modes of transmission associated with Zika virus to be addressed in advance of a Zika HCT.

To assess highly contingent social value, the panel was empowered to gather evidence about active or planned research trajectories for various candidate vaccines from regulators and researchers, and were not confident that results from proposed Zika HCTs were likely to be harnessed in a useful way. To underscore the contingent nature of this judgment, circumstances have changed in the past year. Field trials of Zika vaccines have

become difficult as ongoing outbreaks are unpredictable or short-lived, yet the threat of future (and larger) outbreaks has not gone away (15). Zika HCTs may become the only way to prioritize vaccine candidates, giving them clear and considerable social value.

Some may object that generalizing and institutionalizing this approach could slow valuable research by adding an additional layer for review. However, embedding this process within funding agencies could preempt ethical problems that might otherwise stymie research. Concerns that CERCs might suffer from “mission creep” could be countered by establishing clear charters and triggers for deploying CERCs. Unlike IRBs, their opinions should be publicly available to provide precedent for future research programs or for IRBs evaluating particular protocols at a later date. CERCs should work with existing review bodies, such as by referring dual-use issues to the NSABB or sending protocols not requiring heightened scrutiny back to IRBs for standard research review. Using CERCs to proactively address these critical ethical issues may signal a commitment to ethical research to policy-makers and the public, which could redound to the benefit of researchers through future funding and support. ■

REFERENCES AND NOTES

1. S.K. Shah *et al.*, Ethical considerations for Zika virus human challenge trials: Report & Recommendations (2017); www.niaid.nih.gov/sites/default/files/EthicsZikaHumanChallengeStudiesReport2017.pdf.
2. C.J. Gregory *et al.*, *J. Infect. Dis.* **216** (suppl. 10), S875 (2017).
3. D. Lawrence *et al.*, *J. Virus Erad.* **3**, 69 (2017).
4. National Academies of Sciences, Engineering, and Medicine, *Mitochondrial Replacement Techniques: Ethical, Social, and Policy Considerations* (National Academies Press, Washington, DC, 2016).
5. M.M. Mello, L.E. Wolfe, *N. Engl. J. Med.* **363**, 204 (2010).
6. J. Kimmelman, *Camb. Q. Healthc. Ethics* **16**, 483 (2007).
7. Department of Health and Human Services, Protection of Human Subjects, 45 C.F.R. §46.111 (2009).
8. Institute of Medicine, *Ethical Considerations for Research on Housing-Related Health Hazards Involving Children* (National Academy Press, Washington, DC, 2005), pp. 132.
9. L. Abbott, C. Grady, *J. Empir. Res. Hum. Res. Ethics* **6**, 3 (2011).
10. C. Hooker, A. Capon, J. Leask, *Public Health Res. Pract.* **27**, e2711709 (2017).
11. I.R.F. da Silva, *et al.*, *JAMA Neurol.* **74**, 1190 (2017).
12. B. Bamberg *et al.*, *Public Health Ethics* **9**, 92 (2016).
13. A.H. Beavogui, A. Delamou, M.L. Yansane, *Clin. Trials* **13**, 73 (2016).
14. B. Lo, D. Grady, *JAMA* **302**, 2697 (2009).
15. C. Etienne, T.D. Santos, M.A. Espinal, *Am. J. Trop. Med. Hyg.* **97**, 16 (2017).

ACKNOWLEDGMENTS

This manuscript does not necessarily reflect the positions or policies of the U.S. Department of Health and Human Services, the National Institutes of Health, the National Institute of Allergy and Infectious Diseases, and/or the Department of Defense Walter Reed Army Institute of Research. The authors thank B. Eisinger, H. Marston, C. Paules, C. Dieffenbach, Z. Emanuel, N. Michael, C. Grady, D. Opel, N. Garrison, E. Weiss, J. Yu, J. Krick, E. Berkman, J. Jeavons, S. Kraft, L. Dawson, and an anonymous reviewer for helpful comments. The authors particularly thank F. McCutchan, who also served on the Zika HCT ethics panel, for valuable insights.

10.1126/science.aag0917



BOOKS *et al.*

EVOLUTIONARY BIOLOGY

Adapting to life in the big city

To thrive in rapidly changing urban areas, plants and animals are evolving at astonishing rates

By Arne Mooers

Metal-excreting pigeons, pigeon-eating catfish, cigarette-wielding sparrows, soprano-voiced great tits: The modern city is a fantastical menagerie of the odd and unexpected. Through a series of 20 short but connected chapters that mix natural history vignettes, interviews with visionary scientists, and visits to childhood haunts, science journalist and biology professor Menno Schilthuizen introduces readers to the striking facts of ongoing urban evolution in *Darwin Comes to Town*. But while the prose may be playful (“Cut to the Hollywood bobcats”), the underlying message may cause discomfort.

Two cross-cutting ideas permeate the book. The first is the notion of rampaging sameness. Because we are incessant, but messy, busybodies, Schilthuizen argues, we scatter species across countries and continents. And we move most among cities.

The author does a fine job of conveying this urban sameness when describing a scene along an estuary in Singapore: the house crows and the mynas feeding in the cow grass, the apple snails laying eggs among the mimosa, the red-eared slider turtles dipping into the water, and the peacock bass breaking the surface for a gulp of air. Every one of the species he describes is a non-native, every one is found in countless other cities the world over, and every one is at home in its new habitat. Schilthuizen has even borrowed a name from parasitology for them: anthropophiles.

And the reason for this biological sameness is urban sameness. Cities around the world produce the same sorts of garbage and the same sorts of noise, house the same sorts of skyscrapers, and produce the same fragmented landscapes. They can even generate the same sort of weather via particulate pollution and the heat-island effect.

The book's second major theme is that rapid change is an enduring part of the urban environment. Urban plants and animals evolve and adapt to their novel surroundings at remarkable speed. The city pigeons' darker, more melanic feathers,



Darwin Comes to Town
How the Urban Jungle Drives Evolution
Menno Schilthuizen
Picador, 2018. 304 pp.

The reviewer is at the Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. Email: amooers@sfu.ca

Dark pigments in a pigeon's feathers may help it sequester toxic metals in polluted cities.

for example, sequester poisonous metals; the great tit's new soprano notes are better heard above the city din; and city moths in Europe have become less attracted to deadly artificial lights.

Indeed, the realization that adaptive evolutionary change occurring on human time scales in multicellular species is common, rather than rare, is both fairly new and fairly profound. The ubiquity of the phenomenon has even given rise to a new field known as eco-evolutionary dynamics (1).

It is now clear that adaptation can be so fast as to affect the very environment that sets the stage for those adaptations, leading to possible merry-go-rounds of organism-environment-organism changes through time. The implications of this are still not fully known, but it's safe to assume that this is not what Darwin envisioned from his seat in the Kent countryside. (Perhaps he should have come up to the city more often.)

The fact that urban evolution is surprisingly fast also supports one of the radical ideas promulgated at the very end of this book: a vision of a city engineered to encourage the continued adaptive evolution of other species. We could, Schilthuizen argues, become evolution engineers, promoting the evolution of traits that will stand both us and our anthropophiles in good stead, such as using thriving non-native plants to populate green roofs or actively suppressing genetic mixing. As lineages continue to evolve, we can reengineer their environments as needed.

A project of engineering urban evolution is likely to pique interest, discussion, and perhaps even directed research. But this is clearly a view of nature that is more Abu Dhabi than Amazon rainforest, more engineering than awe.

Schilthuizen tacitly acknowledges that we have bent Earth to our bidding and that the rest of its inhabitants will either adapt or perish. This is undoubtedly true (2). But his view of the future replaces a nature where we decidedly do not meddle (3) with one where we decidedly do. For some, this will seem the very opposite of natural. And that dissonance may put a sting in an otherwise fascinating tale. ■

REFERENCES

1. A. P. Hendry, *Eco-evolutionary Dynamics* (Princeton Univ. Press, Princeton, 2016).
2. C. D. Thomas, *Inheritors of the Earth* (Public Affairs, New York, 2017).
3. D. S. Maier, *What's So Good About Biodiversity?* (Springer Verlag, Berlin, 2012).

10.1126/science.aas9773

MARINE SCIENCE

The future of artisanal fishing

Declining fish populations and policies that favor large operations threaten small fisheries

By Daniel Pauly

Written by Kevin Bailey, a marine fisheries biologist, *Fishing Lessons* is a small book with an urgent plea for readers, seafood consumers, and society in general to pay more attention to the challenges faced by artisanal fisheries, where much of the fish used for human consumption is caught. Although definitions vary among countries, artisanal fishers usually rely on small, owner-operated boats of less than 12 meters and deploy in coastal waters a variety of gears to catch fish for local sale rather than for their own consumption.

Bailey faces a daunting task from the start, as the premise that artisanal fisheries contribute substantially to food security, particularly in developing countries, is difficult to prove with the most commonly used data set. The global fisheries catch statistics that the Food and Agriculture Organization (FAO) compiles, harmonizes, and disseminates do not distinguish between fisheries sectors, meaning that data from large-scale commercial operations can obscure the role played by smaller operations.

Yet artisanal and subsistence fisheries generate about one-third to one-half of the total global catch that is used for direct human consumption, as assessed by catch reconstructions for all maritime countries of the world (1). (Industrial fisheries discard 10% of their catch and send another 30% to be processed as animal feed.)

Using case studies from Italy, Chile, eastern Canada, the west coast of the United States, and the Brazilian Amazon, Bailey introduces us to the challenges faced by artisanal fisheries. These include dwindling fish populations as large-scale industrial vessels move in and take “all the fish” and a lack of governance systems that would provide small fisheries a measure of control over the coastal resources that are accessible to them.

In the Amazon, where Bailey examines the harvesting practices surrounding the

“arapaima”—a giant air-breathing fish—the threat to artisanal fisheries is not industrial fishing but industrial aquaculture. Such ventures are able to deliver preplanned and preordered quantities to international markets, thus marginalizing the value of capture fisheries.

Contrary to the author’s suggestion that arapaima are “easy prey for hunters,” they are, in fact, extremely difficult to harpoon. To do so effectively requires years of training. As such, the potential loss of traditional fisheries represents a cultural loss as well.

However, the biggest challenges for artisanal fisheries—at least in the highly developed areas such as the United States, Canada, and Europe—are explicit government policies that seek to privatize access to what, until recently, were public goods: the fish resources of the coastal seas. These policies are structured such that governments

distribute quotas—called catch shares—to fisheries based on their previous catch history, which means that owners of large fleets get the lion’s share. Such inequality, which is aggravated by the fact that the shares are tradable, gives large-scale fisheries access to a fixed fraction of the total allowable catch (TAC), as determined by a science-based management agency, in perpetuity.

Bailey correctly points out that such policies invariably lead to the quotas or shares

being concentrated in the hands of a few industrial fleet operators, with formerly independent owner-operators (mostly artisanal in nature) having few options but to work as hired crew on the quota owners’ vessels. The seeming intractability of this problem explains the melancholic tone of the book, much of which is dedicated to describing the vanishing of coastal fishing cultures.

However, Bailey also describes some positive developments, including the emergence of community-based artisanal fisheries, where members pay in advance for certain quantities of fresh fish. Such ventures enable artisanal fishers to purchase supplies for a fishing season and to later deliver high-quality fresh fish at lower prices than through conventional outlets. This provides livelihood security for the artisanal fishers and assurance of good-quality fresh seafood supply for the communities involved.

There are several small errors throughout the book (e.g., the main shellfish caught in Chile—the “loco”—is sold in international markets, not local ones), which would have merited better editing. However, overall, *Fishing Lessons* makes a good case for abandoning current practices and policies that marginalize artisanal fisheries and disrupt fishing jobs and communities all over the world. ■

REFERENCE

1. D. Pauly, D. Zeller, *Nat. Commun.* 7, 10244 (2016).

10.1126/science.aas9431



Fishing Lessons
Artisanal Fisheries
and the Future
of Our Oceans

Kevin M. Bailey
University of Chicago
Press, 2018. 252 pp.



Fishermen in Nungwi, Tanzania, prepare to set out to sea.

The reviewer is at the Institute for the Oceans and Fisheries, The University of British Columbia, 2202 Main Mall, Vancouver, BC V6T 1Z4, Canada. Email: d.pauly@oceans.ubc.ca

LETTERS

Edited by **Jennifer Sills**

Editor's note

In her *Working Life* piece “Instagram won’t solve inequality” (16 March, p. 1294), Meghan Wright examined why she feels conflicted reading #scicomm Instagram posts by fellow women scientists. She explained that she recognizes the good they can do, yet it seems unfair that such scientists must devote time to social media outreach to combat systemic inequities. So, she has decided that she prefers to separate her social media use from her scientific activities. Wright named a social media role model at her university—the Science Sam Instagram account run by Samantha Yammine—before detailing why she did not want to participate in this kind of outreach. Although she intended to use Science Sam as an example of social media success, Wright’s critical comments about such outreach were interpreted by some as a sexist and mean-spirited personal attack on Samantha Yammine in particular and women science communicators in general. In this section, Samantha Yammine and colleagues describe the power of social media, the 500 Women Scientists organization responds to the *Working Life* article, and two scientists recognized by AAAS (the publisher of *Science*) for public engagement discuss how outreach and institutional reform can go hand in hand. In the Online Buzz box, we provide several excerpts from the online eletters we received.

Jeremy Berg

Editor-in-Chief

10.1126/science.aat7935

Social media for social change in science

Although we agree with M. Wright (“Instagram won’t solve inequality,” *Working Life*, 16 March, p. 1294) that there are many systemic structures perpetuating the marginalization of women in science, we view social media as a powerful tool in a larger strategy to dismantle such structures. In addition, scientists have been using social media productively to address several other concerns in academia, including engaging with the public about science, increasing science literacy, promoting trust, exploring career options, networking internationally, and influencing policy.

Strong public trust in science contributes to a democratic, civil society. Scientists have a responsibility to engage effectively with society, especially when trust is lacking (1, 2) and scientific knowledge is not equitably accessible (3). Within academic science, much of this outreach is done by women (4) and underrepresented groups (5). Thus, not surprisingly, outreach has been grossly undervalued and sometimes demeaned. Instead of urging academia to stop celebrating this essential service, we should ensure sufficient compensation and recognition for public engagement. Evidence of outreach is increasingly a

component of publicly funded research grants, and public engagement activities should have weight in merit, tenure, and promotion assessments. Whether scientists do outreach themselves or work with communication and media experts, public engagement with science is a responsibility requiring important skills that should be valued accordingly.

Given the other barriers women and

other marginalized scientists must overcome as minorities in science, technology, engineering, and mathematics (STEM) (6), they should not be expected to bear the full responsibility for outreach—nor should they be penalized for choosing to do this work. Diversity among communicators should be encouraged because multiple styles and approaches of science communication can make science more accessible and relatable to more people, including those who may not otherwise seek STEM education. Selfies on Instagram are optional, but they receive 38% more engagement than pictures without a face (7), enabling open dialogue with broad audiences in an effectively personal manner. Further research can determine whether sharing selfies from a research setting helps confer more trust without sacrificing credibility, and these data will inform strategies for improving the public’s lack of trust in scientists (1, 2).

Social media serve an important role in the movement toward increased equity, diversity, and inclusion within STEM because it provides a widely available, readily accessible platform for many to use easily. Social media allow high-throughput networking and exploration of careers, which benefits trainees who may otherwise lack access to professional development (8). Although not free from the bias and prejudice inherent in society, social media can connect diverse groups, enable rapid information exchange, and mobilize like-minded communities.

This connectivity can allow those same groups to challenge traditional structures, identify and call out systemic barriers, and question hierarchies of power. Instagram, for example, allows for visible representation of individuals who are often unseen, and can amplify voices that may go unheard in traditional settings. Furthermore, increased representation of those who break stereotypes and are underrepresented creates a more inviting perception of STEM careers, and these efforts can improve diversity and inclusion in academia (9–11). For a diverse academic community to thrive, inclusion and acceptance of every scientist, regardless of



appearance (whether conventional or not) is necessary.

No single post or person on social media should be expected to change the world, but social media have been instrumental in mobilizing grassroots political movements, including those related to safety in education, research, and equity, such as the March for Our Lives, the March for Science, Black Lives Matter, #MeToo, and the Women's March. Thus, we challenge the false dichotomy that use of social media for public engagement with science and working to change policy and remove systemic barriers to inclusion are mutually exclusive. Rather, they are intrinsically linked, and we need to harness the potential power of social media to create social change. As scientists, we must look to data and evidence to inform our understanding of the benefits and pitfalls of the use of social media for public outreach and policy change, and uphold the same rigor and analysis in determining what has value and what should be celebrated.

Samantha Z. Yammine,^{1*} Christine Liu,² Paige B. Jarreau,^{3,4} Imogen R. Coe⁵

¹Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada. ²Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94704, USA. ³LifeOmic,

Indianapolis, IN 46202, USA. ⁴Louisiana State University, College of Science, Baton Rouge, LA 70802, USA. ⁵Department of Chemistry and Biology, Ryerson University, Toronto, ON M5B 2K3, Canada.

*Corresponding author.

Email: samantha.yammine@mail.utoronto.ca

REFERENCES

1. 3M, "State of Science Index 2018 Global Report" (2018); <https://multimedia.3m.com/mws/media/15152950/presentation-3m-state-of-science-index-2018-global-report-pdf.pdf>.
2. S. T. Fiske, C. Dupree, *Proc. Natl. Acad. Sci.* **111**, 4 (2014).
3. M. Anderson, "The race gap in science knowledge," *Pew Research Centre* (2015); www.pewresearch.org/fact-tank/2015/09/15/the-race-gap-in-science-knowledge/.
4. E. H. Ecklund, S. A. James, A. E. Lincoln, *PLOS ONE* **7**, e36240 (2012).
5. M. Ong, "The mini-symposium on women of color in science, technology, engineering, and mathematics (STEM): A summary of events, findings, and suggestions" (TERC, Cambridge, MA, 2010).
6. N. Gupta, C. Kemelgor, S. Fuchs, H. Etzkowitz, *Curr. Sci.* **89**, 1382 (2005).
7. S. Bakhshi, D. Shamma, E. Gilbert, "Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems" (2014), pp. 965–974; <http://comp.social.gatech.edu/papers/chi14/faces.bakhshi.pdf>.
8. A. L. Gonzales, *Commun. Res.* **44**, 467 (2015).
9. B. J. Drury, J. O. Siy, S. Cheryan, *Psychol. Inquiry* **22**, 265 (2011).
10. S. D. Hermann et al., *Basic Appl. Soc. Psychol.* **38**, 258 (2016).
11. S. Cheryan, J. O. Siy, M. Vichayapai, B. J. Drury, S. Kim, *Soc. Psychol. Person. Sci.* **2**, 656 (2011).

10.1126/science.aat7303

Journal editors should not divide scientists

We're writing to express our disappointment at the poor judgment that led to the publication of "Instagram won't solve inequality" (M. Wright, *Working Life*, 16 March, p. 1294), which singled out and criticized a successful woman science communicator for her Instagram presence promoting and celebrating science. The editor of this piece should have ensured that the message focused on the issues: Women and underrepresented minorities take on a great deal of science communication, mentorship, and outreach work without recognition or professional reward from their institutions. Despite increasing institutional pressure to communicate about science—whether to increase a university's public profile or meet the National Science Foundation's Broader Impact requirements—many institutions expect the work to be done on personal time without compensation or additional resources. Although the piece hinted at these systemic issues, those arguments were undermined when the editors allowed the author to criticize the work of another woman with an

ONLINE BUZZ

SciComm speaks

The *Working Life* "Instagram won't solve inequality" (M. Wright, 16 March, p. 1294) sparked a wide-ranging discussion about the value and purpose of social media in science. Excerpts from readers' reactions to the article are below. Read the full eletters and add your own at <http://science.sciencemag.org/content/359/6381/1294/tab-e-letters>.

A selection of your thoughts:

Not every tweet, post, or YouTube video that happens to feature a woman science communicator is uploaded with the express intent of challenging the status quo or systemic and institutionalized bias.... To assume this...fails to understand the many reasons why women choose to communicate science to the public. There are indeed activists who constantly challenge the institutionalized bias favoring men, people who sporadically participate in collective

events such as Women in Science day, and also science communicators who just happen to be women. We should applaud all of their efforts....

Victoria J. Forster

...Like the author, I strongly believe that women and other underrepresented minorities in science should feel no obligation to take on additional emotional labor for the sake of educating others. I also agree that systemic issues of inequality will likely require systemic solutions to enact lasting change.... It is evident that the author views #scicomm on Instagram as a chore, but for some of us it is a labor of love. If building model satellites out of cake...or posing my dog in front of Apollo 14 moon trees...weren't incredibly fun, I wouldn't be doing it.... Instagram has significant and largely untapped potential as a vehicle for science communication. The visual nature of the platform, in conjunction with the large and diverse userbase,...provides tremendous opportunity to reach nontraditional audiences. I agree with the author that science communication must be performed in a manner authentic to each individual, but my

hope is that we can continue to encourage each other to promote science in a variety of ways. Right now, we need #scicomm more than ever.

Beth R. Gordon

...As an early-career researcher, the first in my family to go to university, social media has provided me with both community and opportunities that would have been unimaginable without it. Having a window into the lives of other academics and scientists from a range of backgrounds has helped me feel I belong and reassured me that there is a place in the academy for people like me.... At the same time, I was recently invited to publish a comment piece...after an editor noticed my tweets. I have also found coauthors on Twitter and used it to keep up with recent publications and research.... I have nonetheless begun to limit time spent on social media, realizing that it...distracts me from important work. But the benefits far outweigh the limitations....

Glen Wright

10.1126/science.aat7933

unabashed tone of condescension and did not give the target of the comments an opportunity to respond.

Rather than address the roadblocks facing women and underrepresented groups in science, technology, engineering, and mathematics (STEM) or grapple with the author's personal misgivings around science communication, the piece was framed as an attack. The tone implied that anything beyond basic research is a frivolous waste of time, belittling meaningful approaches to science communication and public engagement. It offered a false choice between an authentic and relatable social media presence and effective advocacy for institutional change. The choice to run this inflammatory article demonstrates a lack of thoughtfulness on the editors' part.

Pitting one woman scientist against another is destructive and irresponsible, and it perpetuates unreasonable standards for women and underrepresented groups in STEM. It is antithetical to the open, accessible, and inclusive future that we at 500 Women Scientists envision for science. **Maryam Zaringhalam,* Rukmani Vijayaraghavan, Juniper Simonis,**

Kelly Ramirez, and Jane Zelikova, on behalf of 500 Women Scientists
500 Women Scientists, Boulder, CO 80303, USA.
*Corresponding author.
Email: info@500womenscientists.org

10.1126/science.aat6288

Efforts large and small speed science reform

The Working Life article "Instagram won't solve inequality" (M. Wright, 16 March, p. 1294) asserts that science outreach efforts by individual women cannot counteract structural inequities and that women are doing outreach at a cost to their own careers. We concur that collective action and structural change are needed to diversify science and improve meaningful science engagement with the public. However, when such reform is absent or too slow, individual efforts fill the vacuum and should not be condemned.

Along with hundreds of other scientists, we devote time and energy to individual public engagement initiatives, while pushing for institutional reforms to support more scientists who wish to engage effectively. These reforms would provide support and incentives through professional recognition,

financial and logistical resources, networks of support, and an inclusive culture and capacity for public engagement. With support, more scientists could develop collaborative and innovative engagement practices to broaden participation in science. While changing the culture of public engagement, we must similarly push to dismantle other structural barriers to women and minorities in the sciences. To accelerate these changes,

"...when [structural change] is absent or too slow, individual efforts fill the vacuum..."

data collection and learning networks would enable us to improve the effectiveness of our efforts to create a diverse workforce and tackle science-societal challenges. Individual action versus structural change is not an "either/or" question; it is a "yes, and."

Anne J. Jefferson^{1*} and Melissa A. Kenney²
¹Department of Geology, Kent State University, Kent, OH 44242, USA. ²CMNS-Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20742, USA.

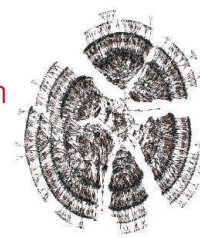
*Corresponding author. Email: ajjeffer9@kent.edu

10.1126/science.aat6341

RESEARCH

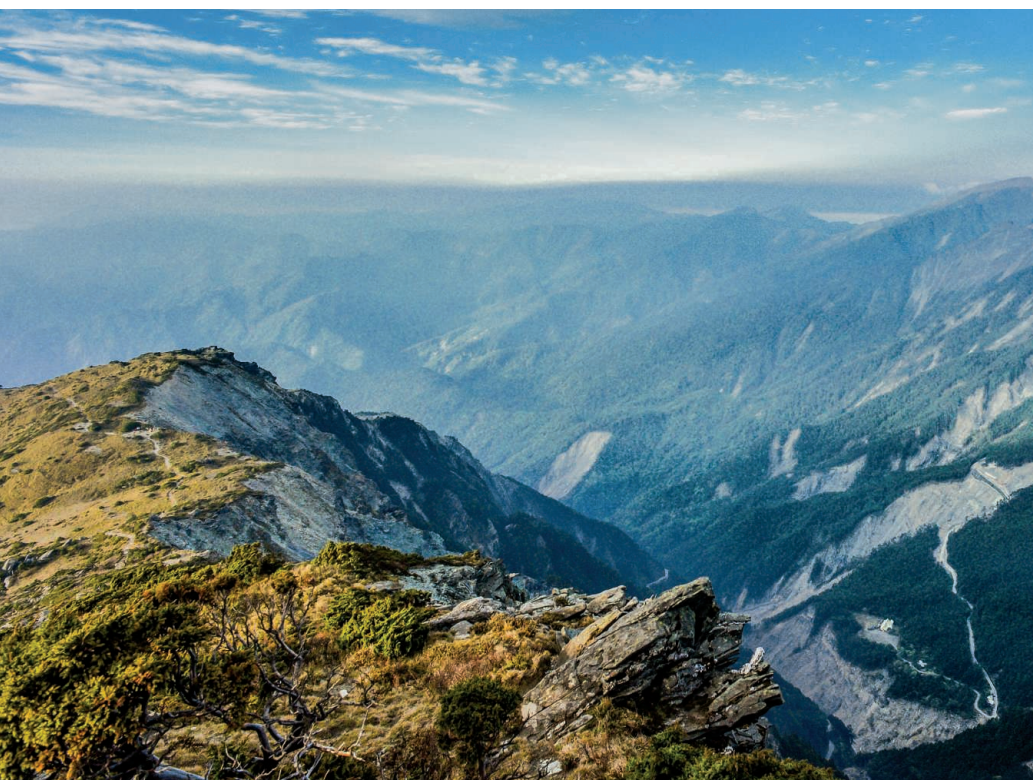
Genetic architecture of human longevity and migration

Kaplanis et al., p. 171



IN SCIENCE JOURNALS

Edited by **Caroline Ash**



CARBON CYCLE

Microbes eat rocks and leave carbon dioxide

The reaction of atmospheric carbon dioxide (CO_2) with silicate rocks provides a carbon sink that helps counterbalance the release of CO_2 by volcanic degassing. However, some types of rocks contain petrogenic organic carbon, the oxidation of which adds CO_2 to the atmosphere, counteracting the drawdown by silicates. Hemingway et al. present evidence from the rapidly eroding Central Range of Taiwan showing that microbes oxidize roughly two-thirds of the petrogenic organic carbon there and that the rate of oxidation increases with the rate of erosion. —HJS

Science, this issue p. 209

Microbes oxidize most of the petrogenic organic carbon in Taiwan's fast-eroding Central Range.

QUANTUM INFORMATION

Scaling up to supremacy

Quantum information scientists are getting closer to building a quantum computer that can perform calculations that a classical computer cannot. It has been estimated that such a computer would need around 50 qubits, but scaling up existing architectures to this number is tricky. Neill et al. explore how increasing the number of qubits from five to nine affects the quality of the output of their superconducting qubit device. If, as the number of qubits grows further, the error continues to increase at the same rate, a quantum computer with about 60 qubits and reasonable fidelity

might be achievable with current technologies. —JS

Science, this issue p. 195

ORGANIC CHEMISTRY

A guide for catalyst choice in the forest

Chemists often discover reactions by applying catalysts to a series of simple compounds. Tweaking those reactions to tolerate more structural complexity in pharmaceutical research is time-consuming. Ahneman et al. report that machine learning can help. Using a high-throughput data set, they trained a random forest algorithm to predict which specific palladium catalysts would best tolerate isoxazoles

(cyclic structures with an N–O bond) during C–N bond formation. The predictions also helped to guide analysis of the catalyst inhibition mechanism. —JSY

Science, this issue p. 186

PLANT SCIENCE

Dormancy by communication shutdown

Trees become dormant in winter, with encapsulated buds protected against harsh conditions. Tylewicz et al. found that, as the days get shorter, communication channels between cells in aspen trees shut down. The blocked plasmodesmata sequester the dormant meristems from growth signals. Growth-promoting

signals can be turned on and off relatively rapidly, but the closed plasmodesmata are not so nimble. Thus, despite the occasional sunny day, the trees stay dormant until spring. —PJH

Science, this issue p. 212

NOROVIRUS

Aiding and abetting norovirus disease

Norovirus is highly infectious and usually causes transient, acute disease. In some individuals, norovirus persists and is associated with inflammatory bowel disorders. While investigating the cell tropism for murine norovirus, Wilen et al. discovered that a rare cell type,

tuft cells, carrying the CD300lf receptor were the virus's specific target. Tuft cells proliferate in response to the type 2 cytokines interleukin-4 and interleukin-25, which thereby amplify norovirus infection. Moreover, infected tuft cells are resistant to immune clearance. This effect may explain the associated persistent disease symptoms that humans can suffer. —CA

Science, this issue p. 204

METROLOGY

Refining the fine-structure constant

The fine-structure constant, α , is a dimensionless constant that characterizes the strength of the electromagnetic interaction between charged elementary particles. Related by four fundamental constants, a precise determination of α allows for a test of the Standard Model of particle physics. Parker *et al.* used matter-wave interferometry with a cloud of cesium atoms to make the most accurate measurement of α to date. Determining the value of α to an accuracy of better than 1 part per billion provides an independent method for testing the accuracy of quantum electrodynamics and the Standard Model. It may also enable searches of the so-called “dark sector” for explanations of dark matter. —ISO

Science, this issue p. 191

STRUCTURAL BIOLOGY

The RNA exosome captured in action

The RNA exosome, a major RNA degradation machine, processes ribosomal RNA (rRNA) precursors and is directly coupled to the protein synthesis machine, the ribosome. Using cryo-electron microscopy, Schuller *et al.* investigated the structure of the precursor large ribosomal subunit from yeast with unprocessed rRNA in complex with the RNA exosome. The structure captures a snapshot of two molecular machines

transiently interacting and explains how the RNA exosome acts on an authentic physiological substrate and remodels the large subunit during ribosome maturation. —SYM

Science, this issue p. 219

PALEONTOLOGY

Early evolution of insect scales

Organisms use tiny structures on their surfaces to produce striking optical effects. The wing scales of butterflies and moths exhibit some of the most diverse physical colors produced by insects, but whether they have always been equipped with photonic structures is unknown. Zhang *et al.* used fossil evidence to establish that these insects possessed color-eliciting structures at least 130 million years earlier than previously thought. They determined the ultrastructure of wing scales from Jurassic Lepidoptera and mid-Cretaceous Tarachoptera. They then used optical modeling to reconstruct the colors that these features would produce. —PJB

Sci. Adv. 10.1126/sciadv.1700988 (2018).

HIV

Zooming in on human lymph nodes

Follicular helper T cells (T_{FH}) play an essential role in shaping B cell-mediated antibody responses. Wendel *et al.* used mass cytometry and T cell receptor sequencing to examine the T_{FH} response in lymph node tissue collected from HIV⁺ individuals. HIV infection altered the clonality of T_{FH} cells, with severe infections associating with pronounced oligoclonal T_{FH} responses. T_{FH} cells in the lymph nodes of HIV⁺ individuals secreted interleukin-21 but were less polyfunctional than T_{FH} cells from healthy individuals. The lack of polyfunctionality correlated with impaired isotype switching of B cells in the lymph nodes. —AB

Sci. Immunol. 3, eaan8884 (2018).

IN OTHER JOURNALS

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



Bill color in waxbills changes with external temperature.

SOCIAL SIGNALS

Sexual signals not so strict

Sexual signals in animals, such as bright plumage, are thought to be predetermined or to be badges of quality that can reflect an animal's current condition. Direct and immediate effects of the environment in which an animal lives are rarely considered to shape these phenotypes. Funghi *et al.*, however, found that in waxbills, bill color—a trait that can change quickly—is not the result of predetermined sexual differences, aggression, or sexual selection, but rather appears to be influenced by changes in the abiotic environment. Bill brightness was reduced in females after a series of lower-temperature nights. The authors suggest that this indicates that environmental conditions place constraints on these types of traits, limiting the degree to which they can reflect quality or be used for social interaction. —SNV

Behav. Ecol. Sociobiol. 10.1007/s00265-018-2486-6 (2018).

MATERIALS SCIENCE

Silicon sheds its harmonicity

The widespread technological uses for silicon make understanding this element's physical properties very important. Kim *et al.* performed inelastic neutron

scattering experiments on single crystals of silicon to measure the vibrational properties up to 1500 K. Silicon has some odd thermal properties at certain temperatures, and these experiments show the need to account for a number of factors to explain the unusual thermal expansion

ALSO IN SCIENCE JOURNALS

Edited by **Caroline Ash**

SYNTHETIC BIOLOGY

Writing a cell's history in its DNA

Recording cellular events could advance our understanding of cellular history and responses to stimuli. The construction of intracellular memory devices, however, is challenging. Tang and Liu used Cas9 nucleases and base editors to record amplitude, duration, and order of stimuli as stable changes in both genomic and extrachromosomal DNA content (see the Perspective by Ho and Bennett). The recording of multiple stimuli—including exposure to antibiotics, nutrients, viruses, and light, as well as Wnt signaling—was achieved in living bacterial and human cells. Recorded memories could be erased and re-recorded over multiple cycles. —SYM

Science, this issue p. 169;
see also p. 150

CELL BIOLOGY

The mitoCPR unclogs mitochondria

The import of proteins into mitochondria is essential for cell viability. How cells respond when mitochondrial protein import is impaired is poorly understood. Weidberg and Amon showed that upon mitochondrial import stress, yeast cells mounted a response known as the mitoCPR. mitoCPR was activated when mitochondrial protein import was impaired and unimported precursors accumulated on the organelle's surface. mitoCPR restored mitochondrial functions by clearing stalled proteins from the import channels. It did this by inducing expression of *Cis1*, which recruited the adenosine triphosphatase *Msp1* to import channels to remove unimported precursors and target them for degradation by the proteasome. —SMH

Science, this issue p. 170

POPULATION BIOLOGY

Quantitative analysis of millions of relatives

Human relationships, as documented by family trees, can elucidate the heritability of a host of medical and biological parameters. Kaplanis *et al.* collected 86 million publicly available profiles from a crowd-sourced genealogy website and used them to examine the genetic architecture of human longevity and migration patterns (see the Perspective by Lussier and Keinan). Various models of inheritance suggested that life span is predominantly attributable to additive genetic effects, with a smaller component from dominant genetic inheritance. The data also suggested that relatedness between individuals is less attributable to advances in human transportation than to cultural changes. —LMZ

Science, this issue p. 171;
see also p. 153

TOPOLOGICAL MATTER

A topological superconductor

A promising path toward topological quantum computing involves exotic quasiparticles called the Majorana bound states (MBSs). MBSs have been observed in heterostructures that require careful nanofabrication, but the complexity of such systems makes further progress tricky. Zhang *et al.* identified a topological superconductor in which MBSs may be observed in a simpler way by looking into the cores of vortices induced by an external magnetic field. Using angle-resolved photoemission, the researchers found that the surface of the iron superconductor $\text{FeTe}_{0.55}\text{Se}_{0.45}$ satisfies the required conditions for topological superconductivity. —JS

Science, this issue p. 182

SINGLE-CELL GENOMICS

Identifying single-cell types in the mouse brain

The recent development of single-cell genomic techniques allows us to profile gene expression at the single-cell level easily, although many of these methods have limited throughput. Rosenberg *et al.* describe a strategy called split-pool ligation-based transcriptome sequencing, or SPLiT-seq, which uses combinatorial barcoding to profile single-cell transcriptomes without requiring the physical isolation of each cell. The authors used their method to profile >100,000 single-cell transcriptomes from mouse brains and spinal cords at 2 and 11 days after birth. Comparisons with *in situ* hybridization data on RNA expression from Allen Institute atlases linked these transcriptomes with spatial mapping, from which developmental lineages could be identified. —LMZ

Science, this issue p. 176

STRUCTURAL BIOLOGY

A close-up view of oligosaccharyltransferase

Many secretory and membrane proteins are modified through the attachment of sugar chains by N-glycosylation. Such modification is required for correct protein folding, targeting, and functionality. In mammalian cells, N-glycosylation is catalyzed by the oligosaccharyltransferase (OST) complex via its STT3 subunit. OST forms a complex with the ribosome and the Sec61 protein translocation channel. Braunger *et al.* combined cryo-electron microscopy approaches to visualize mammalian ribosome-Sec61-OST complexes in order to build an initial molecular model for mammalian OST. —SMH

Science, this issue p. 215

IMMUNOLOGY

Autoantibody redemption through rapid mutations

Antibodies distinguish foreign epitopes from closely related self-antigens by poorly understood mechanisms. In mice, Burnett *et al.* found that a proportion of B cells could cross-react with similar foreign and self-antigens (see the Perspective by Kara and Nussenzweig). Challenge with self-antigen resulted in anergy (i.e., a lack of immune response), which was reversed by exposure to high-density foreign antigen. Mutations that decreased self-affinity were rapidly selected for, whereas selection for epistatic mutations that enhanced foreign reactivity took longer. Self-reactivity, rather than being an impediment to immunization, resulted in higher affinities against a foreign immunogen. —STS

Science, this issue p. 223;
see also p. 152

CLIMATE

Climate effects of aerosol cleanup

Many aerosols emitted by human activities have a cooling effect on the climate and can also change precipitation patterns. In a Perspective, Samset highlights the magnitude of these influences at regional levels. Worldwide, aerosols have reduced the impacts of greenhouse gas emissions on air temperatures. Impacts on precipitation have also been substantial but more variable. Because of the negative impacts of aerosol emissions on health, efforts to reduce them are gathering pace, but this has important implications for future warming and precipitation patterns in many regions of the world. —JFU

Science, this issue p. 148

CANCER

Can wound healing worsen metastasis?

Early metastatic recurrence in breast cancer patients could be caused by tumor cells released into the circulation during primary resection or could be the result of existing metastatic outgrowth. To distinguish between these possibilities, Krall *et al.* used a common wound-healing model in mice harboring breast cancer cells in which the primary tumor bed was not disturbed by surgery. They found that T cells can keep tumor cells in check, but if wound healing is induced, inflammation disrupts this balance. Anti-inflammatory treatment reduced metastasis in the mice. Existing clinical data indicate that perioperative anti-inflammatories reduce early metastatic recurrence in breast cancer patients. By separating surgery from resection, these results may explain this curious clinical occurrence. —LP

Sci. Transl. Med. **10**, eaan3464 (2018).

STRUCTURAL BIOLOGY

Signaling for nitrogen fixation

The nitrogen-fixing bacterium *Bradyrhizobium japonicum* enables high-yield production of soybeans with little use of nitrogen fertilizers, a major source of nutrient pollution. Using structural and modeling techniques, Wright *et al.* generated a model by which a two-component system of this bacterium, comprising the histidine kinase sensor and response regulator, responds to low oxygen to stimulate the expression of genes required for nitrogen fixation. These results may help in the development of plant growth modulators that are unlikely to affect mammalian species, which do not signal through two-component systems. —AV

Sci. Signal. **11**, eaaq0825 (2018).

NANOMATERIALS

Synthesizing graphene nanopores

Nanosize pores in graphene can make its electronic properties more favorable for transistor applications and may also be useful for molecular separations. Moreno *et al.* used Ullmann coupling to polymerize a dibromo-substituted diphenyl-bianthracene on a gold surface (see the Perspective by Sinitskii). Cyclodehydrogenation of the resulting polymer produced graphene nanoribbons, and cross-coupling of these structures created a nanoporous graphene sheet with pore sizes of about 1 nanometer. Scanning tunneling spectroscopy revealed an electronic structure in which semiconductor bands with an energy gap of 1 electron volt coexist with localized states created by the pores. —PDS

Science, this issue p. 199;
see also p. 154

Exposure to pesticide-contaminated wildflowers harms common blue butterflies.



ENVIRONMENT

Wildflower contamination with neonicotinoids

Neonicotinoid pesticides are the most widely used type of insecticides, but there are concerns that they are toxic to nontarget species such as bees and butterflies. Basley and Goulson report on a combined field and laboratory experiment aimed at assessing the impact of neonicotinoids on the common blue butterfly (*Polyommatus icarus*). Wildflowers planted along the margins of fields of neonicotinoid-treated wheat were contaminated with the pesticide at levels similar to those in the treated crops. Common blue butterfly larvae exposed to neonicotinoid-contaminated plants showed increased mortality and reduced growth in the early stages of development. Wildflower margins that specifically aim to boost pollinator populations may chronically expose these species to harmful levels of neonicotinoids. —JFU

Environ. Sci. Technol. **52**, 3990 (2018).

behavior. This in-depth look at silicon helps refine theoretical models and provides a better understanding of this technologically important material. —BG

Proc. Natl. Acad. Sci. U.S.A. **10.1073/pnas.1707745115** (2018).

GENOMICS

Denisovans shaped our genomes, twice

Studies of “molecular relics” from archaic humans in modern human genomes have shown that independent interbreeding events occurred between the ancestors of Eurasians and the Neandertals and Denisovans. Because these archaic admixtures happened after the out-of-Africa migration of the modern human ancestors, comparing present-day non-African and African genomes can reveal introgression events without the need for an archaic reference genome. Using this approach, Browning *et al.* found evidence for two pulses of gene flow from distinct Denisovan populations into modern humans in

East Asian and Papuan genomes. These findings point to at least two populations of Denisovans that contributed genes to modern humans. —SYM

Cell **173**, 53 (2018).

CARBON SEQUESTRATION Reforestation to enhance the soil carbon sink

Soil is a major pool of carbon and hence can play a key role as a carbon sink in strategies to mitigate climate change. For the United States, Nave *et al.* quantified the carbon stocks in forest topsoils, focusing on the potential of reforestation to enhance carbon sequestration. Their estimates indicate that managed reforestation of >500,000 km² would increase the topsoil sink by 1.3 to 2.1 petagrams of carbon within a century, enhancing the forest carbon sink in the United States by 10% annually. Their results also indicate that this enhanced sink would persist for decades, contributing to the offsetting

of greenhouse gas emissions and reversing a decline in the strength of the carbon sink in U.S. forests. —AMS

Proc. Natl. Acad. Sci. U.S.A. **115**, 2776 (2018).

STRUCTURAL BIOLOGY Seeing the clasps that stabilize prion fibrils

A cryo-electron microscopy method called MicroED (micro-electron diffraction) has been used to reveal the core structures of several amyloid fibrils. With this technique, Gallagher-Jones *et al.* determined a 0.72-Å-resolution structure of fibrils formed by a peptide at the core of the infectious scrapie form of mammalian prion protein (proto-PrP^{Sc}). Like the full PrP^{Sc}, the fibril is characterized by unusually high stability. The high-resolution structure shows β -strands that stack into β -sheets, with sheets pairing front-to-back to form fibrils. A network of hydrogen bonds within and between β -strands

forms “polar clasps,” which are shielded by aromatic residues that stack in the fibrils. —VV

Nat. Struct. Mol. Biol. **25**, 131 (2018).

TECHNOLOGY ADOPTION Superstars drive regional drug use

A recent study finds that early adoption of new cancer drugs was geographically influenced by high-profile investigators (“superstars”) on the key clinical trials supporting the drugs. Agha and Molitor combed treatment records and clinical trial publications for 21 newly approved drugs in the United States. Patients in the same region as the lead investigator on the key trial were 36% more likely to use the drug during the first 2 years after it was approved and showed better rates of survival. These findings suggest that policies to promote the adoption of technology may blunt the potential impact if they do not include improved local information. —BW

Rev. Econ. Stat. **100**, 29 (2018).

RESEARCH ARTICLE SUMMARY

SYNTHETIC BIOLOGY

Rewritable multi-event analog recording in bacterial and mammalian cells

Weixin Tang and David R. Liu*

INTRODUCTION: The stable recording of cellular events has the potential to advance our understanding of a cell's history and how cells respond to stimuli. However, the construction of intracellular memory devices that record a history of cellular events has proven challenging.

RATIONALE: We developed two CRISPR-mediated analog multi-event recording apparatus (CAMERA) systems that record cellular events as durable changes in the DNA of bacteria or mammalian cells. In CAMERA 1, Cas9 nucleases are used to shift the ratio of two recording plasmids, and signals are recorded in the form of plasmid ratios. Writing in CAMERA 2 uses base editors to produce single-base modifications at designated positions of plasmid or genomic DNA. Both Cas9 nucleases and base editors can be programmed to target multiple DNA sequences with different guide RNAs, and both are known to function across many cell types. These features enable CAMERA to serve as a multiplexable, analog, rewritable intracellular recording system.

RESULTS: We demonstrate that the ratio of the recording plasmid pair in CAMERA 1 can be stably maintained in bacteria over 144 hours and a dilution ratio of 10^{17} . By using a writing complex of the Cas9 nuclease and a guide RNA to selectively target one of the recording plasmids, we can cause this plasmid ratio to shift in a dose-dependent manner. The presence or absence of a stimulus is recorded in CAMERA 1 by linking to the expression of the writing complex. The analog format of CAMERA 1 enables recording of signal amplitude over a known time scale, or recording of the duration of a signal of known strength. Two resetting methods enable cells harboring CAMERA 1 to function over repeated cycles of recording and erasing.

CAMERA 2 uses base editors to record stimuli of interest as permanent single-base modifications in cellular DNA. Predictable and dose-dependent accumulation of base editing was observed over 68 generations in bacteria. CAMERA 2 achieved analog recording of multiple stimuli of interest, including exposure to antibiotics, nutrients, viruses, and light. When

recording to a high-copy plasmid, CAMERA 2 provides reliable readout by sequencing only 10 to 100 cells and can record event order using an overlapping guide RNA design.

CAMERA 2 also functions in human cells by recording stimuli to safe-harbor genomic loci. We show that CAMERA 2 can be multiplexed, such that two responsive guide RNA expression cassettes can be used to record the presence of two exogenous small molecules in mammalian cells. Finally, we demonstrated CAMERA 2 recording of Wnt signaling, a crucial endogenous mammalian signaling pathway, as a permanent change in genomic DNA in human cells by placing the expression of the writing complex under the control of a Wnt-responsive promoter.

CONCLUSION: Base editors and CRISPR nucleases were used to create "cell data recorders" that enable durable, analog recording of stimuli and cell states. CAMERA systems are sensitive, multiplexable, resettable, and compatible with both bacteria and mammalian cells, and thus may be useful for applications such as recording the presence of extracellular and intracellular signals, mapping cell lineage, and constructing cell state maps. ■

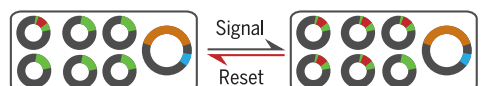
Merkin Institute for Transformative Technologies in Healthcare, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA, and Department of Chemistry and Chemical Biology and Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA.

*Corresponding author. Email: drliu@fas.harvard.edu
Cite this article as W. Tang, D. R. Liu, *Science* **360**, eaap8992 (2018). DOI: [10.1126/science.aap8992](https://doi.org/10.1126/science.aap8992)

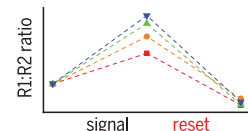
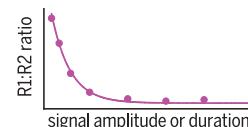
Multiplexed analog cellular recording by CAMERA systems in bacteria and mammalian cells.

CAMERA 1 records stimuli as changes in the ratio of mutually exclusive DNA sequences. CAMERA 2 uses base editors to record the duration or amplitude of signals as single-nucleotide changes. Both systems can be multiplexed to independently record multiple events, including exposure to antibiotics, nutrients, viruses, and light, as well as Wnt signaling.

CAMERA1
Cas9 nuclease-mediated recording
in bacteria



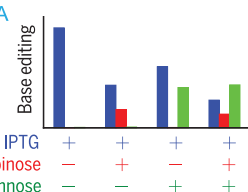
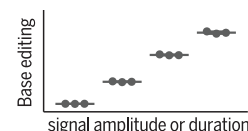
- Writing plasmid: expresses Cas9 + sgRNA
- Recording plasmid 1 (R1): ...CCACCGGC...
- Recording plasmid 2 (R2): ...CCTGAGGC...



CAMERA2
Base editor-mediated recording
in bacteria and human cells



- Writing plasmid: expresses base editor + sgRNA
- Recording sites: ...TCA...TCA...ACG... (plasmid or genomic)
- ...TTA...TTA...ATG...



RESEARCH ARTICLE

SYNTHETIC BIOLOGY

Rewritable multi-event analog recording in bacterial and mammalian cells

Weixin Tang and David R. Liu*

We present two CRISPR-mediated analog multi-event recording apparatus (CAMERA) systems that use base editors and Cas9 nucleases to record cellular events in bacteria and mammalian cells. The devices record signal amplitude or duration as changes in the ratio of mutually exclusive DNA sequences (CAMERA 1) or as single-base modifications (CAMERA 2). We achieved recording of multiple stimuli in bacteria or mammalian cells, including exposure to antibiotics, nutrients, viruses, light, and changes in Wnt signaling. When recording to multicopy plasmids, reliable readout requires as few as 10 to 100 cells. The order of stimuli can be recorded through an overlapping guide RNA design, and memories can be erased and re-recorded over multiple cycles. CAMERA systems serve as “cell data recorders” that write a history of endogenous or exogenous signaling events into permanent DNA sequence modifications in living cells.

Recent technologies have enabled the study of the internal state of cells in exquisite detail, including the sequence of the genome, the status of epigenetic modifications, and the identity and abundance of cellular RNAs, proteins, and metabolites that collectively determine cell state (1, 2). Far less developed are tools to reveal a cell's history and how that history determines present and future cell states, despite the potential impact of such capabilities. For example, detailed information about cell states during division and differentiation could illuminate the process of aging, and recording the presence and duration of exposure to external or internal stresses could yield clues about the emergence of cancer and other diseases. Recording a cell's history in a highly multiplexable, durable, and minimally perturbative form has been a long-standing challenge in the life sciences (3, 4).

Transient recording of environmental signals has been achieved by manipulating transcription and translation in bacteria (5). Information recorded in this manner, however, cannot be passed on to future generations of cells, and the recording process itself is delicate because many factors contribute to transcription and translation efficiencies. In contrast, recombinases can modify designated genomic sequences, and the resulting information stored in DNA can be read even after cell death (6–9). Although individual signals of interest can be stably recorded using recombinase-based memory devices, orthogonal recombinases

are required to record more than one bit of information. Cellular recording devices operated by recombinases have been applied to record the presence or suggest the absence of stimuli, but their use to record signal strength, duration, or order is more challenging (3).

Cellular memory devices can record in digital and analog formats. Whereas digital memory devices store information in one of two distinct states (on or off), analog memory devices leave permanent marks in DNA in a manner that reflects the strength or duration of endogenous or exogenous stimuli. Such recordings, in theory, could illuminate cellular history, reveal how a stimulus dictates downstream responses, and improve our ability to predict cell behavior (3). Recently, Farzadfar and Lu (10) reported synthetic cellular recorders integrating biological events (SCRIBE), an elegant memory device that translates exogenous signals into point mutations in a bacterial genome through beta protein-assisted single-stranded DNA incorporation. Because the production of single-stranded DNA by the adapted retrovirus cassette is not efficient, SCRIBE requires the sampling of large populations of bacteria for both recording and readout (10).

To develop a memory device that is less dependent on a large cell population, we chose the CRISPR (clustered regularly interspaced short palindromic repeats)–Cas9 nuclease (11–14) and Cas9-derived base editors (15, 16) to serve as DNA writing modules. Both Cas9 nuclease and base editors make changes in cellular DNA in an efficient and programmable manner when complexed with guide RNAs (11, 15). If linked to stimuli or cell state changes, these DNA modifications in principle could serve as durable messages that reflect a cell's history and could be read out using modern sequencing technologies, even after cell

death. Here, we present two CRISPR-mediated analog multi-event recording apparatus (CAMERA) systems and demonstrate their ability to simultaneously record multiple cell states, including exposure to antibiotics, nutrients, viruses, light, and a kinase inhibitor that alters endogenous Wnt signaling.

A plasmid compensation system as an information carrier in bacteria

We chose the *Streptococcus pyogenes* Cas9 (SpCas9) nuclease as an initial DNA writing module because it functions robustly across many different cell types in vitro and in vivo (13, 17). SpCas9 makes double-stranded DNA breaks at loci that match the 20-base “spacer region” of a single guide RNA (sgRNA) and that are near an NGG protospacer-adjacent motif (PAM). In mammalian cells, the resulting double-stranded breaks can be repaired by nonhomologous end joining and similar processes to introduce insertions and deletions (indels), or through homology-directed repair by supplying a template strand. In bacteria, however, double-stranded DNA breaks frequently cause cell death or a loss of extrachromosomal DNA (18, 19). To translate DNA loss after double-stranded breaks into durable information, we designed a high-copy number plasmid compensation system to store DNA modification states. This strategy enables analog recording within each cell and thereby avoids dependence on large cell populations.

The plasmid compensation system includes a pair of nearly identical recording plasmids, R1 and R2, that differ only by 3 nucleotides in an *EGFP* gene that encodes enhanced green fluorescent protein (Fig. 1A). The *EGFP* gene in R1 expresses full-length fluorescent protein, whereas the *EGFP* gene in R2 contains a premature stop codon and cannot produce fluorescent protein (Fig. 1A). Because the two plasmids are virtually identical, we hypothesized that their fitness cost to host cells is very similar and that they should coexist in a stable ratio for long periods of time.

The R1/R2 ratio serves as the information carrier that reflects the signal of interest in an analog mode. To convert the signal of interest into an R1/R2 ratio change, a Cas9-sgRNA pair induced by the stimulus cleaves plasmid R1 but not R2 (Fig. 1A). The resulting double-stranded break causes the loss of R1. Because the two recording plasmids share the same origin of replication that controls the total copy number of the plasmids in bacteria, the loss of R1 initiates the replication of the remaining plasmids and the gradual accumulation of R2. A high-copy number plasmid origin (pUC) was chosen to maximize the analog recording range of the system (Fig. 1A).

To test the stability of the plasmid compensation recording system, we cotransformed *Escherichia coli* strain S1030 (20) with R1 and R2 and then isolated two single colonies with different R1/R2 ratios. The colonies were separately grown in LB media at 37°C, and the culture was diluted 500- or 1000-fold six times over 144 hours for a total dilution ratio of 10^{17} (Fig. 1B). The two starting colonies contained 29% R1 and 60% R1, and their

Merkin Institute for Transformative Technologies in Healthcare, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA, and Department of Chemistry and Chemical Biology and Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA.

*Corresponding author. Email: drliu@fas.harvard.edu

R1/R2 ratio was very stably maintained throughout the growth and dilution process (Fig. 1B), ending at 29% R1 and 59% R1, respectively. These results indicate that the R1/R2 ratio can serve as a stable analog information carrier across a range of plasmid ratios.

To assess the potential growth burden that the recording plasmid pair might impose on bacteria, we measured growth curves for the parental *E. coli* strain S1030 and two S1030 colonies containing R1 and R2 in different ratios (29% or 60% R1; fig. S1). The colonies harboring the recording plasmid pair exhibited the same growth rate as the parental strain in the presence or absence of the selection antibiotic, and all bacterial cultures

reached the same final cell density; these results suggest that the recording plasmids do not substantially impair bacterial fitness.

A CRISPR nuclease writing module enables CAMERA 1

We designed a writing module that cleaves R1, but not R2, near the 3-nucleotide region that differs between R1 and R2. This region was chosen to be proximal to the PAM to maximize the selectivity of the writing module (21) (Fig. 1A). The *EGFP* gene fragments from both plasmids were incubated in vitro with the Cas9-sgRNA complex. The functional *EGFP* gene amplified from plasmid R1, but not the mutated *EGFP* gene encoded by

plasmid R2, was cleaved into two fragments (Fig. 1C). These results establish that the writing module can distinguish plasmids R1 and R2 and introduce double-stranded breaks selectively in R1.

Next, we moved the system into live bacteria to test whether we could translate an exogenous signal into a durable change in the DNA content of the cell. We placed a TetO promoter that is inducible with anhydrotetracycline (aTc) upstream of the *Cas9* gene, and placed a constitutive Lac promoter upstream of the R1-targeting sgRNA in writing plasmids W1.0.1 to W1.0.3 (Fig. 1D and fig. S2), thereby forming the CAMERA 1.0 system. Bacteria containing CAMERA 1.0 with an R1/R2

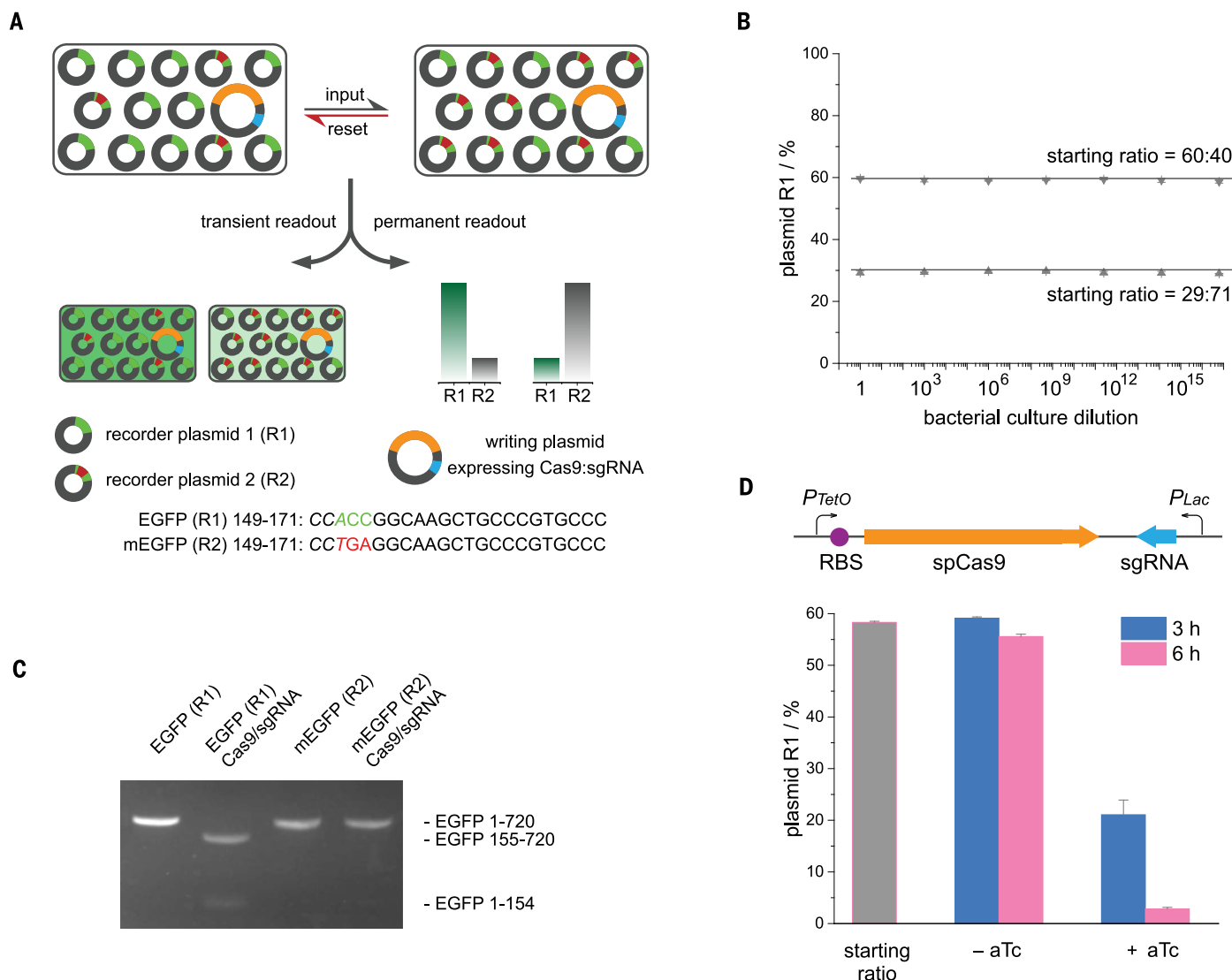


Fig. 1. Recording in CAMERA 1 uses Cas9 nuclease to shift the ratio between a pair of recording plasmids. (A) Schematic representation of CAMERA 1. Recording plasmids R1 and R2 are identical except a 3-nucleotide coding mutation in the *EGFP* gene. The expression of the Cas9-sgRNA complex is controlled by the signal of interest and results in R1 depletion in the bacteria that carry the recording plasmid pair. (B) Stability of the R1/R2 ratio in *E. coli* S1030 cells in

the absence of the writing plasmid. (C) In vitro cleavage of the wild-type and mutated *EGFP* gene by Cas9 in the presence of sgRNA1. The designed spacer sequence targets the distinct region in *EGFP* so the Cas9-sgRNA complex cleaves R1 much faster than R2. (D) Recording the amplitude and duration of aTc by CAMERA 1.0. Values and error bars reflect the mean and SD of three replicate cultures derived from a single bacterial colony.

ratio of 58:42 were used to test aTc-stimulated recording. After being cultured in the presence or absence of aTc for 3 hours and 6 hours, the bacteria were harvested and analyzed for their R1 content by high-throughput sequencing (HTS). In the absence of aTc, R1 content remained steady (59%) after 3 hours and was only slightly lower (56%) after 6 hours (Fig. 1D). This basal level of R1 consumption can be attributed to low-level transcription of the uninduced TetO promoter. In contrast, R1 content responded strongly to the presence of aTc and decreased to 21% in 3 hours, and to 4% after 6 hours (Fig. 1D). Collectively, these results suggest that CAMERA 1.0 can sensitively detect and record the presence of an ex-

ogenous small molecule and the duration of exposure in an analog format.

Using CAMERA 1 derivatives to record multiple stimuli

To enable recording of more than one stimulus, we installed the LacO promoter, which is suppressed by LacI and activated by isopropyl β -D-thiogalactopyranoside (IPTG), upstream of the sgRNA to generate CAMERA 1.1 (Fig. 2A). Both aTc and IPTG are required to initiate recording in CAMERA 1.1. We chose a bacterial colony carrying CAMERA 1.1 with a starting R1 content of 77% and applied different inducer combinations for 3 hours (Fig. 2A). As expected, the R1/R2 ratio

remained stable in the absence of stimuli or in the presence of 0.5 mM IPTG only. A slight decline in R1 content (to 70%) was observed when the bacteria were treated only with aTc (100 ng/ml) (Fig. 2A), consistent with the known leakiness of the LacO promoter in the absence of IPTG (22). However, R1 content decreased to 37% when bacteria were cultured in the presence of both aTc and IPTG (Fig. 2A); this result indicates that both stimuli are required to promote substantial R1/R2 ratio changes, recapitulating an “AND” Boolean logic gate (23).

One advantage of the CAMERA 1 design is that it records signals in an analog format that can capture more information than binary switches.

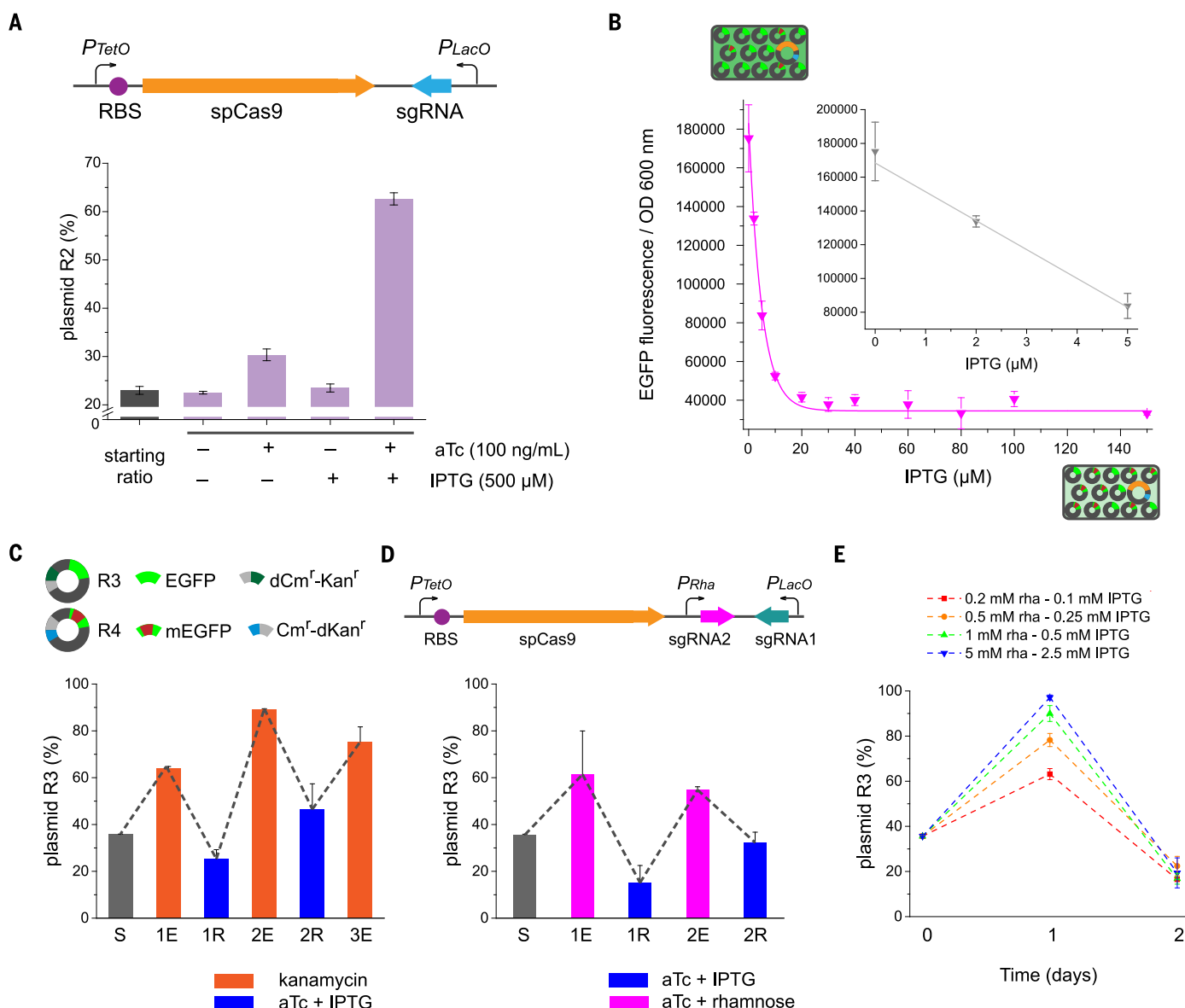


Fig. 2. Multi-event recording and resetting of CAMERA 1 systems.

(A) Construction of a “AND” Boolean logic gate using CAMERA 1.1. Both IPTG and aTc are required for initiation of the recording process. (B) Analog recording of IPTG concentration by CAMERA 1.1 as reported by EGFP fluorescence. (C) Repeated recording and erasing of CAMERA 1.2 by application of the small-

molecule inducers and kanamycin. S, starting state; E, erase (5 to 20 generations); R, record (5 to 10 generations). (D) Repeated recording and erasing of CAMERA 1.3 by inducing different writing complexes. The inducer aTc was constantly supplied at 100 ng/ml. (E) Dose-dependent recording and erasing using CAMERA 1.3. Values and error bars reflect mean and SD of three replicates.

To explore the analog recording capabilities of CAMERA 1.1, we treated the bacterial culture with different doses of IPTG ranging from 0 to 150 μ M with a constant aTc input of 50 ng/ml for 3 hours (Fig. 2B). The R1 content was followed by monitoring EGFP fluorescence and by DNA sequencing. *EGFP* expression was initiated by diluting the bacterial culture with fresh media lacking aTc or IPTG after the recording process was finished. As anticipated, the EGFP signal decreased as the concentration of IPTG increased, reflecting an increased depletion rate of R1, saturating at 30 μ M IPTG (Fig. 2B). The relationship between EGFP loss and IPTG concentration at low dosages (≤ 5 μ M) was predictable and linear (Fig. 2B), which suggests that the R1/R2 ratio can be used to infer signal amplitude in a reliable manner. HTS of the bacterial culture confirmed these dose-dependent changes in R1/R2 ratio (fig. S3). Collectively, these findings establish that CAMERA 1.1 can record multiple stimuli of interest in an analog, dose-dependent, and durable manner.

Erasing and re-recording of CAMERA 1

Memory devices are particularly versatile if they can be erased and rewritten as needed. Instead of using R1 and R2, the CAMERA 1.2 system contains two recording plasmids, R3 and R4, that each confer resistance to different antibiotics. Similar to R1, R3 can be targeted by a writing plasmid expressing Cas9 and an sgRNA to cause a shift in the R3/R4 ratio. To minimize the difference in fitness cost of R3 and R4 to host cells, we fused genes encoding two antibiotic resistance proteins, chloramphenicol acetyltransferase (Cat, which inactivates chloramphenicol), and aminoglycoside-3'-phosphotransferase (Aph3', which targets kanamycin), and incorporated a single amino acid mutation in either of the two domains. R3 expressed inactive Cat H195A (24) fused to wild-type Aph3', whereas R4 expressed inactive Aph3' D208A (25) fused to wild-type Cat (Fig. 2C). Because both plasmids express two nearly identical proteins, their relative fitness cost in the absence of antibiotic should be minimal. In the presence of either antibiotic, R3 and R4 should confer different fitness benefits.

Bacteria containing a starting R3 content of 39% maintained a steady R3/R4 ratio in conditions lacking antibiotic and responded to the presence of chloramphenicol or kanamycin by shifting the plasmid ratio in a dose-dependent manner favoring the plasmid with the corresponding functional resistance domain (fig. S4). These results indicate that the information stored in the R3/R4 ratio can be reset in either direction using exogenous small molecules. By successively exposing cells to media containing either kanamycin (to reset the R3/R4 ratio to a high level) or aTc + IPTG (to induce Cas9 + sgRNA production and cleave R3, lowering the R3/R4 ratio), we performed three successive rounds of erasing and recording using CAMERA 1.2, with strong response levels in each round (Fig. 2C). Hence, this system can be used repeatedly to record and erase exposure to stimuli.

We developed an alternative resetting mechanism in CAMERA 1.3 that is independent of antibiotic resistance by including a second sgRNA circuit. In addition to one guide RNA cassette (sgRNA1) present in writing plasmid W1.2 that targets R3, we incorporated a second guide RNA expression unit (sgRNA2) under the control of a rhamnose-inducible promoter (P_{Rha}) to generate writing plasmid W1.3. The Cas9-sgRNA2 complex targets R4. Similar to the recording process in which the expression of sgRNA1 controlled by IPTG results in the loss of R3, the transcription of sgRNA2, induced by rhamnose, should lead to the cleavage of R4 and thus restore plasmid R3 levels. Indeed, *E. coli* strain S1030 that carried 36% or 77% R3 successfully went through multiple rounds of recording and erasing upon alternating exposure to rhamnose or IPTG (Fig. 2D and fig S5). In addition, the strength of the stimulus (here, the concentration of rhamnose or IPTG) was reflected in the rate of R3/R4 change (Fig. 2E).

HTS analysis of the recording plasmids after the final round of resetting and recording revealed a minimal frequency ($\leq 0.06\%$) of insertions and deletions (indels) (table S1). This suggests that Cas9-mediated DNA cleavage does not substantially induce random mutations in the plasmid compensation system in bacteria, and that both the recording and erasing processes result in minimal loss of future recording or erasing function. Taken together, these results validate CAMERA 1.2 and 1.3 as rewritable, durable cellular memory devices with distinct resetting mechanisms.

Base editing mediates recording in CAMERA 2

We recently developed base editors, chimeric proteins consisting of a DNA base modification enzyme, a catalytically impaired Cas9 nickase, and a base excision repair inhibitor (15, 26–28). Base editors efficiently introduce single C•G \rightarrow T•A mutations at guide RNA-programmed loci in a wide variety of eukaryotic cells and organisms (15, 16, 29–34). Predictable, durable point mutation of genomic or plasmid DNA resulting from base editing has the potential to serve as an ideal information carrier in synthetic memory devices (Fig. 3A). To incorporate a base editor in CAMERA, we first characterized base editing in *E. coli*, as base editors have not been extensively used in prokaryotic cells. Because bacteria lack nick-directed mismatch repair exploited by the third-generation base editor (BE3), we used the second-generation base editor (BE2) that contains a cytidine deaminase fused to a catalytically dead Cas9 (dCas9), rather than to a Cas9 nickase, as the protein component of our writing complex (15).

In writing plasmid 2.0 (W2.0), BE2 expression is induced by aTc and sgRNA1 is constitutively transcribed (Fig. 3A). To test whether CAMERA 2.0, constructed using W2.0 and recording plasmid R1, can faithfully record the amplitude and duration of an exogenous signal, we treated the bacterial culture with aTc at different concentrations and diluted it repeatedly to ensure constant expression of the writing complex. When complexed with sgRNA1, BE2 introduces a C•G \rightarrow T•A

mutation at position 166 of the *EGFP* gene in recording plasmid R1. As anticipated, base editing occurred in an analog mode, and the total percentage of modified base increased with bacterial passage number in a highly linear and remarkably reproducible relationship (Fig. 3B). This observation indicates that base editing with BE2 in bacteria is robust and cumulative, reflecting the duration of exposure to the stimulus that induces expression of the writing complex. Moreover, the rate of editing can be controlled in a dose-dependent manner (Fig. 3B). By the end of the experiment (68 passages), 66% editing was observed with aTc supplied at a concentration of 200 ng/ml and no significant decrease in editing rate was observed as the recording proceeded; these results suggested that given enough time, base editing could approach 100% in bacteria.

Editing at the target locus accumulated at a slow but constant rate when aTc was present at a low concentration of 2 ng/ml (Fig. 3B). Under these low induction conditions, only 12% of the total recording range (C•G \rightarrow T•A conversion at position 166 of the *EGFP* gene) was consumed by bacterial generation 68 (Fig. 3B), which suggests that CAMERA 2.0 can function as a molecular clock that records over hundreds of generations. Collectively, these findings establish CAMERA 2.0 is a highly responsive analog memory device that uses base editing to faithfully record the amplitude of an exogenous signal over a known time scale, or the duration of a signal of known strength, in the form of single nucleotide changes.

Using CAMERA 2 systems to record multiple stimuli

We hypothesized that multiplexed recording could be achieved by CAMERA through the use of multiple responsive guide RNA expression cassettes. To test this possibility, we constructed additional base editor writing plasmids W2.1, W2.2, and W2.3 by replacing the Lac promoter of the guide RNA in writing plasmid W2.0 with promoters regulated by IPTG, arabinose, and rhamnose, respectively, to generate devices CAMERA 2.1, 2.2, and 2.3 (Fig. 3C and fig. S6). Similar to CAMERA 2.0, writing promoted by the BE2-sgRNA1 complex in CAMERA 2.1 occurred in a highly reproducible, predictable, and dose-dependent manner (Fig. 3C). The leaky transcription of the TetO promoter enabled very slow but steady recording in the absence of aTc, whereas the recording space was consumed at a much faster speed in the presence of both IPTG and aTc (Fig. 3C).

To test whether the information recorded in CAMERA could be used to deduce the total exposure time of the device to a stimulus, we passaged bacteria carrying CAMERA 2.0 for 40 generations and treated either the first 20 generations or the second 20 generations with aTc (100 ng/ml; Fig. 3D). The accumulation rate of editing at position 166 of the *EGFP* gene was strongly determined by exposure duration, and the presence or absence of aTc within a certain time window could be determined by comparing the editing rate of the sample with those of control samples that were always exposed to, or

always shielded from, the stimulus (Fig. 3D). Similarly, bacteria carrying CAMERA 2.1 were treated with 0.5 mM IPTG for either the first half or the second half of the total incubation time (Fig. 3E). The editing rate strongly correlated with the presence of IPTG, and the total accumulated editing frequencies in the two groups were nearly identical by the end of the experiment; this indicated that the information recorded by CAMERA 2.1 faithfully reflected the duration of exposure to the signal, regardless of when the exposure took place (Fig. 3E). Collectively, these observations suggest that the rate of base editing at a given time point can be used to deduce the dose of the stimulus, and that the stimulus duration can be calculated from the total base-editing conversion if the stimulus dose is known.

The presence of both aTc and a second stimulus is required for CAMERA 2.1, 2.2, and 2.3 to initiate recording—a process that mimics the behavior of an “AND” gate. Indeed, in the ab-

sence of stimuli, CAMERA 2.2 showed no detectable activity, with $\leq 0.1\%$ C•G \rightarrow T•A editing at position 186 of the *EGFP* gene (fig. S6). Neither arabinose nor aTc by itself increased editing significantly. However, the presence of both inducers resulted in 9.0% C•G \rightarrow T•A conversion after 24 hours, which suggests that CAMERA 2.2 functions as a tightly regulated “AND” gate. Similarly, both rhamnose and aTc were required to initiate recording at position 195 of the *EGFP* gene by CAMERA 2.3 (fig. S6). We tested the recording efficiency at different concentrations of rhamnose in the presence of aTc (200 ng/ml) and confirmed that C•G \rightarrow T•A conversion at position 195 correlated well with the dose of rhamnose, again demonstrating that signal intensity can be faithfully recorded and stored by CAMERA 2.

One advantage of adapting CRISPR technologies to build synthetic memory devices is that multiple stimuli can in theory be recorded using multiple guide RNA units. To test whether

CAMERA could simultaneously record multiple independent signals, we integrated all three small molecule-responsive guide RNA expression circuits from writing plasmids W2.1, W2.2, and W2.3 into writing plasmid W2.4. Bacteria carrying CAMERA 2.4 were treated with different combinations of the four small-molecule inducers, and indeed, editing at the designated *EGFP* positions could be used to infer the presence of the corresponding writing complexes and hence their corresponding stimuli (Fig. 3F and fig. S7). The fidelity of the device is not compromised even in more complicated environments in which more than two stimuli are provided (Fig. 3F and fig. S7). These findings indicate that CAMERA 2 is a versatile and multiplexable memory device.

CAMERA 2 enables recording of event order

Memory devices that are capable of recording the order of biological events are of great interest

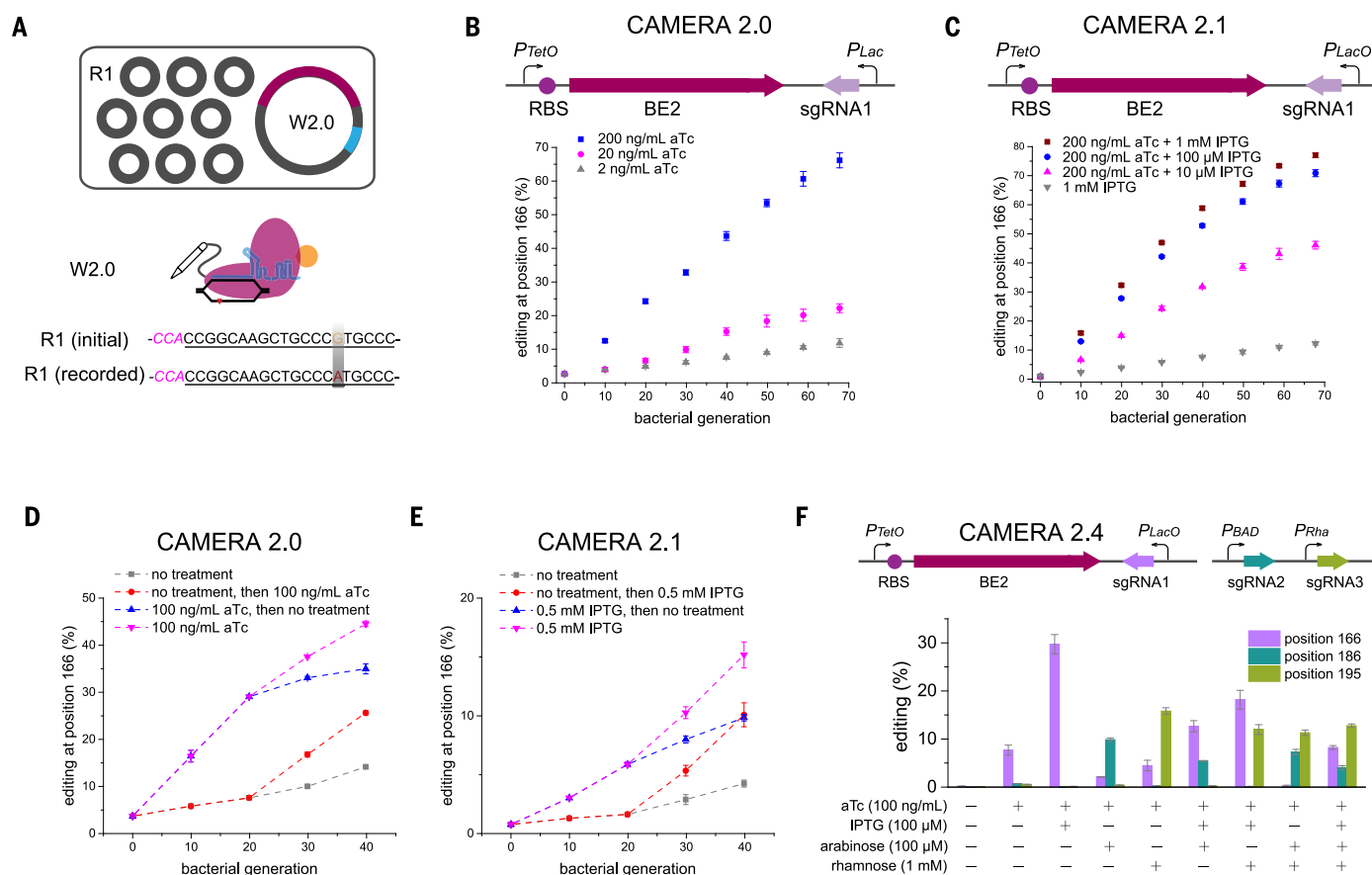


Fig. 3. CAMERA 2 systems use base editing to record the amplitude and duration of exogenous signals. (A) Schematic representation of CAMERA 2. The writing plasmid expresses the writing complex consisting of BE2 and sgRNAs. The recording plasmid is targeted by the writing complex to generate memory in the form of C•G \rightarrow T•A substitutions at guide RNA-specified loci. (B) Recording the concentration of aTc and the treatment duration in analog mode using CAMERA 2.0. (C) Recording the concentration of IPTG in the presence or absence of aTc and the treatment duration in analog mode using CAMERA 2.1. (D) The rate of base editing recorded in CAMERA 2.0 reflects the schedule of exposure to the inducer. (E) CAMERA 2.1 records the total time of

exposure to IPTG, regardless of treatment pattern. (F) Recording four exogenous stimuli using CAMERA 2.4. The presence of each signal, individually or in different combinations, was recorded by base editing at each of three specified positions in the *EGFP* gene. We constructed two mathematical models to simulate the behavior of CAMERA 2.4. A model that accounts for promoter leakage and competition between multiple guide RNAs for BE2 results in a more “digital” CAMERA 2.4 in which the absolute editing level at each position more readily reveals the presence or absence of the corresponding stimulus (fig. S7). Values and error bars reflect the mean and SD of three replicates.

(3), as the order of changes in a cell's environment or in the state of a cell can strongly determine cell fate (35). Murray and co-workers recently described a two-input temporal logic gate that was constructed using integrases to record the order and timing of inputs, but the limited number of possible output states (GFP, RFP, or neither) necessitated the sharing of the same output among five different combinations of ordered inputs, complicating the assignment of multiple cell states (36). We hypothesized that CAMERA 2 systems could record events that occur in a specific order by overlapping two base-editing targets, such that base editing of DNA target 1 mediated by writing complex 1 (BE2-sgRNA5) is required before DNA target 2 can be recognized by writing complex 2 (BE2-sgRNA6). To test this possibility, we constructed CAMERA 2.5 in which the order of exposure to two small-molecule inducers, arabinose and rhamnose, could be recorded (Fig. 4A). The three arabinose-induced C•G → T•A modifications resulting from base editing by writing complex 1 are located within target site 2 near its PAM. Rhamnose-induced sgRNA6 recognizes target site 2 only after modification by writing complex 1, but should not edit this site before base editing by writing complex 1 has taken place (Fig. 4A). Thus, base editing at this sgRNA6-specified position should be initiated if rhamnose (stimulus 2) is provided after arabinose (stimulus 1), but not if the order of stimuli is reversed.

By using an additional target site of sgRNA6 spanning positions 116 to 135 of a modified *EGFP* gene, CAMERA 2.5 is further equipped with the ability to independently record two stimuli (Fig. 4A). Whereas editing at positions 205 to 207 and at position 129 records exposure to arabinose and rhamnose, respectively (Fig. 4B and fig. S8), the ratio of base editing at position 216 versus that at position 129—both promoted by writing complex 2—reflects the order of application of the two stimuli (Fig. 4C). The activating treatment order of arabinose followed by rhamnose resulted in a position 216/position 129 base-editing ratio of 0.54. The ratio was lower by a factor of 6.8 (0.08; Fig. 4C) when the treatment order was reversed such that rhamnose exposure preceded arabinose exposure. Together, these results indicate that CAMERA 2.5 can record cellular events in a strongly order-dependent manner.

Using CAMERA 2 derivatives to record phage infection and light

We further applied the CAMERA 2 architecture in bacteria to sense (i) viral infection of host cells by bacteriophage and (ii) exposure to light. A phage shock promoter (PSP) driving sgRNA1 transcription was included in CAMERA 2.6 (Fig. 4D) (37, 38). Without phage infection, 9% base editing was observed at *EGFP* position 166 (Fig. 4D), consistent with previous reports of background transcriptional activity of PSP in the

absence of phage (39). Base editing at position 166 increased by a factor of 4.7 (to 42%) after infection with phage (Fig. 4D). Similarly, using a light-responsive expression system based on light-inhibited expression of the *cl* repressor gene, CAMERA 2.7 could record the presence of light with a factor of 59 increase in recording site editing efficiency (Fig. 4E) (40). These results collectively demonstrate that the CAMERA 2 is capable of recording, as single-nucleotide changes in bacterial DNA, a wide range of signals including exposure to antibiotics, nutrients, viruses, and light.

In principle, the recording process carried out by CAMERA systems should not require a large population of cells because the recording plasmid is present in hundreds of copies in each cell. To test the possibility of recording and reading CAMERA data in small cell populations, we characterized how light exposure was recorded by CAMERA 2.7 in a handful of cells as well as at the single-cell level (Fig. 4E and fig. S9). As expected, CAMERA 2.7 reliably recorded bacterial exposure to light in bulk cultures, with editing at *EGFP* position 166 in $\sim 10^6$ cells increasing in a linear fashion with light exposure duration (from 1.2% to 57% editing over 3 days; Fig. 4E).

Reliable recording and signal readout were also achieved using only 100-cell or 10-cell samples throughout the 3-day recording process, although larger variations were observed with fewer cells,

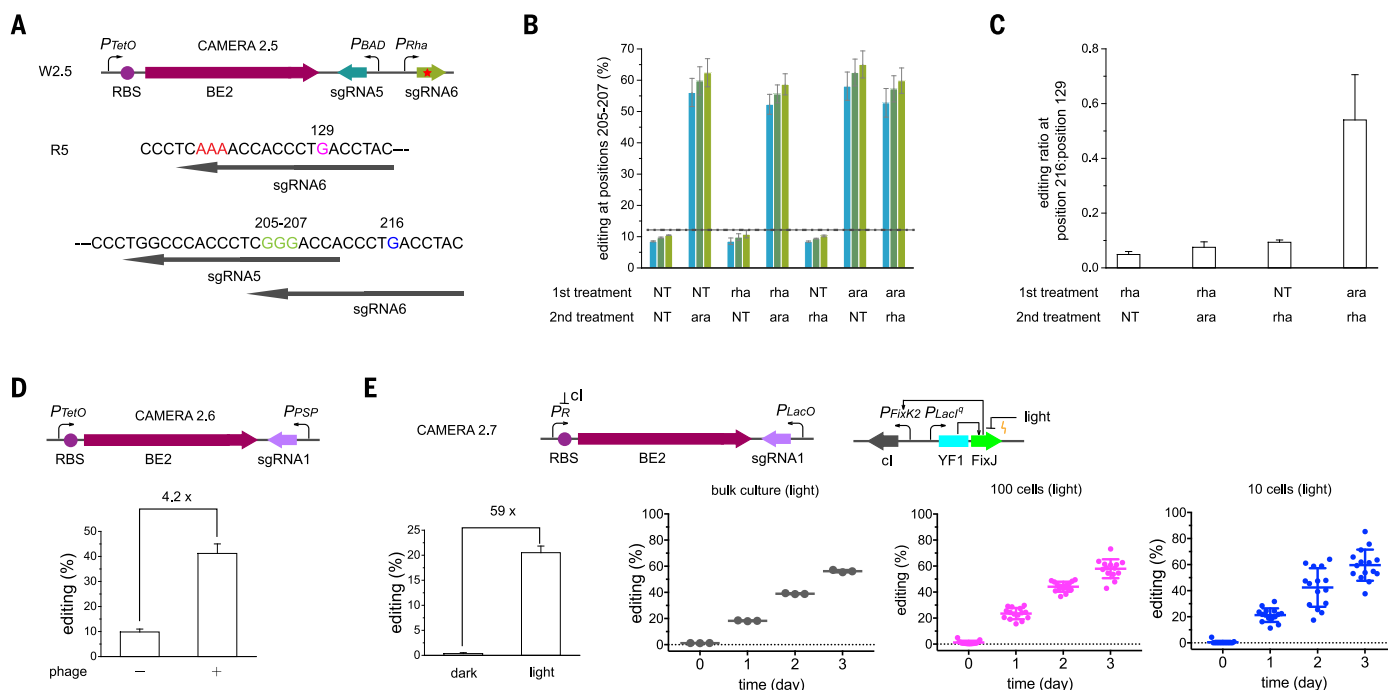


Fig. 4. CAMERA 2 records the order of stimuli and a wide range of environmental signals. (A) Schematic representation of CAMERA 2.5 that records stimuli in an order-dependent manner. (B) CAMERA 2.5 records the presence of arabinose at positions 205 to 207 in format of C•G → T•A mutations. (C) The ratio of base editing at position 216 versus that at position 129 in CAMERA 2.5 indicates the order of exposure to two stimuli. A position 216/position 129 base-editing ratio above 0.1 was only observed when the bacteria were treated first with arabinose and then with

rhamnose, but not when arabinose exposure followed rhamnose exposure.

(D) Phage infection recording by CAMERA 2.6. (E) Light exposure recording with CAMERA 2.7 in bulk culture and in small numbers of cells. Light exposure duration can be recorded faithfully in bulk culture as well as in samples of only 100 or 10 cells. Values and error bars in bar graphs and the dot plot in (E) for bulk cultures reflect the mean and SD of three replicates. Dots and error bars in dot plots in (E) for 100 and 10 cells represent the mean and SD of 15 replicates of randomly sorted sets of 100 and 10 cells.

as expected (Fig. 4E). Even measuring 15 single-cell signals yielded average light duration-dependent editing efficiencies similar to those from bulk cultures (fig. S9). These data demonstrate that CAMERA can support analog-like recording even in very small populations of cells.

CAMERA 2m systems record cell states in mammalian cells

Finally, we tested CAMERA 2 variants (CAMERA 2m systems) in human embryonic kidney (HEK) 293T cells and chose an established human safe-harbor gene, *CCR5* (41), as the recording locus (Fig. 5A). We designed three individual sgRNAs that target different regions of the *CCR5* gene (CAMERA 2m.0; Fig. 5B and fig. S10). Total C•G → T•A editing of 37%, 46%, or 19% was obtained at target positions A, B, or C of the *CCR5* gene when using corresponding guide RNAs A, B, or C with BE3 (fig. S10). The C•G → T•A conversion frequency at each site increased by at least a factor of 270 relative to controls lacking the corresponding guide RNA. Robust multiplexed recording was also achieved using the three sgRNAs in all possible combinations, and less than 0.07% editing was observed at any site for which the corresponding guide RNA was not supplied; this finding suggests that base-editing cross-talk between guide RNAs is minimal in these human cells.

By placing BE3 expression under a doxycycline-controlled promoter, the presence of the drug was recorded in the *CCR5* loci with a base-editing frequency higher than in cells that never encountered doxycycline by a factor of 60 (CAMERA 2m.1, Fig. 5C). In addition, by placing sgRNA expression under TetR- and LacI-suppressed promoters, CAMERA 2m.1 recorded the presence of both doxycycline and IPTG at different positions in the *CCR5* loci (CAMERA 2m.2, Fig. 5D); this result confirms that CAMERA 2m can be multiplexed to record combinations of inputs in human cells.

The Wnt signaling pathway plays a crucial role in embryonic development, and aberrant Wnt signaling is associated with a variety of diseases in humans (42). We sought to record Wnt signaling using CAMERA 2m in human cells. To achieve this goal, we placed the expression of BE3 under a (LEF-TCF)₇ promoter (43) that responds to Wnt signaling to initiate downstream gene expression in CAMERA 2m.3. Cells transfected with CAMERA 2m.3 were treated with LiCl, a GSK3 inhibitor that has been demonstrated to activate Wnt signaling (Fig. 5E) (44). We included a (LEF-TCF)₇-BE3-P2A-Luc construct that expresses a firefly luciferase protein together with BE3 so that we could monitor Wnt both by luminescence and by HTS of the *CCR5* recording locus. As expected, cells transfected with (LEF-TCF)₇-BE3-P2A-Luc exhibited a factor of 140 increase in Wnt signaling-driven luciferase expression upon LiCl treatment (fig. S11). This increase in Wnt signaling was permanently recorded by a factor of 53 increase in base editing at the *CCR5* locus (Fig. 5E). These results demonstrate that Wnt signaling, a major

endogenous mammalian signaling pathway, can be recorded by CAMERA 2m in human cells.

Discussion

We developed synthetic memory devices that record events of interest in live cells by means of two distinct CRISPR-mediated DNA modification mechanisms: Cas9 nuclease-catalyzed double-stranded DNA cleavage and base editor-mediated point mutation. CAMERA records the amplitude of stimuli over a known time scale, or the duration of stimuli with a known amplitude, permanently in the DNA of live cells. The analog nature of both recording systems allows the continuous monitoring of signals of interest and thereby provides more information than canonical digital memory devices.

In CAMERA 1 systems (table S2), information is recorded in the form of plasmid R1/R2 ratios. Because R1 but not R2 expresses a functional fluorescent protein, information stored in CAMERA 1 systems can be read out transiently by monitoring post-recording cellular fluorescence in addition to the permanent readout by HTS. During the development of CAMERA 1, we decreased the ribosome-binding site (RBS) strength of Cas9 by four orders of magnitude (see supplementary text) to slow down the recording process, indicating that these systems can respond quickly and are highly sensitive. This exceptional sensitivity may enable recording of very weak environmental signals that would otherwise be difficult to detect using other methods. Endy and co-workers achieved resetting using recombinase-based synthetic memory devices (6). In this work, we developed two different strategies for CAMERA resetting that also enable repeated cycles of erasing and rewriting.

CAMERA 2 systems (table S2) translate stimuli of interest into single-nucleotide modifications. The devices can be multiplexed by stacking multiple responsive sgRNA units, and we demonstrated that four exogenous signals could be recorded using CAMERA 2.4. Through a ratcheted overlapping protospacer design, CAMERA 2.5 recorded events in an order-dependent manner—a capability that is difficult to achieve using other synthetic memory devices. By including environment-responsive circuits, virus infection and light exposure have also been faithfully recorded using CAMERA 2.6 and 2.7. The cI repressor translation mechanism integrated in CAMERA 2.7 is particularly versatile and can be used to translate a variety of environmental signals. Similar repression systems have been applied to monitor genetic (mutational inactivation of the *cI* gene) and epigenetic (proteolytic inactivation of cI) changes in *E. coli* by Radman and co-workers (45). We also demonstrated that by recording to high-copy plasmids, CAMERA 2 maintains its reliability even in samples containing only 10 to 100 cells. The mammalian cell compatibility of base editing enables CAMERA 2m systems to function in human cells, including its use to record both exposure to external stimuli and flux through an endogenous signaling pathway.

Incorporating the recently developed adenine base editor that mediates A•T → G•C base edit-

ing (27) could expand the versatility of CAMERA 2 systems by adding an additional dimension of recording that can directly reverse the edits introduced by BE3. CRISPR technology has been applied in mammalian cells for molecular recording of exogenous signals and mapping of cell lineage using genomically integrated circuits (46–48). The Cas1/Cas2 DNA sequence capture system has been applied for analog recording in bacteria using large cell populations (49). While this work was under review, Wang and co-workers further developed this recording mechanism and reported an elegant “biological tape recorder” using Cas1/Cas2 and copy-inducible plasmids to follow exogenous signals in a multiplexed format over time (50). In contrast to these devices, CAMERA systems are less dependent on genomic integration of barcoded scratchpads or protospacer arrays that could result in unpredictable changes in the genome and undesired cellular perturbations.

As a synthetic memory device that uses novel recording mechanisms, CAMERA has its own limitations. Because of the sensitivity of the writing process, background recording in the absence of the stimulus can be observed when using less tightly regulated induction circuits. As a result, the sensitivity of CAMERA may need to be tuned for different applications. For example, a weaker RBS can be used to express base editors and Cas9 nucleases when background recording is undesired. Moreover, when recording to genomic loci, CAMERA 2 cannot achieve single-cell readout of analog information and will typically require the analysis of a population of cells.

We demonstrated only the construction of a simple AND logic gate in this work. Additional research is needed to explore more complicated logic gates (23) using CAMERA. The shortage of orthogonal inducible expression cassettes also limits the application of CAMERA in more complex setups. This limitation might be addressed as more inducible transcriptional and translational regulation circuits are developed.

These limitations notwithstanding, the use of base editors in CAMERA 2 systems minimizes stochastic indels and translocations that arise from double-stranded DNA breaks. The capability of recording many endogenous signaling pathways of interest in a minimally perturbative and highly multiplexable manner offers substantial benefits for investigations of mammalian cell states. The small sample size of 10 to 100 cell states that CAMERA requires to achieve faithful analog recording in bacteria may prove especially useful for applications in which limited cellular material is available. We envision that CAMERA can be used for applications such as recording the presence of low-abundance extracellular and intracellular signals, mapping cell lineage, and constructing complex cell state maps.

Materials and methods Cloning and plasmids

Oligonucleotides were ordered from Integrated DNA Technologies. PCR fragments for plasmid construction were amplified using PhuU polymerase (ThermoFisher Scientific) and assembled by USER

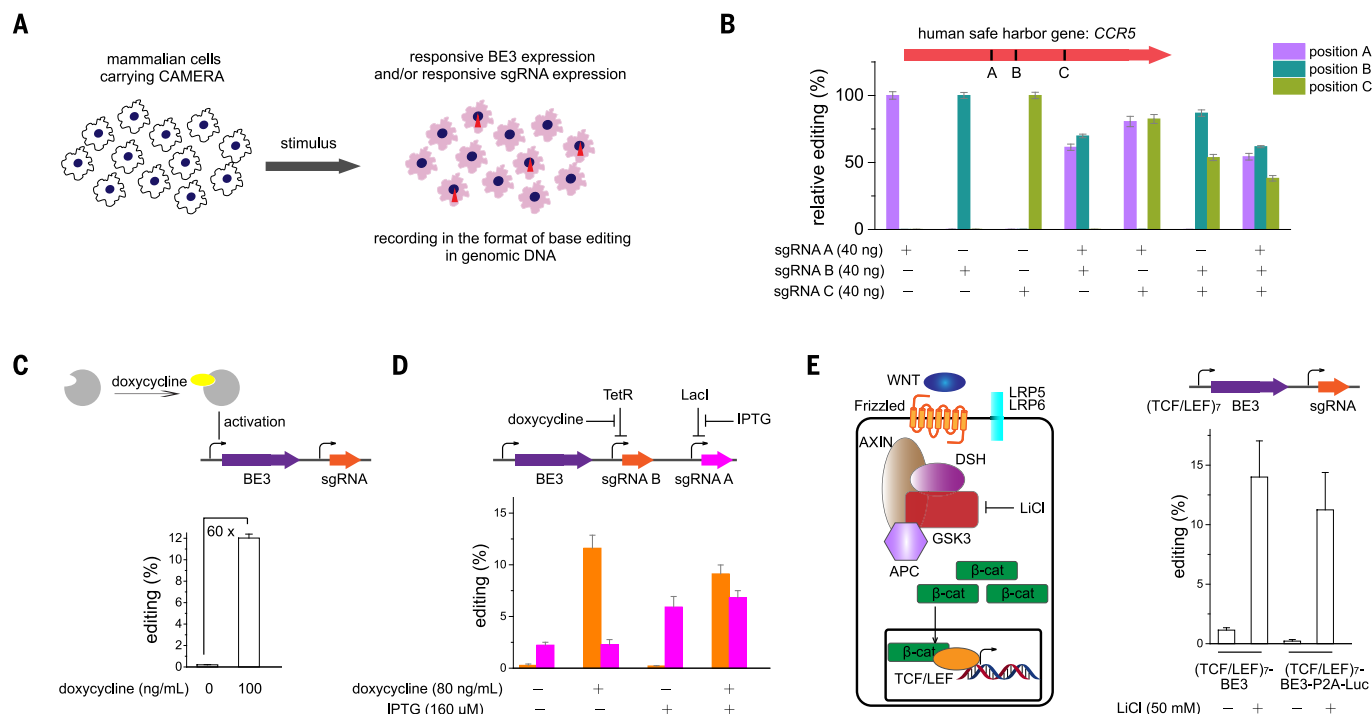


Fig. 5. CAMERA 2m functions in mammalian cells. (A) Schematic representation of CAMERA 2m in mammalian cells. (B) CAMERA 2m.0 records information in the format of C•G→T•A modifications at guide RNA-specified positions of the human safe-harbor gene *CCR5*. (C) The presence of doxycycline is recorded by CAMERA 2m.1 through a doxycycline-controlled transcriptional activator. (D) CAMERA 2m.2 records the

presence of doxycycline and IPTG in a multiplexed manner. Expression of sgRNA A and sgRNA B is repressed by LacI and TetR in the absence of stimuli and can be turned on by the addition of IPTG and doxycycline, respectively. (E) CAMERA 2m.3 responds to Wnt signaling and records the presence of a Wnt signaling stimulus at a target genomic safe-harbor locus. Values and error bars reflect mean editing and SD of three replicates.

enzyme mix (New England Biolabs) according to the manufacturer's instructions. All DNA cloning was performed with NEB Turbo cells (New England Biolabs). Plasmids used in this work (see table S3 for plasmid design specifics) are available from Addgene. Primers used for HTS are listed in table S6.

Strains and chemicals

All bacterial CAMERA devices developed in this work were tested in *E. coli* strain S1030 (19) with the exception of CAMERA 2.6, which was characterized in *E. coli* strain S2063. The complete genotypes of S1030 and S2063 are listed in table S4. Unless otherwise noted, antibiotics were used at the following concentrations: carbenicillin, 100 mg/ml; kanamycin, 50 mg/ml; chloramphenicol, 25 mg/ml; spectinomycin, 100 mg/ml. All chemicals were purchased from Sigma-Aldrich and Fisher Scientific.

Characterization of CAMERA 1.1 in *E. coli* S1030

E. coli S1030 were transformed with a mixture of 500 ng of R1, 500 ng of R2, and 100 ng of W1.1 and plated on LB agar containing carbenicillin and spectinomycin. A total of eight colonies were picked, grown to dense cultures, and analyzed for their R1 content. The bacterial culture carrying CAMERA 1.1 with 77% R1 and 23% R2 was selected for further testing and split into three individual cultures. A bacterial culture was inoculated 1:500 (v/v) into fresh LB media containing (i) no inducer, (ii) aTc (100 ng/ml), (iii) 500 μM

IPTG, and (iv) aTc (100 ng/ml) and 500 μM IPTG. The treated bacteria were allowed to grow at 37°C with shaking for 3 hours, and the R1/R2 ratio was analyzed by amplifying the *EGFP* fragment and sequencing using HTS.

To characterize the analog behavior of CAMERA 1.1, the starting cultures were inoculated 1:100 (v/v) into fresh LB media containing 0, 2, 5, 10, 20, 30, 40, 60, 80, 100, or 150 μM IPTG in the presence of aTc (50 ng/ml). The treated bacteria were allowed to grow at 37°C with shaking for 3 hours and the inducers were removed by diluting the culture in a 1:250 ratio with fresh LB and culturing overnight. The resulting R1/R2 ratio in the bacterial culture was analyzed by amplifying the *EGFP* gene and sequencing in a high-throughput manner. To induce the *EGFP* expression as a transient readout, the bacterial culture was diluted again in a 1:125 ratio with fresh LB containing 0.25 mM arabinose. EGFP fluorescence was measured after 4 hours of induction using a TECAN Infinite M1000 Pro plate reader with excitation/emission wavelengths set to 485/530 nm.

Recording and erasing of CAMERA 1.2

E. coli S1030 were transformed with 500 ng of R3 and 500 ng of R4. The transformed bacteria were plated on LB agar containing kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml) to select for the presence of both plasmids. A total of eight colonies were picked, grown in fresh LB, and analyzed for their R3 content. A bacterial culture containing 38% R3 and 62% R4 was se-

lected to test whether antibiotic treatment could promote the R3/R4 ratio shift. The selected bacterial culture was split into two individual cultures and diluted 1:30 into fresh LB media containing kanamycin (0.4, 0.8, 1.2, or 1.6 mg/ml) or chloramphenicol (100 μg/ml). The process was repeated one more time before the resulting bacteria were analyzed for their R3 content.

To perform recording and device resetting using CAMERA 1.2, *E. coli* S1030 were transformed with 500 ng of R3, 250 ng of R4, and 100 ng of W1.1 and plated on LB agar containing kanamycin (25 μg/ml), chloramphenicol (10 μg/ml), and spectinomycin (100 μg/ml). A bacterial colony carrying CAMERA 1.2 with 36% R3 and 64% R4 was selected for further characterization and split into three independent cultures. To initiate the recording process, the bacterial culture was inoculated 1:30 into fresh LB media containing aTc (50 ng/ml) and 100 μM IPTG, whereas to reset the device, a similar inoculation protocol was performed with fresh LB media containing kanamycin (0.8 mg/ml). The inoculated culture was allowed to grow at 37°C with shaking for 12 to 24 hours to saturation. The process was repeated until a desired R3/R4 ratio was obtained. The R3 content was characterized by HTS analysis of the *EGFP* fragment amplified from the bacterial culture.

Characterization of CAMERAs 2.0 and 2.1 in *E. coli* S1030

E. coli S1030 were transformed with R1 and W2.0 and plated on LB agar containing carbenicillin

and spectinomycin. A single colony was picked and cultured at 37°C with shaking to obtain a dense culture as the starting material of the recording experiments. The split bacterial cultures were diluted 500- or 1000-fold into fresh LB media containing aTc (2, 20, or 200 ng/ml) and were grown in a 96-deep-well plate at 37°C with shaking for 16 to 24 hours before being diluted again. The process was repeated until 68 generations of bacteria were produced. Editing promoted by the BE2-sgRNA1 complex was characterized by amplifying the *EGFP* gene from the bacterial culture and analyzing the amplicon using HTS.

E. coli S1030 carrying CAMERA 2.1 were treated with (i) no inducer, (ii) 1 mM IPTG, or (iii) aTc (200 ng/ml) and 1, 0.1, or 0.01 mM IPTG. Similar repeated diluting and inducing protocol was adapted as that was used for CAMERA 2.0.

To confirm that CAMERA 2.0 could record the duration of a stimulus, *E. coli* S1030 cultures carrying CAMERA 2.0 were diluted 1000-fold into fresh LB media and treated with or without aTc (100 ng/ml). The bacteria were grown in a 24-deep-well plate at 37°C with shaking for 12 hours and diluted 1000-fold again into fresh LB containing the same concentrations of aTc. In the third dilution, bacteria that had not encountered the inducer were split into fresh LB media with or without aTc (100 ng/ml). The process was repeated once in the fourth dilution. Similarly, bacteria that had been treated with aTc were split and treated with or without aTc from generation 20 to 40. *E. coli* S1030 carrying CAMERA 2.1 were tested for IPTG sensing using a similar setup.

Recording in the genomic safe-harbor gene *CCR5* in human cells

HEK293T cells (GenTarget Inc.) were cultured in 48-well plates (collagen-coated, ~40,000 cells seeded per well) in DMEM plus GlutaMAX (Life Technologies) with 10% FBS. Transfection was performed 24 hours after plating when cells reached 60 to 70% confluence. To initiate recording in the human safe-harbor gene *CCR5*, 800 ng of BE3 plasmid and 40 ng of guide RNA plasmid (CAMERA 2m.0; see table S5 for guide RNA sequences) were transfected in each well using 1.2 µl of Lipofectamine 2000 (Life Technologies) following the manufacturer's protocol. To multiplex recording using multiple guide RNAs, each guide RNA plasmid was applied at a dose of 40 ng together with 800 ng of BE3 plasmid. The transfected cells were incubated for an additional 3 days before being harvested for genomic DNA extraction. Base editing was quantified by amplifying the *CCR5* gene fragment from genomic DNA by PCR and analyzing by HTS.

Recording Wnt signaling in the *CCR5* loci of human cells

To enable CAMERA 2m to record Wnt signaling, we installed a (TCF/LEF)₇ promoter upstream of BE3 and BE3-P2A-Luc to generate CAMERA 2m.3 [(TCF/LEF)₇-BE3 and (TCF/LEF)₇-BE3-P2A-Luc].

TOPFlash [(TCF/LEF)₇-Luc] (43) was used as a transient readout of Wnt signaling. A control plasmid that encodes the *Renilla* luciferase was included to normalize transfection efficiency for luminescence readout.

HEK293T cells were cultured in 96-well plates (collagen-coated, ~20,000 cells seeded per well) in DMEM plus GlutaMAX with 10% FBS. Transfection was performed 24 hours after plating when cells reached 60 to 70% confluence. CAMERA 2m.3 were prepared in 5 µl of reduced serum media (Opti-MEM, Life Technologies) with 200 ng of (TCF/LEF)₇-BE3 or (TCF/LEF)₇-BE3-P2A-Luc plasmids, 20 ng of U6-sgRNA B plasmid, and 10 ng of *Renilla* luciferase plasmid, and transfected using 0.5 µl of Lipofectamine 2000. TOPFlash plasmid (200 ng) was transfected using a similar setup without including the guide RNA plasmid. A stock solution of 1 M LiCl was prepared in ddH₂O and added to the media to a final concentration of 50 mM 10 hours after transfection.

Firefly luciferase and *Renilla* luciferase activities were measured 24 hours after LiCl treatment. Luciferase substrates were purchased from Promega. To characterize Wnt-stimulated base editing, we incubated the transfected cells for 3 days before harvesting for genomic DNA extraction. Base editing was quantified by amplifying the *CCR5* gene fragment from genomic DNA by PCR and analyzing by HTS.

REFERENCES AND NOTES

- J. Shendure, H. Ji, Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145 (2008). doi: [10.1038/nbt1486](https://doi.org/10.1038/nbt1486); pmid: [18846087](https://pubmed.ncbi.nlm.nih.gov/18846087/)
- J. M. Heather, B. Chain, The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1–8 (2016). doi: [10.1016/j.ygeno.2015.11.003](https://doi.org/10.1016/j.ygeno.2015.11.003); pmid: [26554401](https://pubmed.ncbi.nlm.nih.gov/26554401/)
- O. Purcell, T. K. Lu, Synthetic analog and digital circuits for cellular computation and memory. *Curr. Opin. Biotechnol.* **29**, 146–155 (2014). doi: [10.1016/j.copbio.2014.04.009](https://doi.org/10.1016/j.copbio.2014.04.009); pmid: [24794536](https://pubmed.ncbi.nlm.nih.gov/24794536/)
- M. C. Inniss, P. A. Silver, Building synthetic memory. *Curr. Biol.* **23**, R812–R816 (2013). doi: [10.1016/j.cub.2013.06.047](https://doi.org/10.1016/j.cub.2013.06.047); pmid: [24028965](https://pubmed.ncbi.nlm.nih.gov/24028965/)
- A. E. Friedland et al., Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009). doi: [10.1126/science.1172005](https://doi.org/10.1126/science.1172005); pmid: [19478183](https://pubmed.ncbi.nlm.nih.gov/19478183/)
- J. Bonnet, P. Subsoontorn, D. Endy, Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8884–8889 (2012). doi: [10.1073/pnas.1202344109](https://doi.org/10.1073/pnas.1202344109); pmid: [22615351](https://pubmed.ncbi.nlm.nih.gov/22615351/)
- L. Yang et al., Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014). doi: [10.1038/nmeth.3147](https://doi.org/10.1038/nmeth.3147); pmid: [25344638](https://pubmed.ncbi.nlm.nih.gov/25344638/)
- T. S. Ham, S. K. Lee, J. D. Keasling, A. P. Arkin, Design and construction of a double inversion recombination switch for heritable sequential genetic memory. *PLOS ONE* **3**, e2815 (2008). doi: [10.1371/journal.pone.0002815](https://doi.org/10.1371/journal.pone.0002815); pmid: [18665232](https://pubmed.ncbi.nlm.nih.gov/18665232/)
- J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, D. Endy, Amplifying genetic logic gates. *Science* **340**, 599–603 (2013). doi: [10.1126/science.1232758](https://doi.org/10.1126/science.1232758); pmid: [23539178](https://pubmed.ncbi.nlm.nih.gov/23539178/)
- F. Farzadfar, T. K. Lu, Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014). doi: [10.1126/science.1256272](https://doi.org/10.1126/science.1256272); pmid: [25395541](https://pubmed.ncbi.nlm.nih.gov/25395541/)
- M. Jinek et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012). doi: [10.1126/science.1225829](https://doi.org/10.1126/science.1225829); pmid: [22745249](https://pubmed.ncbi.nlm.nih.gov/22745249/)
- J. A. Doudna, E. Charpentier, The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014). doi: [10.1126/science.1258096](https://doi.org/10.1126/science.1258096); pmid: [25430774](https://pubmed.ncbi.nlm.nih.gov/25430774/)
- L. Cong et al., Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013). doi: [10.1126/science.1231143](https://doi.org/10.1126/science.1231143); pmid: [23287718](https://pubmed.ncbi.nlm.nih.gov/23287718/)
- P. Mali et al., RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013). doi: [10.1126/science.1232033](https://doi.org/10.1126/science.1232033)
- A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016). doi: [10.1038/nature17946](https://doi.org/10.1038/nature17946); pmid: [27096365](https://pubmed.ncbi.nlm.nih.gov/27096365/)
- K. Nishida et al., Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016). doi: [10.1126/science.aaf8729](https://doi.org/10.1126/science.aaf8729); pmid: [27492474](https://pubmed.ncbi.nlm.nih.gov/27492474/)
- L. A. Gilbert et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013). doi: [10.1016/j.cell.2013.06.044](https://doi.org/10.1016/j.cell.2013.06.044); pmid: [23849981](https://pubmed.ncbi.nlm.nih.gov/23849981/)
- R. Bowater, A. J. Doherty, Making ends meet: Repairing breaks in bacterial DNA by non-homologous end-joining. *PLOS Genet.* **2**, e8 (2006). doi: [10.1371/journal.pgen.0020008](https://doi.org/10.1371/journal.pgen.0020008); pmid: [16518468](https://pubmed.ncbi.nlm.nih.gov/16518468/)
- S. Ayora et al., Double-strand break repair in bacteria: A view from *Bacillus subtilis*. *FEMS Microbiol. Rev.* **35**, 1055–1081 (2011). doi: [10.1111/j.1574-6976.2011.00272.x](https://doi.org/10.1111/j.1574-6976.2011.00272.x); pmid: [21517913](https://pubmed.ncbi.nlm.nih.gov/21517913/)
- J. C. Carlson, A. H. Badran, D. A. Guggiana-Nilo, D. R. Liu, Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat. Chem. Biol.* **10**, 216–222 (2014). doi: [10.1038/nchembio.1453](https://doi.org/10.1038/nchembio.1453); pmid: [24487694](https://pubmed.ncbi.nlm.nih.gov/24487694/)
- V. Pattanayak et al., High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013). doi: [10.1038/nbt.2673](https://doi.org/10.1038/nbt.2673); pmid: [23934178](https://pubmed.ncbi.nlm.nih.gov/23934178/)
- G. L. Rosano, E. A. Ceccarelli, Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **5**, 172 (2014). doi: [10.3389/fmicb.2014.00172](https://doi.org/10.3389/fmicb.2014.00172); pmid: [24860555](https://pubmed.ncbi.nlm.nih.gov/24860555/)
- P. Siuti, J. Yazbek, T. K. Lu, Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013). doi: [10.1038/nbt.2510](https://doi.org/10.1038/nbt.2510); pmid: [23396014](https://pubmed.ncbi.nlm.nih.gov/23396014/)
- A. Lewendon, I. A. Murray, W. V. Shaw, M. R. Gibbs, A. G. Leslie, Replacement of catalytic histidine-195 of chloramphenicol acetyltransferase: Evidence for a general base role for glutamate. *Biochemistry* **33**, 1944–1950 (1994). doi: [10.1021/bi00173a043](https://doi.org/10.1021/bi00173a043); pmid: [7906544](https://pubmed.ncbi.nlm.nih.gov/7906544/)
- D. D. Boehr, P. R. Thompson, G. D. Wright, Molecular mechanism of aminoglycoside antibiotic kinase APH(3')-IIIa: Roles of conserved active site residues. *J. Biol. Chem.* **276**, 23929–23936 (2001). doi: [10.1074/jbc.M100540200](https://doi.org/10.1074/jbc.M100540200); pmid: [11279088](https://pubmed.ncbi.nlm.nih.gov/11279088/)
- Y. B. Kim et al., Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017). doi: [10.1038/nbt.3803](https://doi.org/10.1038/nbt.3803); pmid: [28191901](https://pubmed.ncbi.nlm.nih.gov/28191901/)
- N. M. Gaudelli et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017). doi: [10.1038/nature24644](https://doi.org/10.1038/nature24644); pmid: [29160308](https://pubmed.ncbi.nlm.nih.gov/29160308/)
- A. C. Komor et al., Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017). doi: [10.1038/nature24644](https://doi.org/10.1038/nature24644); pmid: [29160308](https://pubmed.ncbi.nlm.nih.gov/29160308/)
- L. Yang et al., Engineering and optimising deaminase fusions for genome editing. *Nat. Commun.* **7**, 13330 (2016). doi: [10.1038/ncomms13330](https://doi.org/10.1038/ncomms13330); pmid: [27804970](https://pubmed.ncbi.nlm.nih.gov/27804970/)
- H. A. Rees et al., Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017). doi: [10.1038/ncomms15790](https://doi.org/10.1038/ncomms15790); pmid: [28585549](https://pubmed.ncbi.nlm.nih.gov/28585549/)
- K. Kim et al., Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017). doi: [10.1038/nbt.3816](https://doi.org/10.1038/nbt.3816); pmid: [28244995](https://pubmed.ncbi.nlm.nih.gov/28244995/)
- C. Kescu et al., CRISPR-STOP: Gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017). doi: [10.1038/nmeth.4327](https://doi.org/10.1038/nmeth.4327); pmid: [28581493](https://pubmed.ncbi.nlm.nih.gov/28581493/)
- Z. Shimatani et al., Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017). doi: [10.1038/nbt.3833](https://doi.org/10.1038/nbt.3833); pmid: [28346401](https://pubmed.ncbi.nlm.nih.gov/28346401/)
- Y. Zong et al., Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017). doi: [10.1038/nbt.3811](https://doi.org/10.1038/nbt.3811); pmid: [28244994](https://pubmed.ncbi.nlm.nih.gov/28244994/)
- W. Xiong, J. E. Ferrell Jr., A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* **426**, 460–465 (2003). doi: [10.1038/nature02089](https://doi.org/10.1038/nature02089); pmid: [14647386](https://pubmed.ncbi.nlm.nih.gov/14647386/)

36. V. Hsiao, Y. Hori, P. W. Rothmund, R. M. Murray, A population-based temporal logic gate for timing and recording chemical events. *Mol. Syst. Biol.* **12**, 869 (2016). doi: [10.15252/msb.20156663](https://doi.org/10.15252/msb.20156663); pmid: [27193783](https://pubmed.ncbi.nlm.nih.gov/27193783/)
 37. L. Weiner, J. L. Brissette, P. Model, Stress-induced expression of the *Escherichia coli* phage shock protein operon is dependent on $\sigma 54$ and modulated by positive and negative feedback mechanisms. *Genes Dev.* **5**, 1912–1923 (1991). doi: [10.1101/gad.5.10.1912](https://doi.org/10.1101/gad.5.10.1912); pmid: [1717346](https://pubmed.ncbi.nlm.nih.gov/1717346/)
 38. G. Rowley, M. Spector, J. Kormanec, M. Roberts, Pushing the envelope: Extracytoplasmic stress responses in bacterial pathogens. *Nat. Rev. Microbiol.* **4**, 383–394 (2006). doi: [10.1038/nrmicro1394](https://doi.org/10.1038/nrmicro1394); pmid: [16715050](https://pubmed.ncbi.nlm.nih.gov/16715050/)
 39. B. P. Hubbard *et al.*, Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods* **12**, 939–942 (2015). doi: [10.1038/nmeth.3515](https://doi.org/10.1038/nmeth.3515); pmid: [26258293](https://pubmed.ncbi.nlm.nih.gov/26258293/)
 40. R. Ohlendorf, R. R. Vidavski, A. Eldar, K. Moffat, A. Möglich, From dusk till dawn: One-plasmid systems for light-regulated gene expression. *J. Mol. Biol.* **416**, 534–542 (2012). doi: [10.1016/j.jmb.2012.01.001](https://doi.org/10.1016/j.jmb.2012.01.001); pmid: [22245580](https://pubmed.ncbi.nlm.nih.gov/22245580/)
 41. M. Sadelain, E. P. Papapetrou, F. D. Bushman, Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51–58 (2011). doi: [10.1038/nrc3179](https://doi.org/10.1038/nrc3179); pmid: [22129804](https://pubmed.ncbi.nlm.nih.gov/22129804/)
 42. H. Clevers, R. Nusse, Wnt/ β -catenin signaling and disease. *Cell* **149**, 1192–1205 (2012). doi: [10.1016/j.cell.2012.05.012](https://doi.org/10.1016/j.cell.2012.05.012); pmid: [22682243](https://pubmed.ncbi.nlm.nih.gov/22682243/)
 43. V. Korinek *et al.*, Constitutive transcriptional activation by a β -catenin-Tcf complex in APC^{-/-} colon carcinoma. *Science* **275**, 1784–1787 (1997). doi: [10.1126/science.275.5307.1784](https://doi.org/10.1126/science.275.5307.1784); pmid: [9065401](https://pubmed.ncbi.nlm.nih.gov/9065401/)
 44. C. M. Hedgepeth *et al.*, Activation of the Wnt signaling pathway: A molecular mechanism for lithium action. *Dev. Biol.* **185**, 82–91 (1997). doi: [10.1006/dbio.1997.8552](https://doi.org/10.1006/dbio.1997.8552); pmid: [9169052](https://pubmed.ncbi.nlm.nih.gov/9169052/)
 45. Z. Toman, C. Dambly-Chaudière, L. Tenenbaum, M. Radman, A system for detection of genetic and epigenetic alterations in *Escherichia coli* induced by DNA-damaging agents. *J. Mol. Biol.* **186**, 97–105 (1985). doi: [10.1016/0022-2836\(85\)90260-8](https://doi.org/10.1016/0022-2836(85)90260-8); pmid: [3935801](https://pubmed.ncbi.nlm.nih.gov/3935801/)
 46. A. McKenna *et al.*, Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016). doi: [10.1126/science.aaf7907](https://doi.org/10.1126/science.aaf7907)
 47. K. L. Frieda *et al.*, Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017). doi: [10.1038/nature20777](https://doi.org/10.1038/nature20777); pmid: [27869821](https://pubmed.ncbi.nlm.nih.gov/27869821/)
 48. S. D. Perli, C. H. Cui, T. K. Lu, Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511 (2016). doi: [10.1126/science.aag0511](https://doi.org/10.1126/science.aag0511); pmid: [27540006](https://pubmed.ncbi.nlm.nih.gov/27540006/)
 49. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016). doi: [10.1126/science.aaf1175](https://doi.org/10.1126/science.aaf1175)
 50. R. U. Sheth, S. S. Yim, F. L. Wu, H. H. Wang, Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017). doi: [10.1126/science.aao0958](https://doi.org/10.1126/science.aao0958); pmid: [29170279](https://pubmed.ncbi.nlm.nih.gov/29170279/)
- ACKNOWLEDGMENTS**
- We thank H. A. Rees for providing Cas9 protein for the in vitro cleavage assay. The *E. coli* S2063 strain was generated by A. H. Badran by engineering a previously reported strain S1030. We thank M. S. Packer, F. Xiong, J. M. Levy, and J. H. Hu for providing and revising the HTS analysis scripts, L. Wang for modeling the behavior of CAMERA 2.4, and W. Ma for helpful discussions on Wnt signaling. Flow sorting was performed by J. Nelson at the Bauer Core facility of Harvard University. **Funding:** Supported by NIH grants RM1 HG009490, R01 EB022376, and R35 GM118062, DARPA grant HR0011-17-2-0049, and the Howard Hughes Medical Institute. W.T. is an HHMI Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. **Author contributions:** W.T. and D.R.L. designed the research; W.T. prepared materials and performed experiments; W.T. and D.R.L. wrote the manuscript. **Competing interests:** The authors have filed a patent application on aspects of this work. D.R.L. is a consultant and co-founder of Editas Medicine, Beam Therapeutics, and Pairwise Plants, companies that are using genome editing technologies. **Data and materials availability:** Plasmids constructed in this study are available from Addgene. High-throughput sequencing data are available from the NCBI Sequence Read Archive database, accession number SRP132318.
- SUPPLEMENTARY MATERIALS**
- www.sciencemag.org/content/360/6385/eaap8992/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S6
References (51, 52)
- 7 September 2017; resubmitted 22 December 2017
Accepted 8 February 2018
Published online 15 February 2018
[10.1126/science.aap8992](https://doi.org/10.1126/science.aap8992)

RESEARCH ARTICLE SUMMARY

CELL BIOLOGY

MitoCPR—A surveillance pathway that protects mitochondria in response to protein import stress

Hilla Weidberg* and Angelika Amon*

INTRODUCTION: Mitochondria provide cells with energy and numerous essential metabolites such as lipids, amino acids, iron sulfur clusters, and heme. All mitochondrial functions rely on import of proteins into the organelle because the mitochondrial proteome is almost exclusively encoded by nuclear genes. Given the central importance of mitochondria for cell viability, it is not surprising that cells mount a nuclear response when mitochondrial functions are compromised. These mitochondria-to-nucleus signaling pathways include the mtUPR (mitochondrial unfolded protein response), which triggers expression of mitochondrial chaperones when mitochondrial protein folding is defective, and the UPRam (unfolded protein response activated by mistargeting of

proteins) and mPOS (mitochondrial precursor over-accumulation stress) pathways, which reduce translation and induce degradation of unimported proteins in the cytosol when mitochondrial import is impaired. Even though mitochondrial import is central to all mitochondrial functions, no response to protein import defects had been described that protects mitochondria during this stress.

RATIONALE: To determine how cells respond to defects in mitochondrial protein import, we first developed a system in budding yeast with which to specifically inhibit this process. We found that overexpression of proteins that rely on a bipartite signal sequence for their mitochondrial localization inhibited mitochondrial

import and led to the accumulation of mitochondrial precursors. Protease protection and carbonate extraction assays that were performed on isolated mitochondria revealed that these unimported proteins accumulated on the mitochondrial surface and in the import channel known as the translocase.

RESULTS: Having developed a system that allowed us to specifically inhibit mitochondrial protein import, we examined the cellular response to this defect. Transcriptome analysis of cells overexpressing bipartite signal-containing proteins identified a gene expression pattern related to the multi-drug resistance response. We termed this response mitochondrial compromised protein import response (mitoCPR).

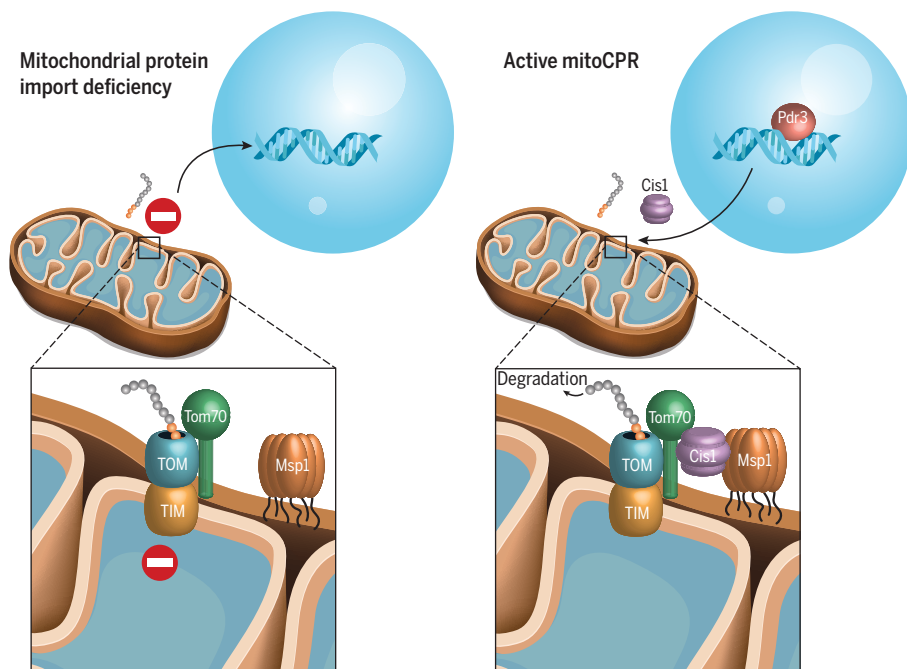
mitoCPR was triggered by protein import defects but not other mitochondrial deficiencies, such as respiratory failure, and was mediated by the transcription factor Pdr3. Our analyses further showed that mitoCPR was critical for the protection of mitochondria during import stress. Cells lacking *PDR3* did not mount a mitoCPR during import stress and accumulated higher levels of unimported proteins on the organelle surface as compared with those of wild-type cells. Consequently, *pdr3Δ* cells exhibited decreased respiratory function and loss of mitochondrial DNA when mitochondrial import was restored. Our results also shed light on the mechanism by which mitoCPR protected mitochondria. Upon mitochondrial import stress, Pdr3 induced expression of *Cis1*. Coimmunoprecipitation analyses showed that *Cis1* recruited the AAA⁺ adenosine triphosphatase Msp1 to the translocase by binding to the translocase receptor Tom70. There, the two proteins mediated the clearance and proteasomal degradation of proteins that failed to be imported into mitochondria.

CONCLUSION: We discovered a mitochondrial import surveillance mechanism in budding yeast. This surveillance mechanism, mitoCPR, is activated when mitochondrial import is stalled in order to induce the removal of mitochondrial proteins accumulating on the mitochondrial surface. Clearance of precursors is critical for maintaining mitochondrial functions during import stress. We propose that mitoCPR could be especially important when the import machinery is overwhelmed, as may occur in situations that require the rapid expansion of the mitochondrial compartment. ■

David H. Koch Institute for Integrative Cancer Research, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*Corresponding author. Email: angelika@mit.edu (A.A.); hillaw@mit.edu (H.W.)

Cite this article as H. Weidberg, A. Amon, *Science* 360, eaan4146 (2018). DOI: 10.1126/science.aan4146



MitoCPR protects mitochondria during import stress. (Left) Mitochondrial protein import deficiency leads to the accumulation of mitochondrial proteins on the organelle's surface and in the translocases. (Right) Pdr3 induces *CIS1* expression. *Cis1* binds to the mitochondrial import receptor Tom70 and recruits Msp1 to mediate clearance of unimported precursors from the mitochondrial surface and their proteasomal degradation. This protects mitochondrial functions during import stress.

RESEARCH ARTICLE

CELL BIOLOGY

MitoCPR—A surveillance pathway that protects mitochondria in response to protein import stress

Hilla Weidberg* and Angelika Amon*

Mitochondrial functions are essential for cell viability and rely on protein import into the organelle. Various disease and stress conditions can lead to mitochondrial import defects. We found that inhibition of mitochondrial import in budding yeast activated a surveillance mechanism, mitoCPR, that improved mitochondrial import and protected mitochondria during import stress. mitoCPR induced expression of *Cis1*, which associated with the mitochondrial translocase to reduce the accumulation of mitochondrial precursor proteins at the mitochondrial translocase. Clearance of precursor proteins depended on the *Cis1*-interacting AAA⁺ adenosine triphosphatase *Msp1* and the proteasome, suggesting that *Cis1* facilitates degradation of unimported proteins. mitoCPR was required for maintaining mitochondrial functions when protein import was compromised, demonstrating the importance of mitoCPR in protecting the mitochondrial compartment.

Mitochondrial function is required for cell viability, producing energy, and many essential biological molecules such as iron-sulfur clusters and heme (1). Even though mitochondria contain their own genome, the vast majority of their proteins are encoded by the nucleus. Import of nuclear-encoded proteins into mitochondria is essential for mitochondrial function and cell viability (1, 2). Defects in mitochondrial protein import are associated with various human diseases, such as deafness-dystonia syndrome and Huntington's disease (3–5). However, even though mitochondrial protein import is essential for all mitochondrial functions, little is known about how cells respond to mitochondrial protein import defects. Recently, two pathways—mPOS (mitochondrial precursor over-accumulation stress) and UPRam (unfolded protein response activated by mistargeting of proteins)—have been identified in yeast that respond to the accumulation of unimported mitochondrial proteins in the cytosol (6, 7). UPRam and mPOS reduce global translation, and UPRam protects the cytosol from proteotoxic effects of unimported proteins by accelerating their degradation. In mammals, the Ubiquitin family of proteins has a similar role in mediating the degradation of mitochondrial transmembrane proteins that fail to get imported and remain in the cytosol (8). Whether mechanisms exist that protect mitochondrial functions in the face of mitochondrial import stress is unclear. We identified a response to mitochondrial protein import defects that protected mitochondrial functions by reducing the accumulation of

precursor proteins at the mitochondrial surface and translocase. This response was brought about by the transcription factor *PDR3*, which has previously been shown to mediate a multidrug resistance (MDR) response.

The canonical MDR response is conserved from bacteria to mammals (9). It protects organisms from xenobiotics and can limit the effectiveness of microbial and cancer chemotherapy (9, 10). In budding yeast, the MDR response is activated by a variety of chemical compounds and is primarily mediated by the two related transcription factors *Pdr1* and *Pdr3* (11–13). They induce the expression of several adenosine 5'-triphosphate (ATP)-binding cassette (ABC) transporters to mediate efflux of xenobiotics (13). A transcriptional response related to the MDR and specifically mediated by *Pdr3* is active in yeast cells with defective mitochondrial DNA (mtDNA) (14). In such cells, *Pdr3* induces the expression of genes encoding ABC transporters, sphingolipid biosynthesis enzymes, and a number of genes of unknown function (15). We show here that *Pdr3* mediates a mitochondrial import defect response.

A system to acutely inhibit mitochondrial protein import

All mitochondrial functions depend on proteins being imported from the cytosol into the organelle. Whether pathways exist that monitor import of proteins into mitochondria and elicit a cellular response under conditions of mitochondrial import stress is unknown. To determine whether cells respond to mitochondrial import stress, we examined the consequences of acutely interfering with mitochondrial protein import. Compounds that uncouple the mitochondrial respiratory chain, such as CCCP (carbonyl cyanide *m*-chlorophenyl hydrazine), prevent mito-

chondrial import, which is dependent on the mitochondrial membrane potential (2). However, these same compounds can also affect potential across other cellular membranes and induce a MDR response, which complicates delineating responses specific to mitochondrial import defects. We hypothesized that acute induction of mitochondrial import stress could be achieved without drugs by overloading the mitochondrial import machinery through overexpression of mitochondrial proteins. We overexpressed a number of mitochondrial proteins from the strong galactose-inducible *GALI-10* promoter and assessed the mitochondrial import of Cox5a, a nuclear-encoded subunit of mitochondrial complex IV. Like most mitochondrial proteins, Cox5a harbors an N-terminal presequence that is cleaved upon import into the mitochondrial matrix (16). In untreated cells, mitochondrial import and precursor cleavage was so efficient that the Cox5a preprotein (Cox5a^{pre}) was not detected (Fig. 1A). Upon disruption of membrane potential and hence protein import with CCCP, Cox5a^{pre} accumulated in cells (Fig. 1A).

Overexpression of the majority of mitochondrial proteins did not affect Cox5a processing, but high levels of *Psd1*, *Ccp1*, *Cyb2*, Cox5a, or *Tim50* led to Cox5a^{pre} accumulation (Fig. 1A). All five proteins use the same mitochondrial import machinery. They contain a bipartite signal that inhibits translocation into the mitochondrial matrix. This results in the lateral release of proteins out of the inner-membrane translocase TIM23 into the inner membrane itself (2). A broad survey of mitochondrial proteins known to contain a bipartite signal confirmed this conclusion (Fig. 1B). By contrast, inner-membrane proteins that use other import mechanisms (for example, the TIM22 pathway) or proteins that translocate across the TIM23 translocase, such as matrix proteins, did not affect Cox5a processing (Fig. 1A). Overexpression of bipartite signal-containing proteins affected import of proteins other than Cox5a. High levels of the bipartite signal-containing protein *Psd1* interfered with the processing of a number of presequence-containing proteins whose import is mediated by the TIM23 complex (Fig. 1C). Thus, saturation of the TIM23 lateral diffusion import pathway leads to the accumulation of mitochondrial preproteins.

The accumulation of mitochondrial preproteins could reflect defects in either translocation into mitochondria or presequence cleavage in the matrix. To test the former possibility, we determined the localization of Cox5a^{pre}. Both the mature and the preprotein forms of Cox5a were detected in mitochondrial but not cytosolic fractions after overexpression of *PSD1* or CCCP treatment (Fig. 2A). Addition of proteinase K to the mitochondrial fractions led to loss of Cox5a^{pre} but not mature Cox5a, which resides in the inner membrane with its C terminus facing the intermembrane space. Because Cox5a was detected by using a C-terminal V5 tag in this analysis, we conclude that at least the C terminus of Cox5a^{pre} resides at the surface of mitochondria that faces the cytosol. These results lead to two important

David H. Koch Institute for Integrative Cancer Research, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*Corresponding author. Email: angelika@mit.edu (A.A.); hillaw@mit.edu (H.W.)

conclusions. First, overexpressed bipartite signal-containing proteins interfere with mitochondrial protein translocation. Second, the C-terminus of Cox5a^{pre} accumulates at the mitochondrial surface when mitochondrial import is impaired.

Cox5a^{pre} could be peripherally associated with the mitochondrial outer membrane by binding to receptors on the mitochondrial surface, be trapped in the translocase, or be incorrectly inserted into the outer membrane via its transmembrane domain. To determine the exact localization of Cox5a^{pre}, we treated mitochondria preparations with sodium carbonate (pH 11), which extracts peripheral membrane proteins from membranes (17). As expected, the inner-membrane localized, mature Cox5a was largely resistant to sodium carbonate extraction, while the peripheral outer membrane protein Cisl dissociated from mitochondria during this treatment (Fig. 2B). Most of Cox5a^{pre} remained associated with mitochondrial membranes during sodium carbonate treatment, indicating that a large fraction of Cox5a^{pre} was either inappropriately integrated into the outer membrane or stalled in the TOM (translocase of the outer membrane) translocase (Fig. 2B). To distinguish between these possibilities, we investigated the localization of Sod2, a mitochondrial matrix protein that lacks any transmembrane domains. Like Cox5a^{pre}, Sod2^{pre} accumulated at the mitochondrial outer membrane after overexpression of *PSD1*; association of the precursor with mitochondrial fractions was sensitive to proteinase K treatment (fig. S1). Sod2^{pre} was also largely resistant to sodium carbonate extraction (Fig. 2C). By contrast, sodium carbonate treatment solubilized mature matrix-localized Sod2. Thus, during import stress, mitochondrial preproteins are tightly bound to the mitochondrial

outer membrane independently of transmembrane domains. This suggests that at least a fraction of the preproteins is stalled in the mitochondrial translocase during import stress.

How do bipartite signal-containing proteins interfere with protein import when overexpressed? To address this question, we determined which bipartite signal element interfered with protein import when overexpressed. Bipartite mitochondrial targeting signals comprise a mitochondrial targeting sequence (MTS) and a hydrophobic segment that directs the protein to the inner membrane. Overexpressed Psd1 lacking its MTS did not inhibit mitochondrial protein import, demonstrating that Psd1 must be imported into mitochondria to interfere with Cox5a import (Fig. 2D). Consistent with this conclusion, Psd1's bipartite signal was sufficient to inhibit Cox5a mitochondrial import, whereas a fusion between only Psd1's MTS and GFP did not (Fig. 2D). Thus, when present in excess, bipartite signal-containing proteins interfere with import only when targeted to the inner membrane. This finding indicates that lateral diffusion out of TIM23 translocase is a rate-limiting step in mitochondrial import that can be saturated by overexpressing proteins imported via this route.

Mitochondrial import defects activate the mitoCPR

Does inhibiting protein import elicit a cellular response? To address this question, we examined the transcriptional consequences of overexpressing *PSD1*. Overexpression of *PSD1* up-regulated 217 genes and down-regulated 11 genes by two-fold or more (table S1). Among the up-regulated

genes was a group of genes previously shown to be induced by the transcription factor Pdr3, but not its close homolog Pdr1, in response to *PSD1* overexpression and loss of mtDNA (14, 18). We identified 19 genes whose induction upon mitochondrial import stress depended on *PDR3* (Fig. 3, A and B, and table S1). This group of genes included MDR response genes such as genes encoding ABC transporters, proteins involved in lipid metabolism and transport, reduced nicotinamide adenine dinucleotide phosphate (NADPH)-dependent enzymes, and a number of proteins of unknown function (Fig. 3A). Other, well-characterized mitochondrial stress responses were, however, not activated by *PSD1* overexpression within the time frame of the experiment. *PSD1*-overexpressing cells did not induce *RTG* (retrograde)-regulated genes, such as *CIT2* and *PDH1*, that are known to be activated in response to defects in Krebs cycle function (table S1) (19). The finding that overexpression of *PSD1* inhibited mitochondrial import suggests that it is mitochondrial import defects that elicit this *PDR3*-mediated transcriptional response. The finding that cells lacking mtDNA, which exhibit severe mitochondrial import defects (20, 21), also show that this transcriptional response is consistent with this idea.

To further explore a potential link between the *PDR3*-mediated transcriptional response and mitochondrial import defects, we first asked whether proteins—which, when overexpressed, inhibited mitochondrial import—also induced the *PDR3*-mediated transcriptional response. This was the case. All mitochondrial proteins that caused protein import defects when overexpressed induced the *PDR3*-mediated transcriptional response as determined by up-regulation of the *PDR3*-responsive

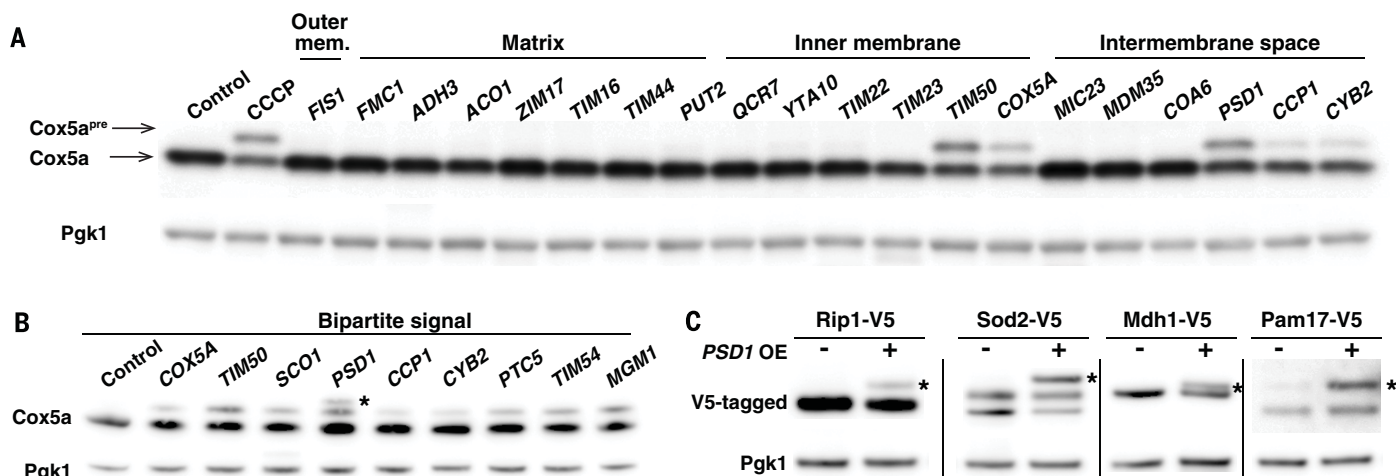


Fig. 1. Overexpression of bipartite signal-containing proteins induces mitochondrial protein import defects. (A) Immunoblot of Cox5a-V5 and Cox5a^{pre}-V5 (Cox5a preprotein) in control cells, CCCP-treated cells (20 μ M, 1 hour), or cells overexpressing mitochondrial proteins through the addition of galactose for 4 hours. Overexpressed proteins are divided by their localization to the outer membrane (outer mem.), matrix, inner membrane, and intermembrane space compartments. Pgk1 was used as a loading control. (B) Same as (A).

Asterisk represents a nonspecific band, result of *PSD1* overexpression. (C) Immunoblot of Rip1-V5, Sod2-V5, Mdh1-V5, and Pam17-V5 (expressed from their endogenous promoter) in control cells or after overexpression of *PSD1* for 4 hours. Asterisks identify the precursor form of the indicated proteins. OE, overexpression. As previously shown (50), Sod2 migrates in SDS-polyacrylamide gel electrophoresis (PAGE) as a doublet under conditions when mitochondria are intact and as a triplet when its cleavage is inhibited.

gene *CIS1*. Conversely, proteins whose overexpression did not interfere with mitochondrial import did not induce *CIS1* (Fig. 3C and fig. S2, A and B). The perfect correlation between the ability to inhibit mitochondrial import and induction of a *PDR3*-mediated transcriptional response was also observed when analyzing cells overexpressing various *PSD1* domains. Cells overexpressing *Psdl* that lacked its mitochondrial targeting signal or that harbored an N-terminal V5 tag to prevent targeting of the protein to mitochondria failed to induce *CIS1* (Fig. 3, D and E, and fig. S2, C and D) or any other *PDR3*-mediated transcripts (table S1). By contrast, GFP that was fused to the complete *Psdl* bipartite signal induced *CIS1* when overex-

pressed, whereas GFP that was fused only to *Psdl*'s MTS did not (Fig. 3D and fig. S2C).

The *PDR3*-mediated transcriptional response was not only induced through acute induction of mitochondrial import defects but was also seen in mutants in which mitochondrial import was constitutively impaired. Cells harboring deletions in mtDNA (*rho*⁻ cells) or lacking mtDNA (*rho0* cells), both of which cause mitochondrial import defects, expressed *CIS1* at an elevated level (Fig. 3F) (22). Cells lacking *TAM41*, a gene encoding a cardiolipin biosynthesis enzyme, have severe mitochondrial import defects but intact mtDNA (23, 24). These cells, too, expressed *CIS1* at high levels (Fig. 3F). Not all mitochondrial defects elicited

the *PDR3*-mediated transcriptional response. Deletion of genes encoding subunits of respiration complexes III and IV results in respiration defects (25, 26) but did not cause induction of *CIS1* expression (fig. S2E). Our results reveal a tight correlation between mitochondrial import defects and induction of a *PDR3*-mediated transcriptional response.

To further test the hypothesis that induction of the *PDR3*-mediated transcriptional response is caused by mitochondrial import defects, we examined the consequences of suppressing mitochondrial import defects on *CIS1* expression. The *ATP1-III* allele increases membrane potential and improves protein import in *rho0* cells by altering the ATP:ADP (adenosine 5'-diphosphate) ratio between the matrix and the intermembrane space (20, 21, 27). Introduction of the *ATP1-III* allele into either *rho0*, *rho*⁻, or *tam41Δ* cells caused a large decrease in *CIS1* expression (Fig. 3G). Thus, either defects in membrane potential or import defects elicit a *PDR3*-mediated transcriptional response. The finding that overexpression of *PSD1* for 4 hours, which is sufficient to induce the *PDR3*-mediated transcriptional response, did not significantly affect mitochondrial membrane potential (Fig. 3B and fig. S2F) suggested that membrane potential defects do not lead to induction of *Pdr3* target genes. Thus, mitochondrial import defects cause a *PDR3*-mediated transcriptional response. We termed this response mitoCPR for mitochondrial compromised protein import response.

The mitoCPR protects mitochondrial functions during import stress.

What is the role of the mitoCPR when mitochondrial protein import is impaired? To address this question, we first determined the consequences of deleting *PDR3* on the fate of *Cox5a*^{pre} under conditions in which protein import is impaired. As shown above, overexpression of *PSD1* led to the accumulation of *Cox5a*^{pre} (Fig. 1A). *Cox5a*^{pre} had a half-life of ~19 min in *PSD1*-overexpressing cells (Fig. 4, A and B). The eventual loss of *Cox5a*^{pre} in *PSD1*-overexpressing cells could be due to import of the preprotein into mitochondria, cytosolic degradation, or both. Deletion of *PDR3* prolonged the half-life of *Cox5a*^{pre} (Fig. 4, A and B). Conversely, overexpression of *PDR3* partially suppressed the accumulation of *Cox5a*^{pre} under conditions of mitochondrial import stress (Fig. 4C). Thus, *PDR3* and by extension mitoCPR are critical for either maintaining some level of mitochondrial import and/or clearing preproteins from the mitochondrial import machinery during import stress.

Next, we determined whether mitoCPR was important for maintaining mitochondrial functions under conditions of import stress. Upon overexpression of *PSD1*, oxygen consumption rate decreased (Fig. 4D and fig. S3A). Deletion of *PDR3* further exaggerated this effect (Fig. 4D and fig. S3A), indicating that *PDR3* is critical for maintaining mitochondrial respiration when mitochondrial import is compromised.

PDR1 and *PDR3* prevent mtDNA loss resulting from mitochondrial fusion defects (28). We tested

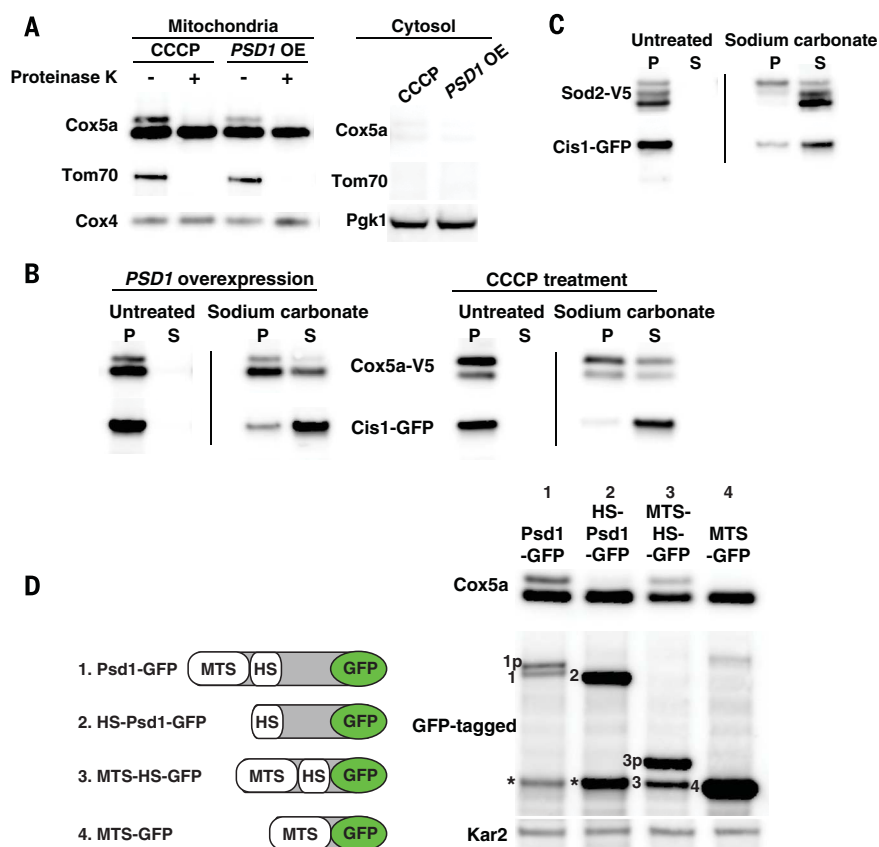


Fig. 2. Mitochondrial precursors accumulate on the surface of the organelle and in the translocase during import stress. (A) Mitochondria were isolated by means of differential centrifugation from cells treated with 20 μM CCCP for 1 hour or cells overexpressing *PSD1* for 6 hours. Cox5a-V5, Tom70-mCherry, and Cox4 or Pgk1 were detected in mitochondria and cytosol fractions. Mitochondria were treated with 50 μg/ml of proteinase K. Tom70 served as an outer-membrane control protein; Cox4 served as a matrix control protein. OE, overexpression. (B) Mitochondria were isolated from cells expressing *COX5a*-V5 and *CIS1*-GFP and overexpressing *PSD1* for 6 hours or cells treated with 20 μM CCCP for 1 hour. Sodium carbonate-treated or -untreated mitochondria were centrifuged so as to separate insoluble proteins [pellet (P)] from soluble proteins [supernatant (S)]. Samples were analyzed by means of immunoblot analysis. Cis1-GFP served as a peripheral outer-membrane protein control. (C) Mitochondria were isolated from cells overexpressing *PSD1* for 6 hours. Mitochondria were treated as in (B) in order to analyze Sod2-V5 by means of immunoblot analysis. (D) (Left) *PSD1*-GFP constructs used in the analysis. MTS, mitochondrial targeting sequence; HS, hydrophobic segment. (Right) Immunoblot blot of Cox5a-V5 and Psdl-GFP fusion proteins after overexpression of *PSD1*-GFP fusion genes for 4 hours. Kar2 was used as a loading control. Numbering on the immunoblot indicates the mature form of the GFP-tagged proteins. The letter "p" following this number identifies the precursor form of proteins. Asterisks identify a proteolytic cleavage product of Psdl known as the α subunit (51).

whether mitoCPR was important for protecting cells from mtDNA loss during import stress. Respiratory competence is a readout of mtDNA integrity. Assaying respiration, however, requires the analysis of colonies. This prerequisite precluded us from inducing mitochondrial import stress through overexpression of *PSD1* because prolonged overexpression of *PSD1* is lethal (fig. S3B). In fact, overexpression of all bipartite signal-containing proteins is lethal (fig. S3C). The mitochondrial Hsp70 chaperone *Ssc1* and its cochaperone *Mge1* are essential for mitochondrial import (29–31). We hypothesized that overexpression of *SSC1* or *MGE1* alone would lead to a mitochondrial import defect because the proper ratio of Hsp70 to its cochaperone is crucial for its chaperone activity in bacteria (32). Overexpression of *MGE1*, although not lethal (fig. S3B), caused a mild protein import defect comparable with that of cells lacking mtDNA. *Cox5a*^{P_{trc}} did not accumulate in *MGE1*-overexpressing cells or *rho0* cells upon induction of *Cox5a* expression from the methionine-regulated promoter (*MET25*) (fig. S3D). Nevertheless, mature *Cox5a* levels were reduced in *GAL-MGE1* and *rho0* cells as compared with control cells, whereas *COX5a* mRNA expression was comparable in all strains (fig. S3, D and E). Thus, less *Cox5a* is imported into mitochondria in *MGE1*-overexpressing cells, and unimported *Cox5a*^{P_{trc}} is rapidly degraded. Consistent with a mild mitochondrial import defect, overexpression of *MGE1* induced a mitoCPR as determined by elevated *CIS1* levels (as did overexpression of *SSC1*) (fig. S3, F and G).

Having established that overexpression of *MGE1* causes a mild mitochondrial import defect that is not lethal, we examined its effects on mtDNA stability. Overexpression of *MGE1* for 24 hours led to an increase in *rho*[−] cells (Fig. 4E). Inactivation of mitoCPR by deleting *PDR3* caused a threefold increase in cells harboring defective mtDNA (Fig. 4E). Because maintenance of mtDNA largely depends on nuclear-encoded genes (33), we conclude that mitochondrial import stress prevented their import. This caused mtDNA damage and the generation of *rho*[−] cells. Furthermore, the mitoCPR protects mtDNA only during import stress. The absence of *PDR3* did not affect respiration or mtDNA maintenance under normal growth conditions. Thus, mitoCPR has a protective role specifically during mitochondrial import stress.

Cis1 protects mitochondria during import stress

One of the most strongly induced genes after mitochondrial import stress is *CIS1* (Fig. 3A) (34). *CIS1* overexpression improves cellular fitness in the presence of citrinin, a mycotoxin that reduces mitochondrial membrane potential (35). The protein itself, however, neither harbors domains with known functions nor has homologs in higher eukaryotes. Cis1 protein only accumulated under conditions of mitochondrial import stress and was unstable even when expressed (fig. S4, A and B). Cis1 associates with mitochondria in high-throughput localization studies (36), which promp-

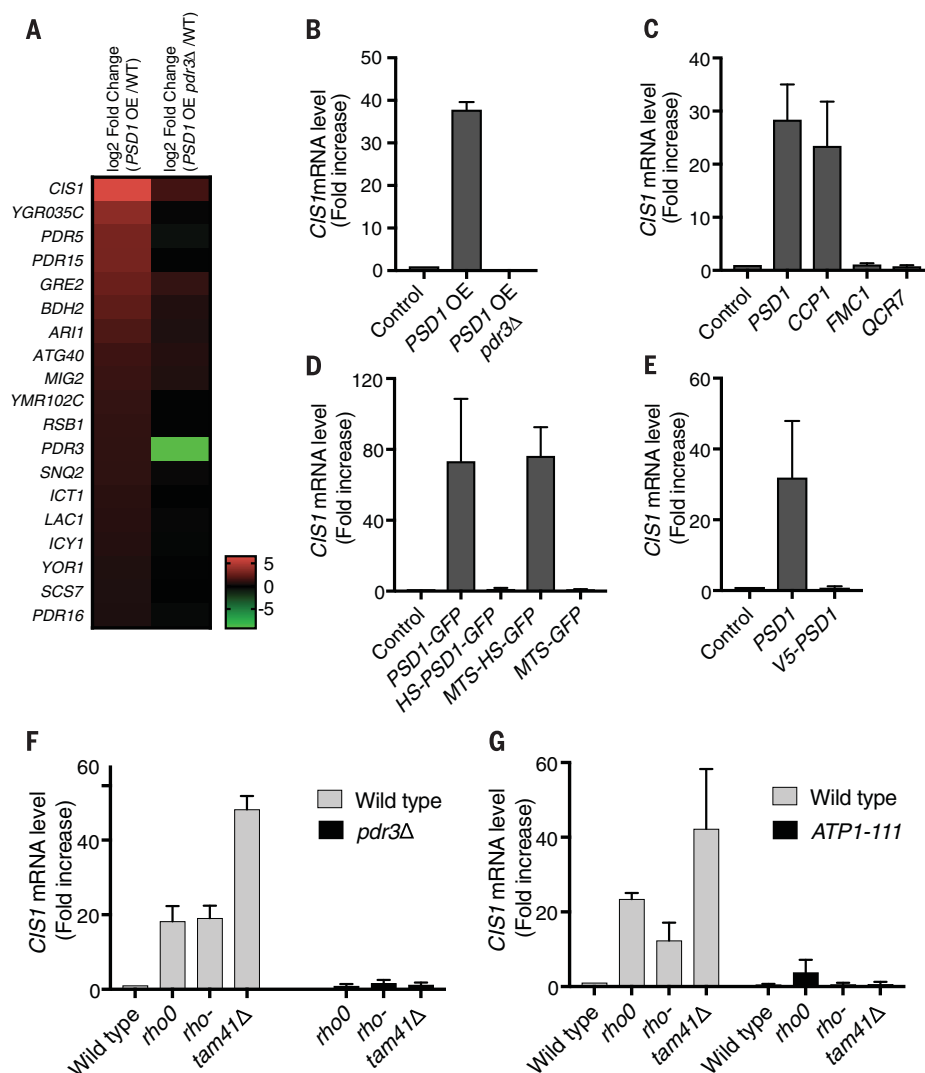


Fig. 3. Inhibition of mitochondrial protein import induces the mitoCPR. (A) Gene expression analysis of control wild-type cells and wild-type or *pdr3Δ* cells that overexpressed *PSD1* for 4 hours through galactose induction. The heat map describes the transcription profiles of cells overexpressing *PSD1* and *pdr3Δ* cells overexpressing *PSD1*. The 19 genes shown met the following criteria: (i) genes that exhibited an increase in expression of at least (log2) 0.6 and adjusted *P* values that are equal to or lower than 0.05 in *PSD1*-overexpressed cells versus *PSD1*-overexpressed cells lacking *PDR3*; (ii) genes that exhibited an increase in expression of at least (log2) 0.6 and adjusted *P* values that are equal to or lower than 0.05 in *PSD1*-overexpressed cells versus control cells. WT, wild type. OE, overexpression. (B) *CIS1* mRNA levels in wild-type cells, cells overexpressing *PSD1*, or *pdr3Δ* cells overexpressing *PSD1*. *PSD1* expression was induced through the addition of galactose for 4 hours. *n* = 3 experiments; data are mean ± SD. (C) *CIS1* mRNA levels in control cells or cells overexpressing mitochondrial proteins through galactose induction (4 hours) were analyzed by means of quantitative reverse transcription polymerase chain reaction (RT-PCR). *n* = 3 experiments; data are mean ± SD. (D) Same as (C), after overexpression of *PSD1*-GFP fusion genes for 4 hours. MTS, mitochondrial targeting sequence; HS, hydrophobic segment. *n* = 3 experiments; data are mean ± SD. (E) Same as (C), after overexpression of *PSD1* or V5-*PSD1* for 4 hours. *n* = 3 experiments; data are mean ± SD. (F) *CIS1* mRNA levels of wild-type, *rho0*, *rho*[−], and *tam41Δ* cells in the presence or absence of *PDR3*. *n* = 3 experiments; data are mean ± SD. (G) Same as (F), in the presence or absence of the *ATP1-111* allele. *n* = 3 experiments; data are mean ± SD.

ted us to investigate whether the protein played a role in protecting mitochondria during import stress. To study Cis1, we placed the gene under the constitutive *TEF2* promoter (fig. S4A). A constitutively expressed Cis1-GFP fusion indeed pre-

dominantly localized to the outer membrane of the organelle (Fig. 5, A to C). Cis1 is not predicted to have a transmembrane domain. We conclude that Cis1 associates with the outer mitochondrial membrane facing the cytosol.

The expression of *Cis1* proved to be important for cells during mitochondrial protein import stress. Although deletion of *CIS1* did not have a noticeable effect on *Cox5a*^{pre} levels (fig. S4C), it did cause a defect in mtDNA maintenance during mitochondrial import stress caused by *MGE1* overexpression (Fig. 5D). The effects of deleting *CIS1* on the mitoCPR were subtle, presumably because proteins acting in parallel could substitute for *CIS1* function. The expression of *CIS1* from the constitutive *TEF2* promoter, however, had a substantial protective effect during mitochondrial import stress. It led to a decrease in *Cox5a*^{pre} levels after *PSD1* overexpression and an increase in the levels of mature *Cox5a* (Fig. 5E).

Drugs such as CCCP could not be used to study the role of *PDR3* during mitochondrial import stress because the drug caused *PDR3*-independent expression of mitoCPR genes, including *CIS1*. *TEF2* is, however, not controlled by any MDR response, which allowed us to explore the role of *CIS1* expressed from the *TEF2* promoter in the mitoCPR using CCCP. We induced expression of *COX5a* from the *MET25* promoter and simultaneously treated cells with CCCP. CCCP treatment partially blocked mitochondrial import, causing *Cox5a*^{pre} to accumulate. Constitutive expression of *CIS1* prevented this accumulation (Fig. 5, F to H). Constitutive *Cis1* had the same effect on the matrix proteins Rmd9, Ily2, and Mss116 (fig. S4,

D and E). Thus, high levels of *Cis1* affect precursor levels of many and perhaps all mitochondrial proteins. Constitutive *CIS1* (tagged and untagged) also protected mtDNA during mitochondrial import stress caused by overexpression of *MGE1* and even partially suppressed the detrimental effects of deleting *PDR3* on mtDNA maintenance (Fig. 5I and fig. S4F). Thus, *CIS1* is an important effector of the mitoCPR. *Cis1* reduces the levels of unimported proteins and protects mitochondrial functions during mitochondrial import stress.

Cis1 and Msp1 mediate mitochondrial preprotein clearance during mitochondrial import stress

Our results indicate that during mitochondrial import stress, *Cox5a*^{pre} accumulated on the surface of mitochondria and appeared to be stalled in the translocase (Fig. 2, A and B). *Cis1* aided in the import of preproteins, facilitated their degradation at the mitochondrial surface, or contributed to both. To test whether *Cis1* promoted the degradation of unimported proteins, we asked whether down-regulation of *Cox5a*^{pre} brought about by constitutive *CIS1* expression depended on the proteasome. Although constitutive *CIS1* prevented the accumulation of *Cox5a*^{pre} in wild-type cells treated with CCCP (Fig. 5, F to H), it failed to do so in cells that carried the temperature-sensitive *rpn6-1* allele and thus had compromised proteasome function (Fig. 6, A and B). *MET25-COX5a* was likely induced before methionine depletion in the *rpn6-1* mutant because the transcription factor responsible for activating *MET25* is a proteasome substrate (37). Thus, *CIS1* promotes proteasomal degradation of unimported proteins that accumulate at the mitochondrial surface.

How does *Cis1* promote the degradation of unimported proteins? The AAA-adenosine triphosphatase (ATPase) *Msp1* is a dislocase that extracts endoplasmic reticulum (ER) and peroxisome membrane proteins mistargeted to the mitochondrial outer membrane for proteasomal degradation (38–41). Our results show that *Msp1* has a similar function in reducing preprotein accumulation during mitochondrial import stress. Cells lacking *MSP1* accumulated high levels of *Cox5a*^{pre} when *Cox5a* expression was induced under conditions of mitochondrial import stress (CCCP treatment) (Fig. 6, C and D). Furthermore, accumulation of mature *Cox5a* was significantly delayed, suggesting that less *Cox5a* was imported into mitochondria (Fig. 6, C and E). Cells lacking *MSP1* neither accumulated *Cox5a*^{pre} nor induced mitoCPR under normal growth conditions (Fig. 6F and fig. S5A), excluding the possibility that *msp1Δ* cells were generally defective in importing proteins into mitochondria. An effect on *Cox5a*^{pre} was also observed when the *msp1-E193Q* allele was overexpressed from the *GALI-10* promoter in cells lacking endogenous *MSP1* (Fig. 6, G and H, and fig. S5B). The *E193Q* substitution, located in the Walker B motif of the ATPase domain, is predicted to disrupt ATPase activity and stabilizes ER- and peroxisome-mistargeted proteins in the outer membrane of mitochondria (38–40). Thus,

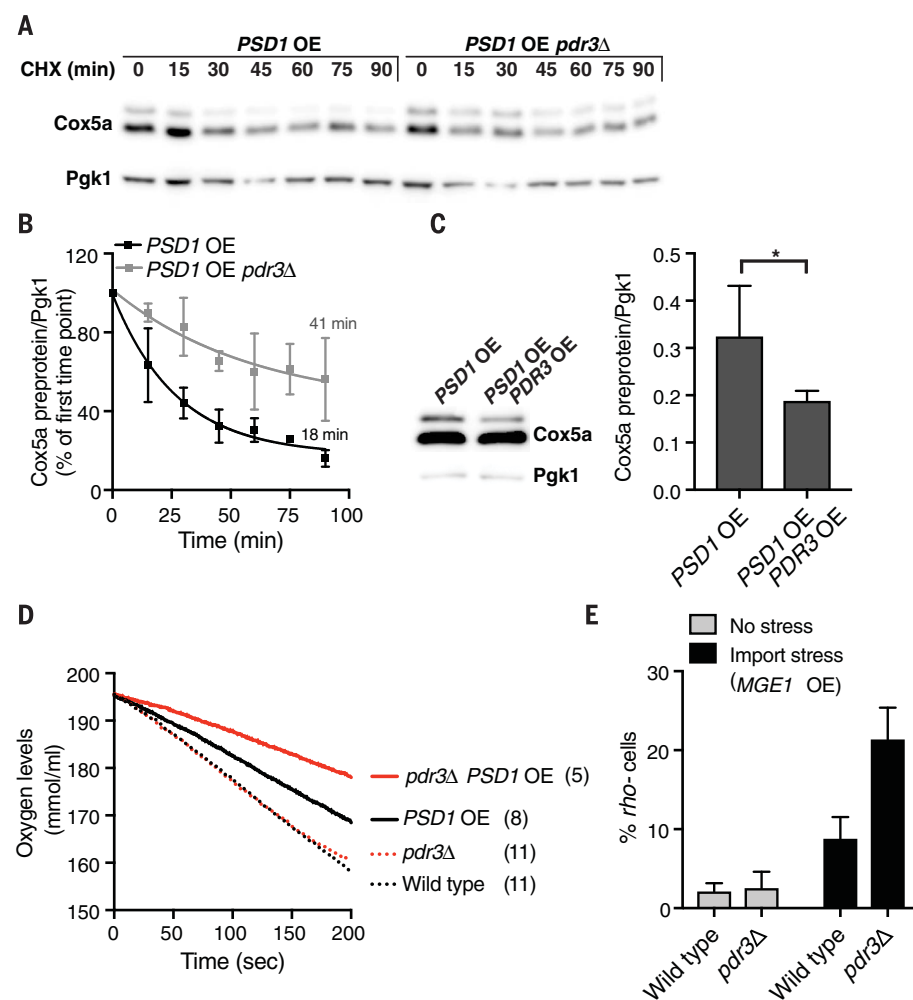
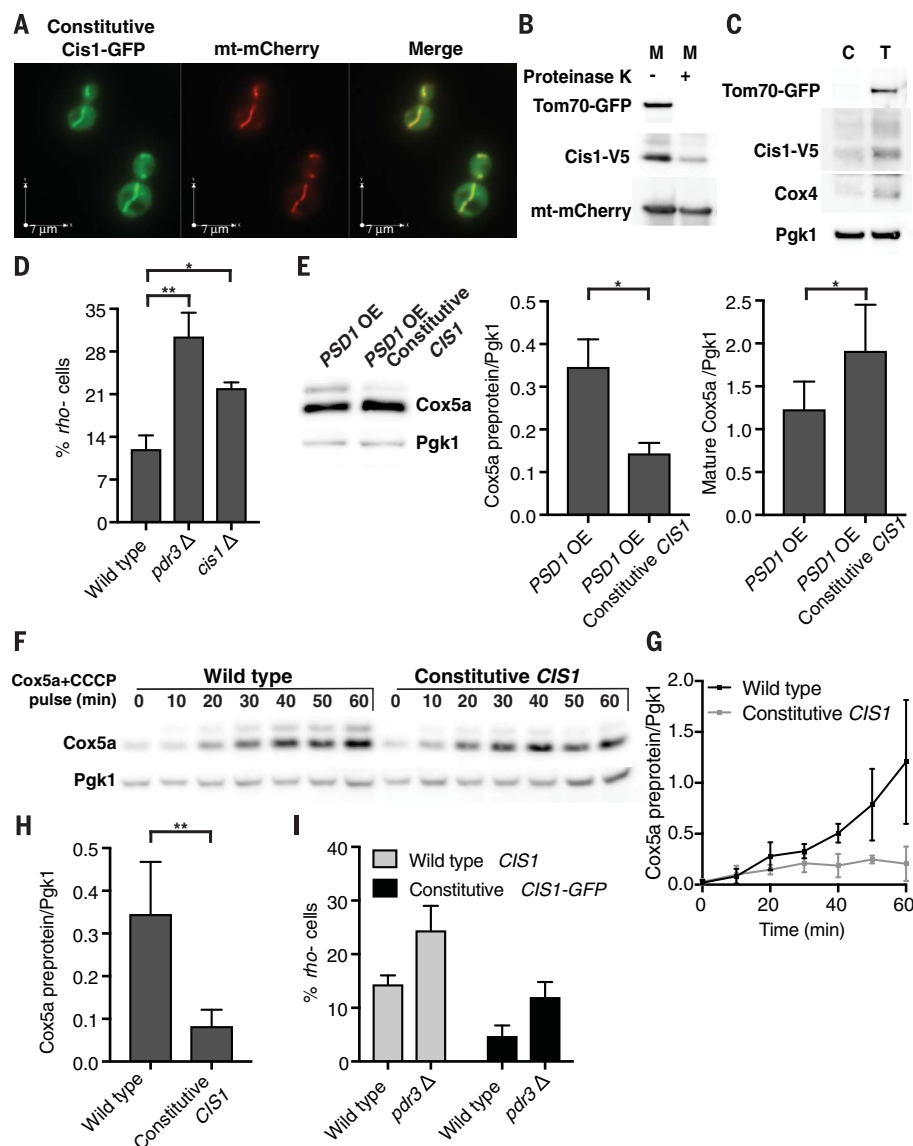


Fig. 4. The mitoCPR protects mitochondrial functions during import stress. (A) *PSD1* was overexpressed for 6 hours, and the half-life of *Cox5a* preprotein was examined after cycloheximide (0.5 mg/ml) addition in wild-type or *pdr3Δ* cells. CHX, cycloheximide; OE, overexpression. *Pgk1* served as a loading control. (B) Quantification of (A): *Cox5a* preprotein half-life. *n* = 4 experiments; data are mean ± SD. (C) Immunoblot of *Cox5a*-V5 from *GAL-PSD1* cells or *GAL-PSD1* cells overexpressing *PDR3* (*TEF2-PDR3*) 6 hours after galactose induction. Quantification of *Cox5a* preprotein from three independent experiments is depicted on the right. Data are mean ± SD. Statistics were performed by using the Student's *t* test; **P* ≤ 0.05. (D) Oxygen consumption of wild-type and *pdr3Δ* cells that did or did not overexpress *PSD1* for 4 hours. The oxygen consumption rate (nmol s⁻¹ ml) of this experiment is shown in parentheses. (E) *GAL-MGE1* and *GAL-MGE1 pdr3Δ* cells were grown for 24 hours in the presence or absence of galactose so as to induce *GAL-MGE1*. Mitochondrial DNA loss was analyzed by the appearance of *rho*⁻ colonies on 1% yeast extract, 2% peptone (YEP) plates containing 2% ethanol and 0.3% glucose. *n* = 4 experiments; data are mean ± SD.

Fig. 5. Cis1 maintains mitochondrial function during protein import stress.

(A) Live cell fluorescence imaging of cells expressing *TEF2-CIS1-GFP* and mitochondrial targeted mCherry (mt-mCherry). (B) Mitochondria were isolated from cells expressing *TEF2-CIS1-V5* that were grown in 3% glycerol. Mitochondria (M) (\pm proteinase K) are shown. mt-mCherry-matrix control protein, Tom70-GFP—outer membrane control protein. (C) Cytosolic fraction of cells presented in (B). Cytosolic (C) fraction as well as total cell lysate (T) are shown. Pgk1 served as a cytosol control protein, Tom70-GFP served as an outer-membrane control protein, and Cox4 served as a control matrix protein. (D) Wild-type, *pdr3 Δ* , and *cis1 Δ* cells were grown for 48 hours in the presence of galactose so as to induce *GAL-MGE1*. Mitochondrial DNA loss was analyzed through the appearance of *rho⁻* (petite) colonies. $n = 3$ experiments; data are mean \pm SD. Student's t test was used; $*P \leq 0.05$, $**P \leq 0.005$. (E) Immunoblot analysis of Cox5a from *GAL-PSD1* or *GAL-PSD1 TEF2-CIS1* cells after *PSD1* overexpression (6 hours). OE, overexpression. Quantifications of Cox5a preprotein (middle) and mature Cox5a (right) are shown. $n = 3$ experiments; data are mean \pm SD. Student's t test was used; $*P \leq 0.05$. (F) Wild-type or *TEF2-CIS1* cells were grown in the presence of methionine. *MET25-COX5a* was then induced through methionine removal in the presence of CCCP. Cox5a-V5 protein levels were analyzed at the indicated times (Pgk1, loading control). (G) Quantification of (F); Cox5a preprotein. $n = 4$ experiments; data are mean \pm SD. (H) Quantification of Cox5a preprotein levels 60 min after induction of *MET25-COX5a* in the presence of CCCP. $n = 6$ experiments; data are mean \pm SD. Student's t test was used; $**P \leq 0.005$. (I) Wild-type and *pdr3 Δ* cells (\pm *TEF2-CIS1-GFP*) were grown for 24 hours in the presence of galactose so as to induce *GAL-MGE1*. Mitochondrial DNA loss was analyzed as in (D). $n = 4$ experiments; data are mean \pm SD.



like Cis1, Msp1 limits the accumulation of un-imported precursor proteins.

Next, we determined the epistatic relationship between *MSP1* and *CIS1*. We asked whether *CIS1*'s ability to limit the accumulation of Cox5a^{pre} required *MSP1*. Whereas *TEF2-CIS1* prevented the accumulation of Cox5a^{pre} in wild-type cells (Fig. 5, F and G), it failed to do so in cells lacking *MSP1* (Fig. 6, I and J). Thus, Cis1's effect on preprotein clearance depended on *MSP1*.

Having established that Cis1 and Msp1 both function in preprotein clearance during mitochondrial import stress, we next asked whether the two proteins act in the same pathway. Cis1 expressed from the *TEF2* promoter coimmunoprecipitated with Msp1-E193Q-FLAG, and vice versa (Fig. 7A and fig. S6A). We were not able to detect binding between Cis1 and wild-type Msp1 most likely because this interaction is transient (fig. S6B). We did, however, obtain genetic evidence to indicate that the two proteins interact. In cells lacking *GET1*, ER membrane proteins

accumulate in the mitochondrial outer membrane (38, 39). These conditions did not induce the mitoCPR but caused a growth defect at 37°C (fig. S5, A and C) (38). Overexpression of *CIS1*, like deletion of *MSP1*, enhanced this growth defect (fig. S5C), suggesting that high levels of Cis1 reduce the interaction of Msp1, with ER proteins mistargeted to the mitochondrial outer membrane.

The observation that the association of preproteins with mitochondrial membranes was resistant to sodium carbonate treatment suggested that preproteins accumulate at translocases during mitochondrial import stress (Fig. 2, B and C). We therefore asked whether Cis1 was also found at translocases. This appeared to be the case. Localization of Cis1 to mitochondria was dependent on Tom70, a receptor of the outer-membrane translocase (Fig. 7B). Furthermore, Cis1 interacted with Tom70 as well as with Msp1 as assessed with coimmunoprecipitation analysis (Fig. 7C).

Because Cis1 is only expressed during mitochondrial import stress (fig. S4A), we conclude

that Cis1 is recruited to mitochondria under import stress, during which it interacts with both Tom70 and Msp1. Consistent with this conclusion, the interaction between Tom70 and Msp1 was enhanced during mitochondrial import stress (Fig. 7, D and E). We propose that upon recruitment to the translocase via Cis1, Msp1 evicts preproteins from the translocase and the mitochondrial surface to target them for proteasomal degradation. Our results do not exclude the possibility that Cis1 and Msp1 also improve import efficiency. We have some evidence to suggest that this may in fact be the case. Overexpression of *CIS1* caused an increase in mature Cox5a levels during prolonged mitochondrial import stress brought about by high levels of Psd1 (Fig. 5E). Similarly, *msp1 Δ* cells accumulated less mature Cox5a after CCCP treatment (Fig. 6, C and E).

Discussion

Here, we describe the discovery of a surveillance mechanism, mitoCPR, that detects mitochondrial

import stress and protects mitochondrial functions in response. We propose that the mitoCPR effector Cis1 recruits Msp1 to the outer-membrane translocase to clear stalled proteins from the translocase, and consequently improve mitochondrial import (Fig. 7F). This response is essential to protect mitochondrial functions and to maintain the mitochondrial genome during import stress. Recently, it was discovered that translation by ribosomes at the surface of mitochondria can stall (42). Whether the Msp1-Cis1 complex can clear preproteins from ribosomes during cotranslational import or whether the complex only recognizes posttranslationally imported proteins has yet to be determined. We also do not yet know whether Cis1 and Msp1 improve mitochondrial import solely by clearing unimported proteins. Our data suggest that they may also aid in the import process itself. Mitochondrial preproteins must be kept unfolded in order to translocate into mitochondria (2). A delay in mitochondrial import could result in premature folding and perhaps even aggregation of preproteins at the organelle's surface. We speculate that Msp1, whose

ATPase domain faces the cytosol, could unfold prematurely folding or aggregated preproteins, giving them a second chance to translocate into mitochondria or, when this does not occur, target them for degradation (Fig. 7F).

The mitoCPR likely performs additional functions. Mitochondrial import defects lead to widespread mitochondrial dysfunction. Up-regulation of NADPH-dependent enzymes suggests a potential role for mitoCPR in restoring redox potential. Induction of genes involved in lipid metabolism argues for an effort to compensate for lipid biosynthesis disruption. Last, up-regulation of ABC transporter gene expression may be indicative of detoxification efforts aimed at removing toxic metabolic intermediates that could accumulate in the cytosol as a result of mitochondrial dysfunction.

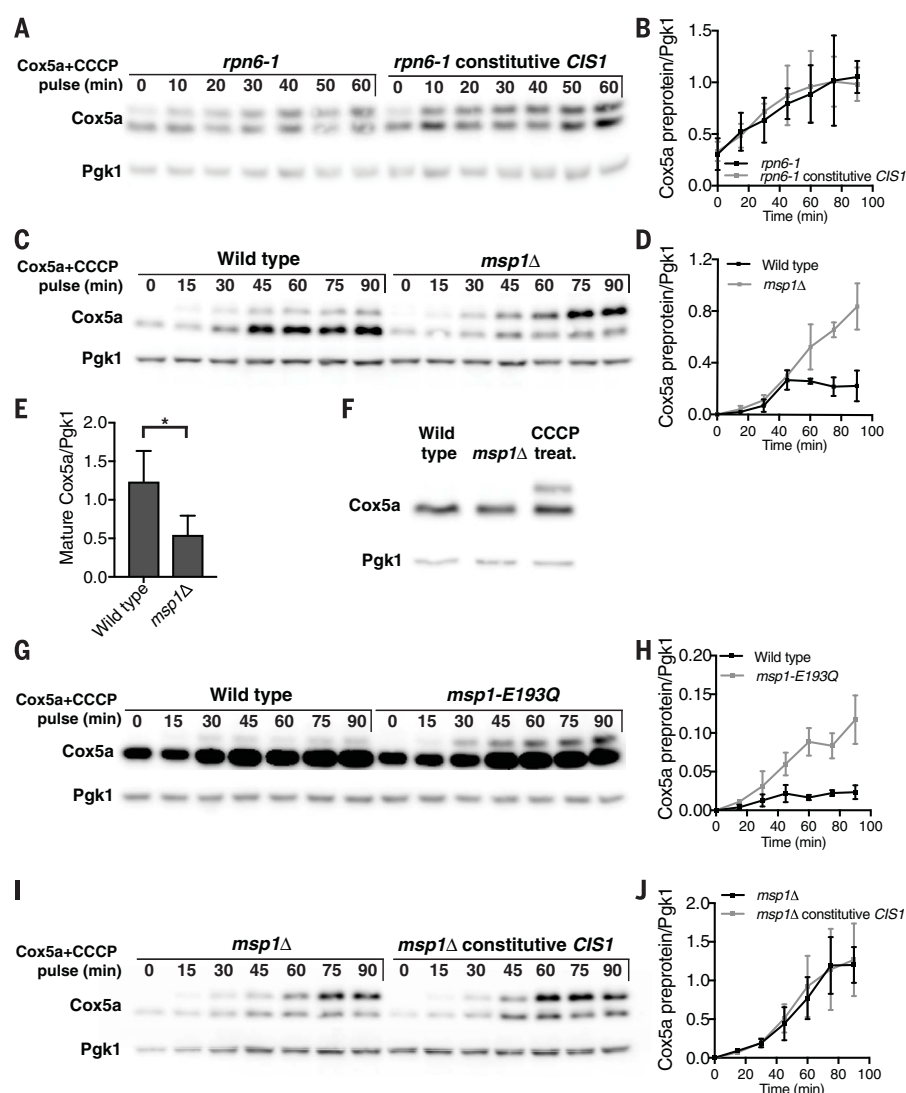
We have not yet been able to identify the signal (or signals) that activates the mitoCPR. We can thus only speculate as to how the pathway is activated. In the MDR, Pdr1 and Pdr3 are activated by binding to xenobiotics (43). Mitochondrial dysfunction resulting from defects in

mitochondrial import could lead to accumulation of metabolic intermediates in the cytoplasm, which in turn bind to and activate Pdr3. It is also possible that specific unimported proteins activate Pdr3. Such mechanisms have been described for the mitochondrial unfolded protein response and the recognition of damaged mitochondria in mammals (44, 45).

We have studied the mitoCPR in response to overexpression of bipartite signal-containing proteins. Although this is unlikely to occur under physiological conditions, budding yeast cells are exposed to microorganisms that produce compounds known to interfere with mitochondrial import in the wild (35). Import defects could also result from disease or mitochondrial stress conditions such as high levels of reactive oxygen species. *CIS1* and other mitoCPR genes are induced during diauxic shift, a physiological state defined as the switch from glycolysis to respiration that occurs when fermentable carbon sources become limiting (46). Switch to respiratory growth requires an expansion of the mitochondria compartment. We propose that this increase in mitochondrial

Fig. 6. Cis1 and Msp1 are required for preprotein clearance after mitochondrial import stress.

(A) *rpn6-1* or *rpn6-1 TEF2-CIS1* cells were grown at room temperature in the presence of methionine. Cells were then transferred into medium lacking methionine with 20 μ M CCCP at 30°C. The accumulation of Cox5a-V5 preprotein (encoded by *MET25-COX5a*) is shown. (B) Quantification of (A); Cox5a preprotein levels from four independent experiments. Data are mean \pm SD. (C) Wild-type or *msp1* Δ cells were grown at 30°C with methionine and treated as in (A). (D) Quantification of (C); Cox5a preprotein levels from four independent experiments. Data are mean \pm SD. (E) Quantification of (C); Mature Cox5a levels 60 min after induction. $n = 4$ experiments; data are mean \pm SD. Statistics were determined by using the Student's *t* test. * $P \leq 0.05$. (F) Immunoblot analysis of Cox5a-V5 from wild-type cells, wild-type cells treated with 20 μ M CCCP for 1 hour, or *msp1* Δ cells. (G) Wild-type cells or cells expressing *msp1-E193Q* from the inducible *GALI-10* promoter were grown in the presence of galactose for 6 hours. Cells were then transferred to medium lacking methionine and containing 20 μ M CCCP, and the accumulation of inducible Cox5a-V5 preprotein (encoded by *MET25-COX5a*) was examined. Cox5a levels were higher in this experiment because *MET25-COX5a* expression is higher in medium containing raffinose/galactose than in glucose (fig. S5B). (H) Quantification of (G); Cox5a preprotein levels from three independent experiments. Data are mean \pm SD. (I) *msp1* Δ cells or *msp1* Δ cells expressing *TEF2-CIS1* were treated as in (C). The experiment shown in (C) was performed in parallel, and results can thus be directly compared. (J) Quantification of (I); Cox5a preprotein levels from three independent experiments. Data are mean \pm SD.



mass, which requires increased mitochondrial import, leads to mitochondrial import stress. Mitochondria of multicellular eukaryotes are less likely to be exposed to mitochondrial poisons in the environment but do undergo increased biogenesis in specific tissues and during development. Whether a mitochondrial import stress response exists in higher eukaryotes has yet to be determined.

Materials and methods

Yeast strains and growth conditions

All strains are derivatives of W303 (AA2587) and are listed in table S2. Cells were grown overnight in YPD (1% yeast extract, 2% peptone, 2% glucose) at 30°C to saturation, then diluted in fresh YPD (OD₆₀₀ = 0.1) and grown until they reached logarithmic phase. To induce the *GALI-10* promoter, cells were grown overnight at 30°C in minimal selective medium containing 2% raffinose or in YPR (1% yeast extract, 2% peptone, 2% raffinose). Cells were then diluted to OD = 0.3 or OD = 0.1 and recovered for an hour or 3 hours, respectively, following the addition of galactose to a final concentration of 1% for 4 hours (for measuring

mRNA levels) or 6 hours (for protein analysis). To induce *MET25-COX5a*, cells were grown overnight in YPD supplemented with 8 mM methionine. Cells were diluted to OD = 0.1, grown for a few hours and then switched to medium lacking methionine [Complete supplement mixture w/o methionine (CSM, MP Biomedicals), yeast nitrogen base w/o amino acids (Difco), 2% glucose, titrated to pH 7]. CCCP was added to a final concentration of 20 μM.

Wild-type cells were incubated in the presence of 5 μg/ml ethidium bromide in YPD for 72 hours to obtain *rho0* cells. *rho0* state was verified by DAPI staining. *rho⁻* cells were obtained by deletion of the mitochondrial ribosomal subunit *MRPL16* (47). The *mrpl16Δ* strain was confirmed to be *rho⁻* by its inability to grow on medium lacking a fermentable carbon source as a haploid and as a diploid following mating with *rho0* cells. The presence of mitochondrial DNA in *mrpl16Δ* cells was tested by DAPI.

The plasmid pRS426 was used as an empty plasmid control. A plasmid expressing mt-mCherry and integrated into the *LEU2* locus was cloned from

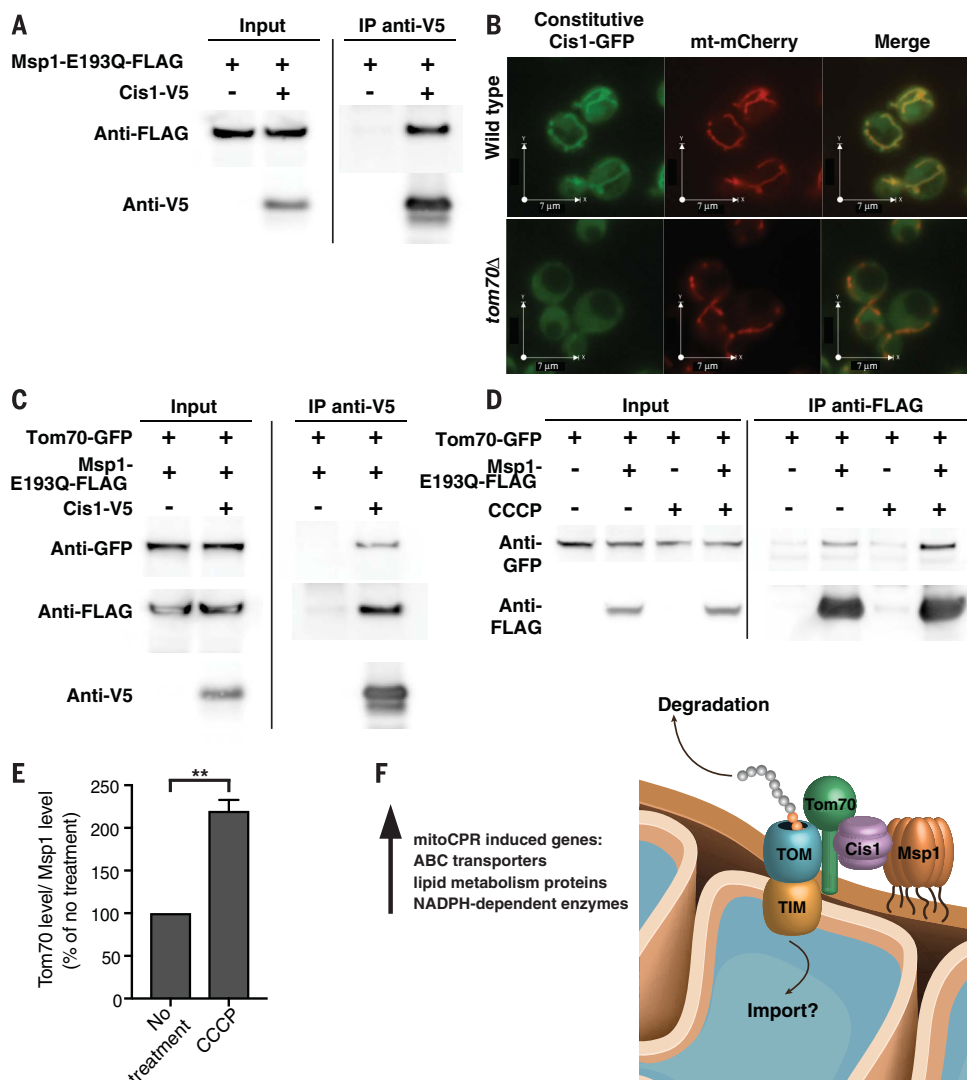
plasmid pHS12-mCherry (a gift from Benjamin Glick, Addgene plasmid # 25444).

Immunoblot analysis

For immunoblot analyses, ~2 OD₆₀₀ units of cells were harvested and treated with 5% trichloroacetic acid overnight at 4°C. The acid was washed away with acetone and the cell pellet was subsequently dried. The cell pellet was pulverized with glass beads in 100 μl of lysis buffer (50 mM Tris-HCl at pH 7.5, 1 mM EDTA, 2.75 mM DTT) using a bead-beater. 3 × SDS sample buffer was added and the cell homogenates were boiled. Samples were separated by SDS-PAGE, blotted onto nitrocellulose membranes, and subsequently incubated with anti-V5 antibodies (1:2000 dilution; Life Technologies), anti-3-Phosphoglycerate Kinase antibodies (1:5000 dilution; Invitrogen), anti-GFP antibodies (1:1000; Clontech, JL-8), anti-Kar2 (1:200,000 dilution; kindly provided by Mark Rose), anti-Myc antibodies (1:1000 dilution; Sigma, 9E10), anti-Cox4 antibodies (1:1000; Abcam) or anti-FLAG antibodies (1:1000; Sigma). HRP-linked sheep anti-mouse antibodies and

Fig. 7. Cis1 interacts with Msp1 and with the outer-membrane translocase. (A) Cells expressing

msp1-E193Q-FLAG and cells expressing *msp1-E193Q-FLAG* and *TEF2-CIS1-V5* were grown in yeast extract, peptone, and glucose (YPD). Cells were lysed, and Cis1-V5 was immunoprecipitated by using antibodies to V5. (B) Live cell fluorescence imaging of wild-type or *tom70Δ* cells expressing *TEF2-CIS1-GFP* and mitochondrial-targeted mCherry (mt-mCherry). (C) Cis1-V5 (encoded by *TEF2-CIS1-V5*) was immunoprecipitated by using antibodies to V5 from *TOM70-GFP*– and *msp1-E193Q-FLAG*–expressing cells. Cells expressing only *TOM70-GFP* and *msp1-E193Q-FLAG* were used as control. (D) Cells expressing *TOM70-GFP* or *msp1-E193Q-FLAG* and *TOM70-GFP* were grown in YPD in the presence or absence of 20 μM CCCP for 1 hour. Msp1-E193Q-FLAG was immunoprecipitated by using antibodies to FLAG. (E) Quantification of (D); coimmunoprecipitated Tom70 levels (normalized to coimmunoprecipitated Msp1 levels) in nontreated and CCCP-treated cells from three independent experiments. No treatment was set to 100%. Data are mean ± SD. Statistics were performed by using the Student's *t* test; ***P* ≤ 0.005. (F) A model for how Cis1 and Msp1 affect mitochondrial import during import stress. IMS, intermembrane space.



HRP-linked donkey anti-rabbit antibodies (GE Healthcare) were used as secondary antibodies. Statistics were performed using the Student's *t* test. The protein half-life in Fig. 4B was analyzed as a one-phase exponential decay chart using Prism software.

Fluorescence microscopy

Cells were grown overnight in minimal medium at 30°C, diluted to OD = 0.1 and grown to logarithmic phase. Images were acquired with a DeltaVision Elite microscope (GE Healthcare Bio-Sciences, Pittsburgh, PA). Images were taken with a 100× plan-Apo objective, an InsightSSI solid-state light source, and a CoolSNAP HQ2 camera.

Real-time PCR

Total RNA was isolated using the RNeasy mini-kit (Qiagen). RNA (750 ng) was used to generate cDNAs using the SuperScript III first strand synthesis system (Life Technologies). Quantitative PCR was performed using a SYBR green mix (Life Technologies) and amplified using a LightCycler 480 II (Roche). Signals were normalized to *ACT1* transcript levels and are presented as fold increase of control conditions.

Gene expression analysis

For RNA expression analysis, *PSD1* was over-expressed for 4 hours. Total yeast RNA was isolated using the RNeasy mini-kit (Qiagen) and samples were sequenced on a HiSeq 2000. *S. cerevisiae* RNA-seq reads were aligned to the sacCer3 genome with STAR version 2.5.3a and Ensembl transcripts were quantified using rsem version 1.3.0. Differential expression analysis was performed using deseq2 version 1.16.1 running under R version 3.4.0. Default options were selected for deseq2 runs except cooksCutoff and independent Filtering were both set to false during results preparation and unmoderated fold changes were used. RNA sequencing data can be accessed via the following link: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107784.

Mitochondrial oxygen consumption

Cells were grown overnight at 30°C in minimal selective medium with 2% raffinose. The cells were then diluted to OD = 0.3 and recovered for an hour following the addition of galactose to a final concentration of 1% for 4 hours. Cells were then transferred to YPG (1% yeast extract, 2% peptone, 3% glycerol) and incubated for 20 min. Oxygen consumption rate was measured from 0.75 OD (1 ml) cells in YPG using an Oxytherm instrument (Hansatech) for 3 min at 25°C. The slope of the linear range of oxygen depletion was used to measure oxygen consumption rate of 3 experiments. Statistics were performed using the Student's *t* test.

Mitochondrial DNA maintenance assay

The analysis of mtDNA maintenance was described previously (48). Cells were grown overnight at 30°C in minimal selective medium with 2% glucose. Cells were then diluted to OD = 0.15

in minimal selective medium with 2% raffinose and were grown for 3 hours following the addition of galactose to a final concentration of 1% for 24 or 48 hours. Within these 24 hours (8 hours after induction) the cells were diluted 1:20 into the same medium. Yeast cells (~200) were spread on plates containing 1% yeast extract, 2% peptone, 0.3% glucose, 2% ethanol and were grown at 30°C for 3 days until all colonies could be detected. The percentage of small *rho*⁻ (petite) colonies was determined from 3 different experiments.

Membrane potential measurements

Cells lacking *PDR1*, *PDR3*, and *PDR5* (to prevent efflux of dyes out of the cells) bearing either an empty plasmid (for control and CCCP treatment) or a *GAL-PSD1* containing plasmid and expressing a mitochondria-targeted mCherry (mt-mCherry) were grown overnight at 30°C in minimal selective medium containing 2% raffinose. The cells were diluted to OD = 0.3 and recovered for an hour following the addition of galactose to a final concentration of 1% for 4 hours. CCCP (20 μM) was added for 1 hour. Cells were then transferred to 1 ml dye buffer (10 mM Hepes pH 7.2 and 5% glucose) and incubated with 2.5 μM Rhodamine 123 (Thermo Fisher Scientific) for 15 min at room temperature. Cells were washed 5 times in 1.5 ml dye buffer. Mitochondria were identified by mt-mCherry labeling. Membrane potential was analyzed by the following equation: (mitochondrial fluorescence intensity – cytosolic fluorescence intensity)/cytosolic fluorescence intensity cytosol.

Mitochondria isolation

Cells were grown to logarithmic phase, collected by centrifugation and washed once with water. Cells were then resuspended in 0.1 M Tris pH 9.4, 10 mM DTT and incubated for 20 min at 30°C. Cell walls were disturbed by incubation in 1.2 M sorbitol, 20 mM K₂HPO₄ pH 7.4, 1% zymolyase for 1 hour at 30°C. Dounce homogenization was used to lyse the cells in 0.6 M sorbitol, 10 mM Tris pH 7.4, 1 mM EDTA, fatty acid free 0.2% BSA and 1 mM PMSF. Mitochondria were then isolated by differential centrifugation as described previously (49) and resuspended in SEM buffer (0.25 M sucrose, 10 mM MOPS KOH pH 7.2 and 1 mM EDTA). Proteinase K was added to a final concentration of 50 μg/ml for 5 min at 37°C and the reaction was stopped by the addition of 4 mM PMSF for 15 min on ice.

For sodium carbonate extraction, 40 μg of mitochondria were pelleted and resuspended in 500 μl of 100 mM sodium carbonate pH 11 or in SEM buffer for the untreated control. The samples were kept on ice for 30 min followed by centrifugation at 90,000 g for 30 min. Supernatants and pellets were incubated with 12.5% TCA overnight at 4°C and separated by SDS-PAGE.

Coimmunoprecipitation assays

Cells were grown in YPD to OD = 0.9 when not treated or to OD = 0.7 following treatment with 20 μM CCCP for 1 hour. Approximately 50 OD units of cells were collected, washed once with

water and frozen. Cells were lysed with Silica Beads using a FastPrep instrument (speed 6.5, 45 s, 3 cycles) with 200 μl IGEPAL buffer [50 mM Tris pH 7.5, 150 mM NaCl, 1% IGEPAL and Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific)]. Lysates were brought up to 1.5 ml with IGEPAL buffer containing 0.2% BSA. Lysates were clarified by centrifugation at 20,000 g for 10 min at 4°C. Twenty μl of Anti-V5 agarose affinity gel antibody (Sigma) or Anti-FLAG M2 affinity gel (Sigma) were added and lysates were incubated for 2 hours at 4°C. Beads were then washed 5 times with IGEPAL buffer containing 0.2% BSA. Sample buffer was added to the beads, which were then boiled. Final eluates and two percent of the lysates were separated by means of SDS-PAGE.

REFERENCES AND NOTES

1. A. Chacinska, C. M. Koehler, D. Milenkovic, T. Lithgow, N. Pfanner, Importing mitochondrial proteins: Mechanisms and mechanisms. *Cell* **138**, 628–644 (2009). doi: [10.1016/j.cell.2009.08.005](https://doi.org/10.1016/j.cell.2009.08.005); pmid: [19703392](https://pubmed.ncbi.nlm.nih.gov/19703392/)
2. W. Neupert, J. M. Herrmann, Translocation of proteins into mitochondria. *Annu. Rev. Biochem.* **76**, 723–749 (2007). doi: [10.1146/annurev.biochem.76.052705.163409](https://doi.org/10.1146/annurev.biochem.76.052705.163409); pmid: [17263264](https://pubmed.ncbi.nlm.nih.gov/17263264/)
3. J. A. MacKenzie, R. M. Payne, Mitochondrial protein import and human health and disease. *Biochim. Biophys. Acta* **1772**, 509–523 (2007). doi: [10.1016/j.bbdis.2006.12.002](https://doi.org/10.1016/j.bbdis.2006.12.002); pmid: [17300922](https://pubmed.ncbi.nlm.nih.gov/17300922/)
4. A. B. Harbauer, R. P. Zahedi, A. Sickmann, N. Pfanner, C. Meisinger, The protein import machinery of mitochondria—A regulatory hub in metabolism, stress, and disease. *Cell Metab.* **19**, 357–372 (2014). doi: [10.1016/j.cmet.2014.01.010](https://doi.org/10.1016/j.cmet.2014.01.010); pmid: [24561263](https://pubmed.ncbi.nlm.nih.gov/24561263/)
5. H. Yano et al., Inhibition of mitochondrial protein import by mutant huntingtin. *Nat. Neurosci.* **17**, 822–831 (2014). doi: [10.1038/nn.3721](https://doi.org/10.1038/nn.3721); pmid: [24836077](https://pubmed.ncbi.nlm.nih.gov/24836077/)
6. L. Wrobel et al., Mistargeted mitochondrial proteins activate a proteostatic response in the cytosol. *Nature* **524**, 485–488 (2015). doi: [10.1038/nature14951](https://doi.org/10.1038/nature14951); pmid: [26245374](https://pubmed.ncbi.nlm.nih.gov/26245374/)
7. X. Wang, X. J. Chen, A cytosolic network suppressing mitochondria-mediated proteostatic stress and cell death. *Nature* **524**, 481–484 (2015). doi: [10.1038/nature14859](https://doi.org/10.1038/nature14859); pmid: [26192197](https://pubmed.ncbi.nlm.nih.gov/26192197/)
8. E. Itakura et al., Ubiquitins chaperone and triage mitochondrial membrane proteins for degradation. *Mol. Cell* **63**, 21–33 (2016). doi: [10.1016/j.molcel.2016.05.020](https://doi.org/10.1016/j.molcel.2016.05.020); pmid: [27345149](https://pubmed.ncbi.nlm.nih.gov/27345149/)
9. B. Sarkadi, L. Homolya, G. Szakács, A. Váradi, Human multidrug resistance ABCB and ABCG transporters: Participation in a chemoinnate defense system. *Physiol. Rev.* **86**, 1179–1236 (2006). doi: [10.1152/physrev.00037.2005](https://doi.org/10.1152/physrev.00037.2005); pmid: [17015488](https://pubmed.ncbi.nlm.nih.gov/17015488/)
10. S. Kumar, M. M. Mukherjee, M. F. Varela, Modulation of bacterial multidrug resistance efflux pumps of the major facilitator superfamily. *Int. J. Bacteriol.* **2013**, 1–15 (2013). doi: [10.1155/2013/204141](https://doi.org/10.1155/2013/204141); pmid: [25750934](https://pubmed.ncbi.nlm.nih.gov/25750934/)
11. B. Akache, S. MacPherson, M.-A. Sylvain, B. Turcotte, Complex interplay among regulators of drug resistance genes in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**, 27855–27860 (2004). doi: [10.1074/jbc.M403487200](https://doi.org/10.1074/jbc.M403487200); pmid: [15123673](https://pubmed.ncbi.nlm.nih.gov/15123673/)
12. N. V. C. Coorey, J. H. Matthews, D. S. Bellows, P. H. Atkinson, Pleiotropic drug-resistance attenuated genomic library improves elucidation of drug mechanisms. *Mol. Biosyst.* **11**, 3129–3136 (2015). doi: [10.1039/C5MB00406C](https://doi.org/10.1039/C5MB00406C); pmid: [26381459](https://pubmed.ncbi.nlm.nih.gov/26381459/)
13. W. S. Moye-Rowley, Transcriptional control of multidrug resistance in the yeast *Saccharomyces*. *Prog. Nucleic Acid Res. Mol. Biol.* **73**, 251–279 (2003). doi: [10.1016/S0079-6603\(03\)01008-0](https://doi.org/10.1016/S0079-6603(03)01008-0); pmid: [12882520](https://pubmed.ncbi.nlm.nih.gov/12882520/)
14. T. C. Hallstrom, W. S. Moye-Rowley, Multiple signals from dysfunctional mitochondria activate the pleiotropic drug resistance pathway in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **275**, 37347–37356 (2000). doi: [10.1074/jbc.M007338200](https://doi.org/10.1074/jbc.M007338200); pmid: [10980204](https://pubmed.ncbi.nlm.nih.gov/10980204/)
15. W. S. Moye-Rowley, Retrograde regulation of multidrug resistance in *Saccharomyces cerevisiae*. *Gene* **354**, 15–21 (2005). doi: [10.1016/j.gene.2005.03.019](https://doi.org/10.1016/j.gene.2005.03.019); pmid: [15896930](https://pubmed.ncbi.nlm.nih.gov/15896930/)

16. M. G. Cumsky, C. E. Trueblood, C. Ko, R. O. Poyton, Structural analysis of two genes encoding divergent forms of yeast cytochrome c oxidase subunit V. *Mol. Cell. Biol.* **7**, 3511–3519 (1987). doi: [10.1128/MCB.7.10.3511](#); pmid: [2824989](#)
17. Y. Fujiki, A. L. Hubbard, S. Fowler, P. B. Lazarow, Isolation of intracellular membranes by means of sodium carbonate treatment: Application to endoplasmic reticulum. *J. Cell Biol.* **93**, 97–102 (1982). doi: [10.1083/jcb.93.1.97](#); pmid: [7068762](#)
18. K. Gulshan, J. A. Schmidt, P. Shahi, W. S. Moye-Rowley, Evidence for the bifunctional nature of mitochondrial phosphatidylserine decarboxylase: Role in Pdr3-dependent retrograde regulation of PDR5 expression. *Mol. Cell. Biol.* **28**, 5851–5864 (2008). doi: [10.1128/MCB.00405-08](#); pmid: [18644857](#)
19. C. B. Epstein *et al.*, Genome-wide responses to mitochondrial dysfunction. *Mol. Biol. Cell* **12**, 297–308 (2001). doi: [10.1091/mbc.12.2.297](#); pmid: [11179416](#)
20. R. D. Appleby *et al.*, Quantitation and origin of the mitochondrial membrane potential in human cells lacking mitochondrial DNA. *Eur. J. Biochem.* **262**, 108–116 (1999). doi: [10.1046/j.1432-1327.1999.00350.x](#); pmid: [10231371](#)
21. J. R. Veatch, M. A. McMurray, Z. W. Nelson, D. E. Gottschling, Mitochondrial dysfunction leads to nuclear genome instability via an iron-sulfur cluster defect. *Cell* **137**, 1247–1258 (2009). doi: [10.1016/j.cell.2009.04.014](#); pmid: [19563757](#)
22. T. C. Hallstrom *et al.*, Coordinate control of sphingolipid biosynthesis and multidrug resistance in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 23674–23680 (2001). doi: [10.1074/jbc.M101568200](#); pmid: [11323424](#)
23. M. R. Gallas, M. K. Dienhart, R. A. Stuart, R. M. Long, Characterization of Mmp37p, a *Saccharomyces cerevisiae* mitochondrial matrix protein with a role in mitochondrial protein import. *Mol. Biol. Cell* **17**, 4051–4062 (2006). doi: [10.1091/mbc.E06-04-0366](#); pmid: [16790493](#)
24. Y. Tamura *et al.*, Identification of Tam41 maintaining integrity of the TIM23 protein translocator complex in mitochondria. *J. Cell Biol.* **174**, 631–637 (2006). doi: [10.1083/jcb.200603087](#); pmid: [16943180](#)
25. D. K. Woo, T. L. Phang, J. D. Trawick, R. O. Poyton, Multiple pathways of mitochondrial-nuclear communication in yeast: Intergenomic signaling involves ABF1 and affects a different set of genes than retrograde regulation. *Biochim. Biophys. Acta* **1789**, 135–145 (2009). doi: [10.1016/j.bbaggm.2008.09.008](#); pmid: [18977319](#)
26. P. Golik, N. Bonnefoy, T. Szczepanek, Y. Saint-Georges, J. Lazowska, The Rieske FeS protein encoded and synthesized within mitochondria complements a deficiency in the nuclear gene. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8844–8849 (2003). doi: [10.1073/pnas.1432907100](#); pmid: [12837937](#)
27. B. R. Francis, K. H. White, P. E. Thorsness, Mutations in the Atp1p and Atp3p subunits of yeast ATP synthase differentially affect respiration and fermentation in *Saccharomyces cerevisiae*. *J. Bioenerg. Biomembr.* **39**, 127–144 (2007). doi: [10.1007/s10863-007-9071-4](#); pmid: [17492370](#)
28. N. Mutlu, G. Garipiler, E. Akdoğan, C. D. Dunn, Activation of the pleiotropic drug resistance pathway can promote mitochondrial DNA retention by fusion-defective mitochondria in *Saccharomyces cerevisiae*. *G3 (Bethesda)* **4**, 1247–1258 (2014). doi: [10.1534/g3.114.010330](#); pmid: [24807265](#)
29. S. Laloraya, B. D. Gambill, E. A. Craig, A role for a eukaryotic GrpE-related protein, Mge1p, in protein translocation. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6481–6485 (1994). doi: [10.1073/pnas.91.14.6481](#); pmid: [8022808](#)
30. J. Ostermann *et al.*, Precursor proteins in transit through mitochondrial contact sites interact with hsp70 in the matrix. *FEBS Lett.* **277**, 281–284 (1990). doi: [10.1016/0014-5793\(90\)80865-G](#); pmid: [2176621](#)
31. P. E. Scherer, U. C. Krieg, S. T. Hwang, D. Vestweber, G. Schatz, A precursor protein partly translocated into yeast mitochondria is bound to a 70 kd mitochondrial stress protein. *EMBO J.* **9**, 4315–4322 (1990). pmid: [2265609](#)
32. S. Sugimoto, K. Saruwatari, K. Higashi, K. Sonomoto, The proper ratio of GrpE to DnaK is important for protein quality control by the DnaK-DnaJ-GrpE chaperone system and for cell division. *Microbiology* **154**, 1876–1885 (2008). doi: [10.1099/mic.0.2008/017376-0](#); pmid: [18599817](#)
33. H. Zhang, K. K. Singh, Global genetic determinants of mitochondrial DNA copy number. *PLOS ONE* **9**, e105242 (2014). doi: [10.1371/journal.pone.0105242](#); pmid: [25170845](#)
34. F. Devaux, E. Carvajal, S. Moye-Rowley, C. Jacq, Genome-wide studies on the nuclear PDR3-controlled response to mitochondrial dysfunction in yeast. *FEBS Lett.* **515**, 25–28 (2002). doi: [10.1016/S0014-5793\(02\)02387-6](#); pmid: [11943188](#)
35. S. Naranjo *et al.*, Dissecting the genetic basis of a complex cis-regulatory adaptation. *PLOS Genet.* **11**, e1005751 (2015). doi: [10.1371/journal.pgen.1005751](#); pmid: [26713447](#)
36. W.-K. Huh *et al.*, Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003). doi: [10.1038/nature02026](#); pmid: [14562095](#)
37. A. Rouillon, R. Barbey, E. E. Patton, M. Tyers, D. Thomas, Feedback-regulated degradation of the transcriptional activator Met4 is triggered by the SCF(Met30) complex. *EMBO J.* **19**, 282–294 (2000). doi: [10.1093/emboj/19.2.282](#); pmid: [10637232](#)
38. Y.-C. Chen *et al.*, Msp1/ATAD1 maintains mitochondrial function by facilitating the degradation of mislocalized tail-anchored proteins. *EMBO J.* **33**, 1548–1564 (2014). doi: [10.15252/emboj.201487943](#); pmid: [24843043](#)
39. V. Okreglak, P. Walter, The conserved AAA-ATPase Msp1 confers organelle specificity to tail-anchored proteins. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8019–8024 (2014). doi: [10.1073/pnas.1405755111](#); pmid: [24821790](#)
40. M. L. Wohlever, A. Mateja, P. T. McGilvray, K. J. Day, R. J. Keenan, Msp1 is a membrane protein dislocase for tail-anchored proteins. *Mol. Cell* **67**, 194–202.e6 (2017). doi: [10.1016/j.molcel.2017.06.019](#); pmid: [28712723](#)
41. N. R. Weir, R. A. Kamber, J. S. Martenson, V. Denic, The AAA protein Msp1 mediates clearance of excess tail-anchored proteins from the peroxisomal membrane. *eLife* **6**, e28507 (2017). doi: [10.7554/eLife.28507](#); pmid: [28906250](#)
42. T. Izawa, S.-H. Park, L. Zhao, F. U. Hartl, W. Neupert, Cytosolic protein Vms1 links ribosome quality control to mitochondrial and cellular homeostasis. *Cell* **171**, 890–903.e18 (2017). doi: [10.1016/j.cell.2017.10.002](#); pmid: [29107329](#)
43. J. K. Thakur *et al.*, A nuclear receptor-like pathway regulating multidrug resistance in fungi. *Nature* **452**, 604–609 (2008). doi: [10.1038/nature06836](#); pmid: [18385733](#)
44. R. J. Youle, D. P. Narendra, Mechanisms of mitophagy. *Nat. Rev. Mol. Cell Biol.* **12**, 9–14 (2011). doi: [10.1038/nrm3028](#); pmid: [21179058](#)
45. A. M. Nargund, M. W. Pellegrino, C. J. Fiorese, B. M. Baker, C. M. Haynes, Mitochondrial import efficiency of ATFS-1 regulates mitochondrial UPR activation. *Science* **337**, 587–590 (2012). doi: [10.1126/science.1223560](#); pmid: [22700657](#)
46. M. J. Brauer, A. J. Saldanha, K. Dolinski, D. Botstein, Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures. *Mol. Biol. Cell* **16**, 2503–2517 (2005). doi: [10.1091/mbc.E04-11-0968](#); pmid: [15758028](#)
47. A. M. Myers, L. K. Pape, A. Tzagoloff, Mitochondrial protein synthesis is required for maintenance of intact mitochondrial genomes in *Saccharomyces cerevisiae*. *EMBO J.* **4**, 2087–2092 (1985). pmid: [3905388](#)
48. E. Baruffini, I. Ferrero, F. Foury, In vivo analysis of mtDNA replication defects in yeast. *Methods* **51**, 426–436 (2010). doi: [10.1016/j.jymeth.2010.02.023](#); pmid: [20206271](#)
49. N. Pfanner, C. Meisinger, B. Turcotte, In *Yeast Protocols*, W. Xiao, Ed. (Springer, 2006), vol. 313, pp. 33–39.
50. D. Mossmann *et al.*, Amyloid- β peptide induces mitochondrial dysfunction by inhibition of preprotein maturation. *Cell Metab.* **20**, 662–669 (2014). doi: [10.1016/j.cmet.2014.07.024](#); pmid: [25176146](#)
51. S. E. Horvath *et al.*, Processing and topology of the yeast mitochondrial phosphatidylserine decarboxylase 1. *J. Biol. Chem.* **287**, 36744–36755 (2012). doi: [10.1074/jbc.M112.398107](#); pmid: [22984266](#)

ACKNOWLEDGMENTS

We are grateful to T. Fox, C. Koehler, P. Perlman, F. Solomon, M. Vander Heiden, J. Rutter, and V. Denic for discussions and insights and for critical reading of the manuscript. We thank A. Sandıkçı for figure preparation, X. Zhou for help with data analysis, and N. Gebert for technical help. We thank D. Gottschling for the *ATP1-111* plasmid and P. Walter for the *GAL-msp1-E193Q* plasmid. We thank C. Whittaker of the Barbara K. Ostrom (1978) Bioinformatics and Computing Facility at Koch Institute in the Swanson Biotechnology Center for analyzing the gene expression. We thank J. Falk, E. MacDuffie, and the members of the Amon laboratory for critical reading of the manuscript.

Funding: This work was supported by the National Institutes of Health (grant GM 118066 to A.A.) and by the Koch Institute Support (core) Grant P30-CA 14051 from the National Cancer Institute. A.A. is also an investigator of the Howard Hughes Medical Institute and the Glenn Foundation for Biomedical Research. H.W. was supported by the Jane Coffin Childs Memorial Fund, the European Molecular Biology Organization Long-Term Fellowship, and the Israel National Postdoctoral Program for Advancing Women in Science.

Author contributions: A.A. and H.W. designed and interpreted the experiments and drafted and revised the manuscript. H.W. acquired and analyzed the data. **Competing interests:** None declared. **Data and materials availability:** RNA sequencing data can be accessed via the following link: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107784. All other data needed to evaluate the conclusions in this paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/eaan4146/suppl/DC1
Figs. S1 to S6
Tables S1 and S2

9 April 2017; resubmitted 19 December 2017

Accepted 19 February 2018

10.1126/science.aan4146

RESEARCH ARTICLE

BIG DATA

Quantitative analysis of population-scale family trees with millions of relatives

Joanna Kaplanis,^{1,2*} Assaf Gordon,^{1,2*} Tal Shor,^{3,4} Omer Weissbrod,⁵ Dan Geiger,⁴ Mary Wahl,^{1,2,6} Michael Gershovits,² Barak Markus,² Mona Sheikh,² Melissa Gymrek,^{1,2,7,8,9} Gaurav Bhatia,^{10,11} Daniel G. MacArthur,^{7,9,10} Alkes L. Price,^{10,11,12} Yaniv Erlich^{1,2,3,13,14†}

Family trees have vast applications in fields as diverse as genetics, anthropology, and economics. However, the collection of extended family trees is tedious and usually relies on resources with limited geographical scope and complex data usage restrictions. We collected 86 million profiles from publicly available online data shared by genealogy enthusiasts. After extensive cleaning and validation, we obtained population-scale family trees, including a single pedigree of 13 million individuals. We leveraged the data to partition the genetic architecture of human longevity and to provide insights into the geographical dispersion of families. We also report a simple digital procedure to overlay other data sets with our resource.

Family trees are mathematical graph structures that can capture mating and parent-hood among humans. As such, the edges of the trees represent potential transmission lines for a wide variety of genetic, cultural, sociodemographic, and economic factors. Quantitative genetics is built on dissecting the interplay of these factors by overlaying data on family trees and analyzing the correlation of various classes of relatives (1–3). In addition, family trees can serve as a multiplier for genetic information through study designs that leverage genotype or phenotype data from relatives (4–7), analyzing parent-of-origin effects (8), refining heritability measures (9, 10), or improving individual risk assessment (11, 12). Beyond classical genetic ap-

plications, large-scale family trees have played an important role across disciplines, including human evolution (13, 14), anthropology (15), and economics (16).

Despite the range of applications, constructing population-scale family trees has been a labor-intensive process. Previous approaches mainly relied on local data repositories such as churches or vital-records offices (14, 17, 18). But these approaches have limitations (19, 20): They require nontrivial resources to digitize the records and organize the data, the resulting trees are usually limited in geographical scope, and the data may be subject to strict usage protections. These challenges reduce demographic accessibility and complicate fusion with information such as genomic or health data.

Constructing and validating population-scale family trees

Here, we leveraged genealogy-driven social media data to construct population-scale family trees. To this end, we focused on Geni.com, a crowdsourcing website in the genealogy domain. Users can create individual profiles and upload family trees. The website automatically scans profiles to detect similarities and offers the option to merge the profiles when a match is detected. By merging, larger family trees are created that can be collaboratively managed to improve their accuracy. After obtaining relevant permissions, we downloaded approximately 86 million publicly available profiles (21). The input data consisted of millions of individual profiles, each of which describes a person; for 43 million of these profiles, the data also included any putative connections to other individuals in the data set.

Similar to other crowdsourcing projects (22), a small group of participants contributed the majority of genealogy profiles (fig. S1).

We organized the profiles into graph topologies that preserve the genealogical relationships between individuals (Fig. 1A). Biology dictates that a family tree should form a directed acyclic graph, where each individual has an in-degree that is less than or equal to 2. However, 0.3% of the profiles resided in invalid biological topologies that included cycles (e.g., a person who is both the parent and child of another person) or an individual with more than two parents. We developed an automated pipeline to resolve local conflicts and prune invalid topologies (fig. S2) and benchmarked the performance of the pipeline against human genealogists (21). This resulted in >90% concordance between the pipeline and human decisions to resolve conflicts, thereby generating 5.3 million disjoint family trees.

The largest family tree in the processed data spanned 13 million individuals who were connected by shared ancestry and marriage (Fig. 1B). On average, the tree spanned 11 generations between each terminal descendant and their founders (fig. S3). The size of this pedigree fits what is expected as familial genealogies coalesce at a logarithmic rate compared to the size of the population (23).

We evaluated the structure of the tree by inspecting the genetic segregation of unilineal markers. We obtained mitochondrial DNA (mtDNA) and Y-chromosome short tandem repeat (Y-STR) haplotypes to compare multiple pairs of relatives in our graph (21). The mtDNA data were available for 211 lineages and spanned a total of 1768 transmission events (i.e., graph edges), whereas the Y-STR data were available for 27 lineages that spanned 324 total transmission events. Using a prior of no more than a single nonpaternity event per lineage, we estimated a nonmaternity rate of 0.3% per meiosis and nonpaternity rate of 1.9% per meiosis. This rate of nonpaternity matched previous rates of Y-chromosome studies (24, 25) and the nonmaternity rate was close to historical rates of adoption of an unrelated member in the United States (26). Taken together, these results show that millions of genealogists can produce high-quality population-scale family trees.

Extracting demographic data

We found that life span in the Geni.com profiles was largely concordant with reports generated by traditional demographic approaches. First, we extracted demographic information from the collected profiles with exact birth and death dates, thereby avoiding the problems inherent in profiles with only year resolution for these events, such as heaping at round years (fig. S4). The data reflected historical events and trends, such as elevated death rates at military age during the American Civil War and First and Second World Wars and a reduction in child mortality during the 20th century (Fig. 2A). We compared the average life span in our collection to a worldwide historical analysis covering

¹New York Genome Center, New York, NY 10013, USA.

²Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. ³MyHeritage, Or Yehuda 6037606, Israel.

⁴Computer Science Department, Technion-Israel Institute of Technology, Haifa 3200003, Israel. ⁵Computer Science Department, Weizmann Institute of Science, Rehovot 7610001, Israel. ⁶Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA.

⁷Harvard Medical School, Boston, MA 02115, USA. ⁸Harvard-MIT Program in Health Sciences and Technology, Cambridge, MA 02142, USA. ⁹Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA.

¹⁰Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

¹¹Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA. ¹²Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA. ¹³Department of Computer Science, Fu Foundation School of Engineering, Columbia University, New York, NY, USA. ¹⁴Center for Computational Biology and Bioinformatics, Department of Systems Biology, Columbia University, New York, NY, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: erlichya@gmail.com

the years 1840 to 2000 (27). We found an R^2 value of 0.95 between the expected life span from historical data and the Geni data set (Fig. 2B) and a 98% concordance with historical distributions reported by the Human Mortality Database (HMD) (Fig. 2C and fig. S5).

Next, we extracted the geographic locations of life events by two approaches: an automated geoparsing pipeline and structured text manually curated and approved by genealogists (21) (fig. S6A). Overall, we were able to place about 16 million profiles into longitude/latitude coordinates, typically at fine-scale geographic resolution, without major differences in quality between the automated geoparsing and manual curations for subsequent analyses (fig. S6B) (21). The profiles were distributed across a wide range of locations in the Western world (Fig. 2D and fig. S7), with 55% from Europe and 30% from North America. We analyzed profiles in 10 cities across the globe and found that the first appearance of profiles was only after the known first settlement date for nearly all of the cities, suggesting good spatiotemporal assignment of profiles (Fig. 2E). Movie S1 presents the place of birth of individuals in the Geni data set in 5-year intervals from 1400 to 1900 along with known migration events.

We were concerned that the Geni.com profiles might suffer from certain socioeconomic ascertainment biases and therefore would not reflect the local population. To evaluate this concern,

we collected ~80,000 publicly available death certificates from the Vermont Department of Health for every death in that state between 1985 and 2010. These records have extensive information for each individual, including education level, place of birth, and a cause of death in an ICD-9 code. About 1000 individuals in Geni overlapped this death certificate collection. We compared the education level, birth state, and ICD-9 code between these ~1000 Geni profiles and the entire Vermont collection. For all three parameters, we found >98% concordance between the distribution of these key sociodemographic attributes in the Geni profiles in Vermont and the entire state of Vermont (tables S1 to S3). Overall, this high level of consistency argues against severe socioeconomic ascertainment. Table S4 reports key demographic and genetic attributes for various familial relationships from parent-child via great-great-grandparents to fourth cousins.

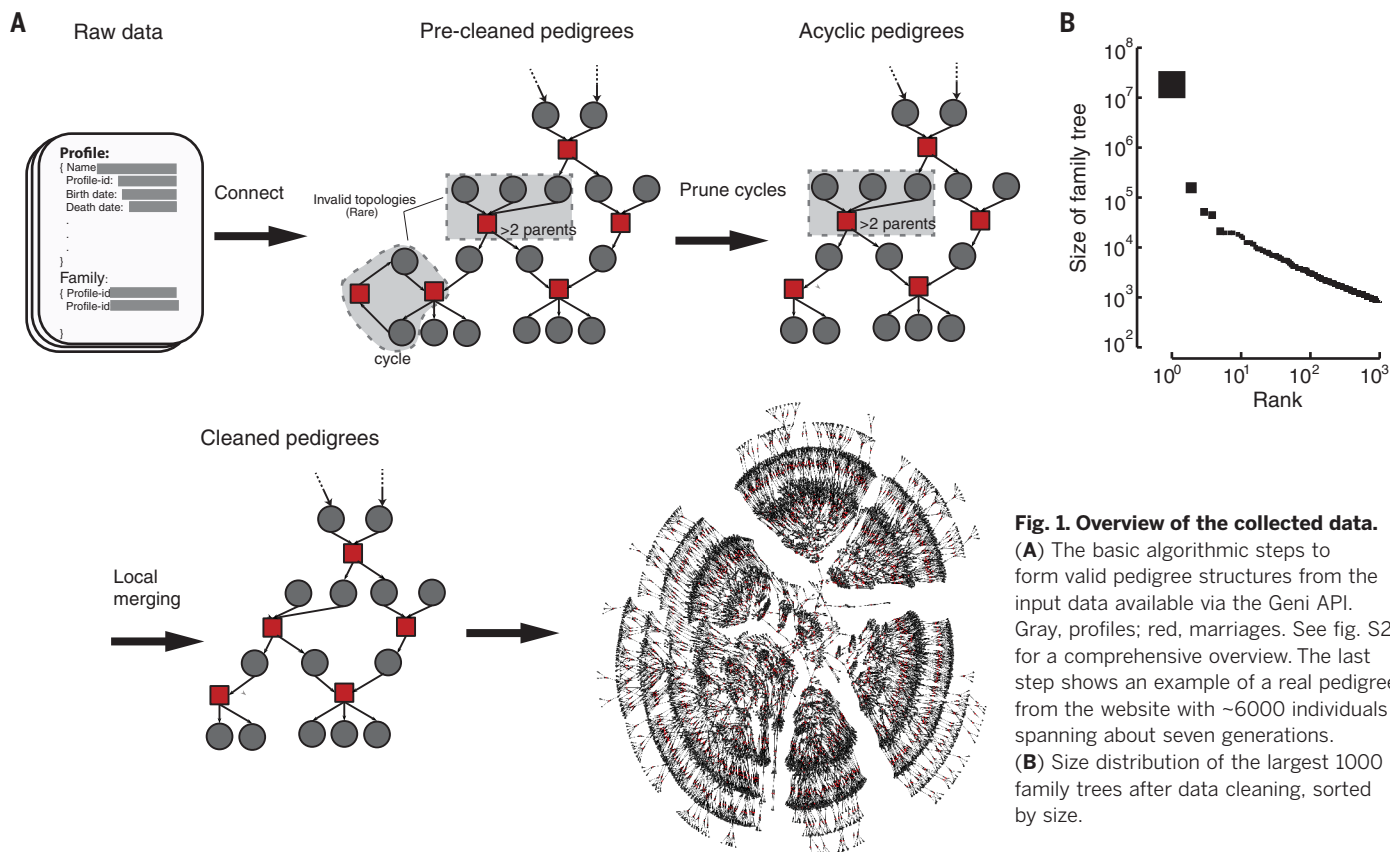
Characterizing the genetic architecture of longevity

We leveraged the Geni data set to characterize the genetic architecture of human longevity, which exhibits complex genetics likely to involve a range of physiological and behavioral endophenotypes (28, 29). Narrow-sense heritability (h^2) of longevity has been estimated to be around 15 to 30% (table S5) (30–35). Genome-wide association studies have had limited success in identifying genetic variants associated with longevity (36–38). This

relatively large proportion of missing heritability can be explained by the following: (i) Longevity has nonadditive components that create upward bias in estimates of heritability (39), (ii) estimators of heritability are biased as a result of unaccounted environmental effects (10), and (iii) the trait is highly polygenic and requires larger cohorts to identify the underlying variants (40). We thus sought to harness our resource and build a model for the sources of genetic variance in longevity that jointly evaluates additivity, dominance, epistasis, shared household effects, spatiotemporal trends, and random noise.

We adjusted longevity to be the difference between age of death and expected life span, using a model that we trained with 3 million individuals. Our model includes spatiotemporal and sex effects and was the best among 10 different models that adjusted various spatiotemporal attributes (fig. S8). We also validated this model by estimating h^2 according to the mid-parent design (41) with nearly 130,000 parent-child trios. This process yielded $h^2_{\text{mid-parent}} = 12.2\%$ (SE = 0.4%) (Fig. 3A), which is on the lower end but in the range of previous heritability estimates (table S5). Consistent with previous studies, we did not observe any temporal trend in mid-parent heritability (Fig. 3B).

We partitioned the source of genetic variance of longevity using more than 3 million pairs of relatives from full sibling to fourth cousin (21). We measured the variance explained by an



additive component, a pairwise epistatic model, three-way epistasis, and dominance (Fig. 3C). These 3 million pairs were all sex-concordant to address residual sex differences not accounted for by our longevity adjustments (fig. S9) and do not include relatives who are likely to have died because of environmental catastrophes or in major wars (fig. S10); this mitigated correlations due to nongenetic factors. We also refined the genetic correlation of the relatives by considering multiple genealogical paths (figs. S11 to S13).

The analysis of longevity in these 3 million pairs of relatives showed a robust additive genetic component, a small impact of dominance, and no detectable epistasis (Fig. 3D and table S6) (27). Additivity was highly significant ($P_{\text{additive}} < 10^{-318}$) with an estimated $h^2_{\text{sex-concordant/relatives}} = 16.1\%$ (SE = 0.4%), similar to the heritability estimated from sex-concordant parent-child pairs, $h^2_{\text{concordant/parent-child}} = 15.0\%$ (SE = 0.4%). The maximum-likelihood estimate for dominance was around 4%, but the epistatic terms converged to zero despite the substantial amount of data. Other model selection procedures, such as mean squared error analysis and Bayesian information criterion, argued against a pervasive epistatic contribution to longevity variance in the population (21).

We tested the ability of our model to predict the longevity correlation of an orthogonal data set of 810 monozygotic twin pairs collected by the Danish Twin Registry (Fig. 3D) (42). Our inferred model for longevity accurately predicted the observed correlation of this twin cohort with 1% difference, well within the sampling error for the mean twin correlation (SE = 3.2%). We also evaluated an extensive array of additional analyses that included various adjustments for environmental components and other confounders (figs. S14 and S15) (21). In all cases, additivity explained 15.8 to 16.9% of the longevity estimates, dominance explained 2 to 4%, and no evidence for epistatic interactions could be detected using our procedure.

We also estimated the additive and epistatic components using a method that allows rapid estimation of variance components of extremely large relationship matrices, called sparse Cholesky factorization linear mixed models (Sci-LMM) (43). This method takes into account a kinship coefficient matrix of 250 million pairs of related individuals in the Geni data set and includes adjustments for population structure, sex, and year of birth. We observed an additivity of 17.8% (SE = 0.84%) and a pairwise epistatic component that was not significantly different from zero (21).

Taken together, our results across multiple study designs (fig. S16) indicate that the limited ability of genome-wide association studies so far to associate variants with longevity cannot be attributed to statistical epistasis. Note that this does not rule out the existence of molecular interactions between genes contributing to this trait (44–47). On the basis of a large number of data points and study designs, we measured an additive component ($h^2 \approx 16\%$) that is considerably smaller than the 25% figure that is generally cited in the literature. These results indicate that previous studies are likely to have overestimated the heritability of longevity. As such, we should lower our expectations about our ability to predict longevity from genomic data and presumably to identify causal genetic variants.

Assessment of theories of familial dispersion

Familial dispersion is a major driving force of various genetic, economic, and demographic processes (48). Previous work has primarily relied on vital records from a limited geographical scope (49, 50) or used indirect inference from genetic data sets that mainly illuminate distant historical events (51).

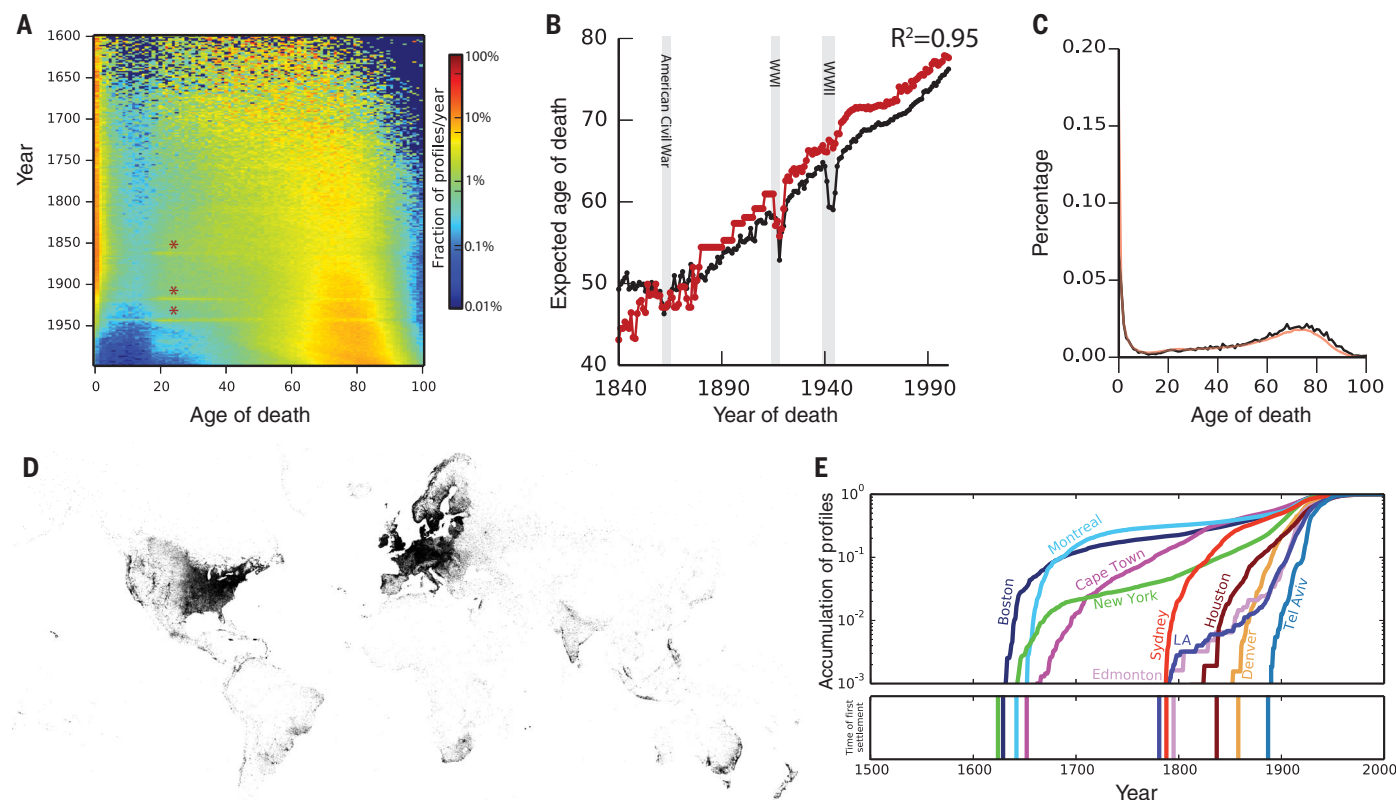


Fig. 2. Analysis and validation of demographic data. (A) Distribution of life expectancy per year. Colors correspond to the frequency of profiles of individuals who died at a certain age for each year. Asterisks indicate deaths at military age in the Civil War and First and Second World Wars. (B) Expected life span in Geni (black) and the Oeppen and Vaupel study [red (27)] as a function of year of death. (C) Comparison

of the life-span distributions versus Geni (black) and HMD (red). See also fig. S5A. (D) Geographic distribution of the annotated place-of-birth information. Every pixel corresponds to a profile in the data set. (E) Validation of geographical assignment by historical trends. Top: Cumulative distribution of profiles since 1500 for each city on a logarithmic scale as a function of time. Bottom: Year of first settlement in the city.

We harnessed our resource to evaluate patterns of human migration. First, we analyzed sex-specific migration patterns (21) to resolve conflicting results regarding sex bias in human migration (52). Our results indicate that in Western societies, females migrate more than males but over shorter distances. Median mother-child distances were significantly larger than median father-child distances by a factor of 1.6 (Wilcoxon, one-tailed, $P < 10^{-90}$) (Fig. 4A). This trend appeared throughout the 300 years of our analysis window, including

in the most recent birth cohort, and was observed both in North American duos (Wilcoxon, one-tailed, $P < 10^{-23}$) and European duos (Wilcoxon, one-tailed, $P < 10^{-87}$). On the other hand, we found that average mother-child distances (fig. S17) were significantly shorter than average father-child distances (t test, $P < 10^{-90}$), which suggests that long-range migration events are biased toward males. Consistent with this pattern, fathers displayed a significantly ($P < 10^{-83}$) higher frequency than mothers to be born in a different

country than their offspring (Fig. 4B). Again, this pattern was evident when restricting the data to North American or European duos. Taken together, males and females in Western societies show different migration distributions; patrilocal events occur only in relatively local migration events, and large-scale events that usually involve a change of country are more common in males than in females.

Next, we inspected the marital radius (the distance between mates' places of birth) and its

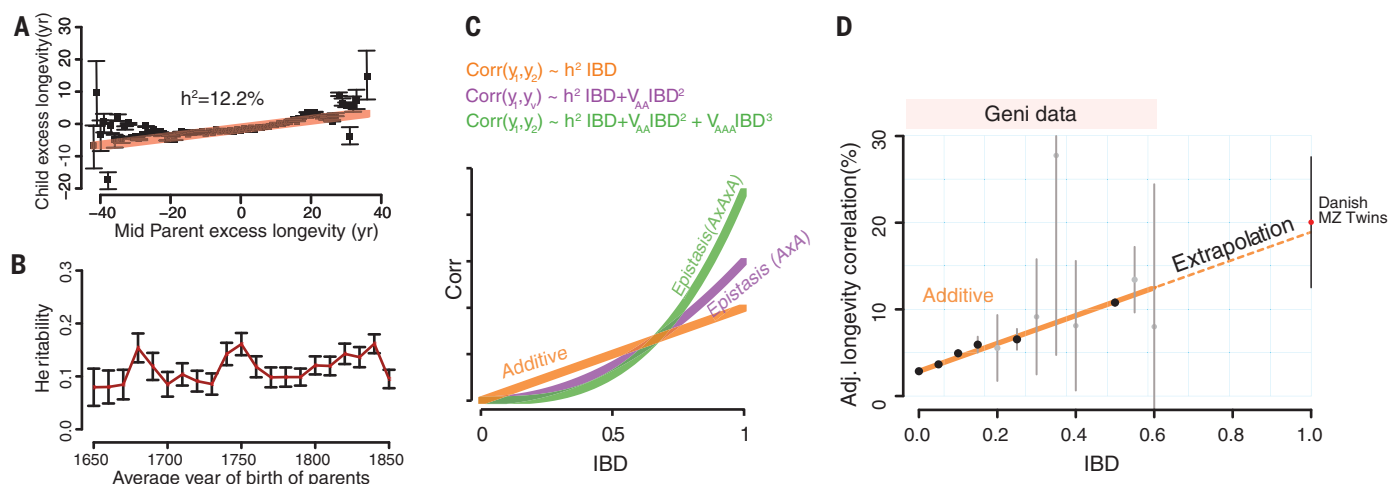


Fig. 3. The genetic architecture of longevity. (A) Regression (red) of child longevity on its mid-parent longevity (defined as difference between age of death and expected life span). Black squares, average longevity of children binned by the mid-parent value; gray bars, estimated 95% confidence interval (CI). (B) Estimated narrow-sense heritability (red) with 95% confidence intervals (black bars) obtained by the mid-parent design stratified by the average decade of birth of the parents.

(C) Correlation of a trait as a function of IBD under strict additive (h^2 , orange), squared (V_{AA} , purple), and cubic (V_{AAA} , green) epistasis architectures after dominance adjustments. (D) Average longevity correlation as a function of IBD (black circles) grouped in 5% increments (gray: 95% CI) after adjusting for dominance. A dashed line denotes the extrapolation of the models toward monozygotic twins from the Danish Twin Registry (red circle).

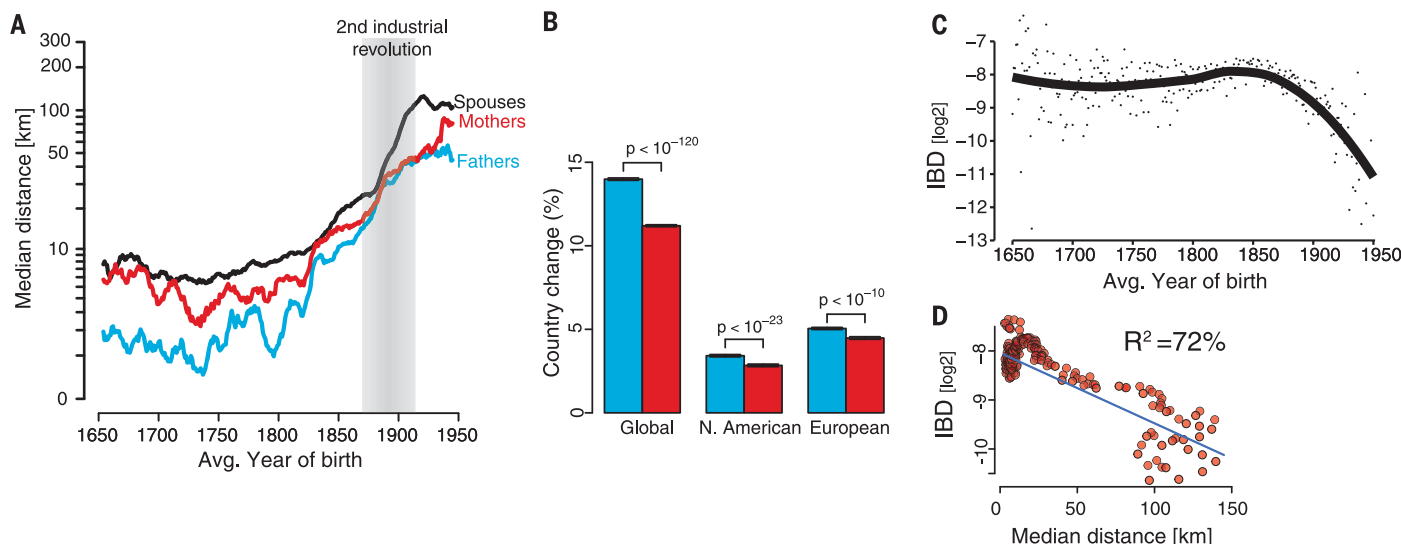


Fig. 4. Analysis of familial dispersion. (A) Median distance [$\log_{10}(x + 1)$] of father-offspring places of birth (cyan), mother-offspring (red), and marital radius (black) as a function of time (average year of birth). (B) Rate of change in the country of birth for father-offspring (cyan) or mother-offspring (red) stratified by major geographic areas. (C) Average IBD (\log_2) between

couples as a function of average year of birth. Individual dots represent the measured average per year; the black line denotes the smooth trend using locally weighted regression. (D) IBD of couples as a function of marital radius. Each dot represents a year between 1650 to 1950. The blue line denotes the best linear regression line in log-log space.

effect on the genetic relatedness of couples (21). The isolation-by-distance theory of Malécot predicts that increases in the marital radius should exponentially decrease the genetic relatedness of individuals (53). But the magnitude of these forces is also a function of factors such as taboos against cousin marriages (54).

We started by analyzing temporal changes in the birth locations of couples in our cohort. Before the Industrial Revolution (earlier than 1750), most marriages occurred between people born only 10 km from each other (Fig. 4A, black line). Similar patterns were found when analyzing European-born individuals (fig. S18) or North American-born individuals (fig. S19). After the beginning of the second Industrial Revolution (1870), the marital radius rapidly increased and reached ~100 km for most marriages in the birth cohort in 1950. Next, we analyzed the expected identity-by-descent (IBD) of couples as measured by tracing their genealogical ties (Fig. 4C). Between 1650 and 1850, the average IBD of couples was relatively stable and on the order of fourth cousins, whereas IBD exhibited a rapid decrease after 1850. Overall, the median marital radius for each year showed a strong correlation ($R^2 = 72\%$) with the expected IBD between couples. Every 70-km increase in the marital radius correlated with a decrease in the genetic relatedness of couples by one meiosis event (Fig. 4D). This correlation matches previous isolation-by-distance forces in continental regions (55). However, this trend is not consistent over time and exhibits three phases. For the pre-1800 birth cohorts, the correlation between marital distance and IBD was insignificant ($P > 0.2$) and weak ($R^2 = 0.7\%$) (fig. S20A). Couples born around 1800 to 1850 showed a doubling of their marital distance, from 8 km in 1800 to 19 km in 1850. Marriages usually occur about 20 to 25 years after birth, and around this time (1820 to 1875) rapid transportation changes took place, such as the advent of railroad travel in most of Europe and the United States. However, the increase in marital distance was significantly ($P < 10^{-13}$) coupled with an increase in genetic relatedness, contrary to the isolation-by-distance theory (fig. S20B). Only for the cohorts born after 1850 did the data match ($R^2 = 80\%$) the theoretical model of isolation by distance (fig. S20C).

Taken together, the data show a 50-year lag between the advent of increased familial dispersion and the decline of genetic relatedness between couples. During this time, individuals continued to marry relatives despite the increased distance. From these results, we hypothesize that changes in 19th-century transportation were not the primary cause for decreased consanguinity. Rather, our results suggest that shifting cultural factors played a more important role in the recent reduction of genetic relatedness of couples in Western societies.

Discussion

In this work, we leveraged genealogy-driven media to build a data set of human pedigrees of massive

scale that covers nearly every country in the Western world. Multiple validation procedures indicated that it is possible to obtain a data set that has similar quality to traditionally collected studies, but at much greater scale and lower cost.

We envision that this and similar large data sets can address quantitative aspects of human families, including genetics, anthropology, public health, and economics. Our tree and demographic data are available in a de-identified format, enabling static analysis of the Geni data set. We also offer a dynamic method that enables fusing other data sets with our data, based on digital consent of participants using the Geni application programming interface (API) (fig. S21) (21). We have been using this one-click mechanism to overlay thousands of genomes with family trees on DNA.Land (56). Other projects can use a similar strategy to add large pedigrees to their existing data collection.

More generally, similar to previous studies (57, 58), our work demonstrates the synergistic power of a collaboration between basic research and consumer genetic genealogy data sets. With ever-growing digitization of humanity and the rise of consumer genetics (59), we believe that such collaborative efforts can be a valuable path to reach the scale of information needed to address fundamental questions in biomedical research.

REFERENCES AND NOTES

- R. A. Fisher, *Trans. R. Soc. Edinb.* **52**, 399–433 (1919).
- S. Wright, *J. Agric. Res.* **20**, 557–585 (1921).
- A. Tenesa, C. S. Haley, *Nat. Rev. Genet.* **14**, 139–149 (2013).
- A. Kong et al., *Nat. Genet.* **40**, 1068–1075 (2008).
- J. K. Lowe et al., *PLOS Genet.* **5**, e1000365 (2009).
- D. F. Gudbjartsson et al., *Nat. Genet.* **47**, 435–444 (2015).
- J. Z. Liu, Y. Erlich, J. K. Pickrell, *Nat. Genet.* **49**, 325–331 (2017).
- A. Kong et al., *Nature* **462**, 868–874 (2009).
- C. Ober, M. Abney, M. S. McPeck, *Am. J. Hum. Genet.* **69**, 1068–1079 (2001).
- N. Zaitlen et al., *PLOS Genet.* **9**, e1003520 (2013).
- R. Valdez, P. W. Yoon, N. Qureshi, R. F. Green, M. J. Khoury, *Annu. Rev. Public Health* **31**, 69–87 (2010).
- C. B. Do, D. A. Hinds, U. Francke, N. Eriksson, *PLOS Genet.* **8**, e1002973 (2012).
- M. Lahdenperä, V. Lummaa, S. Helle, M. Tremblay, A. F. Russell, *Nature* **428**, 178–181 (2004).
- C. Moreau et al., *Science* **334**, 1148–1150 (2011).
- A. Helgason, S. Pálsson, D. F. Gudbjartsson, T. Kristjánsson, K. Stefánsson, *Science* **319**, 813–816 (2008).
- J. Modalsli, “Multigenerational persistence: Evidence from 146 years of administrative data” (Statistics Norway, 2016); <https://EconPapers.repec.org/RePEc:ssb:dispa:850>.
- J. R. Gulcher, K. Stefánsson, in *Encyclopedia of Life Sciences* (Wiley, 2001).
- L. A. Cannon-Albright, *Hum. Hered.* **65**, 209–220 (2008).
- L. A. Cannon-Albright, in *AMIA Annual Symposium Proceedings* (American Medical Informatics Association, 2006), p. 1161.
- V. Stefánsson et al., *J. Community Genet.* **4**, 1–7 (2013).
- See supplementary materials.
- A. Kittur, E. Chi, B. A. Pendleton, B. Suh, T. Mytkowicz, *World Wide Web* **1**, 19 (2007).
- J. T. Chang, *Adv. Appl. Probab.* **31**, 1002–1026 (1999).
- K. Anderson, *Curr. Anthropol.* **47**, 513–520 (2006).
- T. E. King, M. A. Jobling, *Mol. Biol. Evol.* **26**, 1093–1102 (2009).
- P. Maza, *Child Welf. Res. Notes* **9**, 1–11 (1984).
- J. Oeppen, J. W. Vaupel, *Science* **296**, 1029–1031 (2002).
- P. Sebastiani, T. T. Perls, *Front. Genet.* **3**, 277 (2012).
- R. E. Marioni et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13366–13371 (2016).
- P. Philippe, J. M. Opitz, *Am. J. Med. Genet.* **2**, 121–129 (1978).
- P. J. Mayer, *Am. J. Hum. Biol.* **3**, 49–58 (1991).
- B. Ljungquist, S. Berg, J. Lanke, G. E. McClearn, N. L. Pedersen, *J. Gerontol. A* **53**, M441–M446 (1998).
- A. M. Herskind et al., *Hum. Genet.* **97**, 319–323 (1996).
- B. D. Mitchell et al., *Am. J. Med. Genet.* **102**, 346–352 (2001).
- R. A. Kerber, E. O'Brien, K. R. Smith, R. M. Cawthon, *J. Gerontol. A* **56**, B130–B139 (2001).
- P. Sebastiani et al., *PLOS ONE* **7**, e29848 (2012).
- J. Deelen et al., *Hum. Mol. Genet.* **23**, 4420–4432 (2014).
- G. A. Erikson et al., *Cell* **165**, 1002–1011 (2016).
- O. Zuk, E. Hechter, S. R. Sunyaev, E. S. Lander, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1193–1198 (2012).
- E. A. Boyle, Y. I. Li, J. K. Pritchard, *Cell* **169**, 1177–1186 (2017).
- P. M. Visscher, W. G. Hill, N. R. Wray, *Nat. Rev. Genet.* **9**, 255–266 (2008).
- A. Skytthe, K. O. Kyvik, N. V. Holm, K. Christensen, *Scand. J. Public Health* **39** (suppl.), 75–78 (2011).
- T. Shor, D. Geiger, Y. Erlich, O. Weissbrod, *BioRxiv* 256396 [Preprint], 30 January 2018. <https://doi.org/10.1101/256396>.
- W. Li, J. Reich, *Hum. Hered.* **50**, 334–349 (2000).
- P. C. Phillips, *Nat. Rev. Genet.* **9**, 855–867 (2008).
- H. J. Cordell, *Nat. Rev. Genet.* **10**, 392–404 (2009).
- W.-H. Wei, G. Hemani, C. S. Haley, *Nat. Rev. Genet.* **15**, 722–733 (2014).
- L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, 1994).
- E. M. Wijsman, L. L. Cavalli-Sforza, *Annu. Rev. Ecol. Syst.* **15**, 279–301 (1984).
- R. Labouriau, A. Amorim, *Genetics* **178**, 601–603 (2008).
- K. R. Veeramah, M. F. Hammer, *Nat. Rev. Genet.* **15**, 149–162 (2014).
- L. J. Lawson Handley, N. Perrin, *Mol. Ecol.* **16**, 1559–1578 (2007).
- G. Malécot, *The Mathematics of Heredity* (Freeman, 1970).
- L. L. Cavalli-Sforza, A. Moroni, G. Zei, *Consanguinity, Inbreeding, and Genetic Drift in Italy* (Princeton Univ. Press, 2004).
- J. H. Relethford, E. R. Brennan, *Hum. Biol.* **54**, 315–327 (1982).
- J. Yuan et al., *Nat. Genet.* **50**, 160–165 (2018).
- J. K. Pickrell et al., *Nat. Genet.* **48**, 709–717 (2016).
- E. Han et al., *Nat. Commun.* **8**, 14238 (2017).
- R. Khan, D. Mittelman, *Genome Biol.* **14**, 139 (2013).

ACKNOWLEDGMENTS

We thank D. Zielsinski, G. Japhet, and J. Novembre for valuable comments, the Erlich lab members for constant support in pursuing this project, and the Vermont Health Department for providing all death certificates. This study was supported by a generous gift from Andria and Paul Heafy (Y.E.), the Burroughs Wellcome Fund Career Awards at the Scientific Interface (Y.E.), the Broad Institute's SPARC: Catalytic Funding for Novel Collaborative Projects award (Y.E. and D.G.M.), NIH grants R01 MH101244 and R03 HG006731 (A.L.P.), and Israeli Science Foundation grant 1678/12 (D.G.). Author contributions: A.G. and Y.E. conducted the downloading, indexing, and organizing of the data; J.K., A.G., M.W., B.M., M.Ge., M.S., and Y.E. developed the procedures to clean the family trees and extract demographic information; J.K., T.S., O.W., D.G., M.G., G.B., D.G.M., A.L.P., and Y.E. were involved in analyzing the genetic architecture of longevity; J.K., M.W., and Y.E. conducted the analysis of human migration; and J.K., T.S., O.W., D.G.M., A.L.P., and Y.E. wrote the manuscript. T.S. and Y.E. became employees of MyHeritage.com, the parent company of Geni.com, during the course of this study. The other authors do not declare relevant competing interests. The Geni data set without names is available from Y.E. under the terms described on FamilyTree.com. The code for the API integration is available at <https://github.com/TeamErlich/geni-integration-example>, the code for Sci-LMM is available at <https://github.com/TalShor/SciLMM>, and the code to download Geni profiles is available at <https://github.com/erlichya/geni-download>. The Human Mortality Database (HMD) is available at www.mortality.org. The Danish Twin Registry (DTR) data are available upon request from the University of Southern Denmark (www.sdu.dk/en/om_sdu/institutter_centre/ist_sundhedstjenesteforsk/centre/dtr). The findings, opinions, and recommendations expressed herein are those of the authors and are not necessarily those of the DTR. The Vermont Death Certificate collection was obtained upon request from the Chief of Public Health Statistics, Vermont Department of Health (www.healthvermont.gov/stats).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/171/suppl/DC1
Materials and Methods
Figs. S1 to S21
Tables S1 to S6
Movie S1
References (60–79)

7 February 2017; resubmitted 2 November 2017
Accepted 7 February 2018
Published online 1 March 2018
10.1126/science.1239309

SINGLE-CELL GENOMICS

Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding

Alexander B. Rosenberg,^{1*†} Charles M. Roco,^{2*} Richard A. Muscat,¹ Anna Kuchina,¹ Paul Sample,¹ Zizhen Yao,³ Lucas T. Graybuck,³ David J. Peeler,² Sumit Mukherjee,¹ Wei Chen,⁴ Suzie H. Pun,² Drew L. Sellers,^{2,5} Bosiljka Tasic,³ Georg Seelig^{1,4,6†}

To facilitate scalable profiling of single cells, we developed split-pool ligation-based transcriptome sequencing (SPLiT-seq), a single-cell RNA-seq (scRNA-seq) method that labels the cellular origin of RNA through combinatorial barcoding. SPLiT-seq is compatible with fixed cells or nuclei, allows efficient sample multiplexing, and requires no customized equipment. We used SPLiT-seq to analyze 156,049 single-nucleus transcriptomes from postnatal day 2 and 11 mouse brains and spinal cords. More than 100 cell types were identified, with gene expression patterns corresponding to cellular function, regional specificity, and stage of differentiation. Pseudotime analysis revealed transcriptional programs driving four developmental lineages, providing a snapshot of early postnatal development in the murine central nervous system. SPLiT-seq provides a path toward comprehensive single-cell transcriptomic analysis of other similarly complex multicellular systems.

More than 300 years have passed since van Leeuwenhoek first described living cells, yet we still do not have a complete catalog of cell types or their functions. Recently, transcriptomic profiling of individual cells has emerged as an essential tool for characterizing cellular diversity (*1–3*). Single-cell RNA-sequencing (scRNA-seq) methods have profiled tens of thousands of individual cells (*4–6*), revealing new insights about cell types within both healthy (*7–14*) and diseased tissues (*15–18*). Unfortunately, since these methods require cell sorters, custom microfluidics, or microwells, throughput is still limited and experiments are costly. We introduce split-pool ligation-based transcriptome sequencing (SPLiT-seq), a low-cost, scRNA-seq method that enables transcriptional profiling of hundreds of thousands of fixed cells or nuclei in a single experiment. SPLiT-seq does not require partitioning single cells into individual compartments (droplets, microwells, or wells) but relies on the cells themselves as compartments. The entire workflow before sequencing consists just of pipetting steps, and no complex instruments are needed.

In SPLiT-seq, individual transcriptomes are uniquely labeled by passing a suspension of formaldehyde-fixed cells or nuclei through four

rounds of combinatorial barcoding. In the first round of barcoding, cells are distributed into a 96-well plate, and cDNA is generated with an in-cell reverse transcription (RT) reaction using well-specific barcoded primers. Each well can contain a different biological sample, thereby enabling multiplexing of up to 96 samples in a single experiment. After this step, cells from all wells are pooled and redistributed into a new 96-well plate, where an in-cell ligation reaction appends a second well-specific barcode to the cDNA. The third-round barcode, which also contains a unique molecular identifier (UMI), is then appended with another round of pooling, splitting, and ligation. After three rounds of barcoding, the cells are pooled and split into sublibraries, and sequencing barcodes are introduced by polymerase chain reaction (PCR). This final step provides a fourth barcode, while also making it possible to sequence different numbers of cells in each sublibrary. After sequencing, each transcriptome is assembled by combining reads containing the same four-barcode combination (Fig. 1A and fig. S1A).

Four rounds of combinatorial barcoding can yield 21,233,664 barcode combinations (three rounds of barcoding in 96-well plates followed by a fourth round with 24 PCR reactions), enough to uniquely label over 1 million cells. Even larger numbers of barcode combinations can be achieved by performing experiments in 384-well plates or through additional rounds of barcoding (fig. S1B). In addition, by performing the first step in a 384-well plate, up to 384 different biological samples could be combined in a single experiment.

SPLiT-seq validation

To test SPLiT-seq's ability to generate uniquely barcoded cells (UBCs), we performed a species-

mixing experiment. We mixed cells from one mouse and two human cell lines (NIH/3T3, HEK293, and HeLa-S3), fixed them, and used SPLiT-seq to generate a scRNA-seq library with 1758 UBCs. The library was sequenced, and reads were aligned to a combined mouse-human genome. Nearly all (99.9%) of the UBCs were unambiguously assigned to a single species (>90% of reads aligned to a single genome), with the remaining 0.1% of UBCs representing barcode collisions between mouse and human cells (Fig. 1B). At saturating read coverage (>500,000 reads per cell), we identified a median of 15,365 UMIs and 5498 genes per human cell and 12,243 UMIs and 4497 genes per mouse cell. The species purity in both human and mouse UBCs was high: 99.6% of reads in human UBCs and 99.0% of reads in mouse UBCs aligned to their respective genomes. We also performed single-nucleus RNA-seq (snRNA-seq) experiments using SPLiT-seq with freshly prepared nuclei, as well as nuclei and cells that had been preserved at -80°C for 2 weeks. In all samples, we detected similar numbers of transcripts and genes per cell (Fig. 1C, fig. S2, and table S1). Gene expression was highly correlated between preserved and freshly prepared cells (Fig. 1D and fig. S2) (Pearson r , 0.987), as well as between cells and nuclei (fig. S2) (Pearson r , 0.952). We also examined gene and UMI detection at different sequencing depths and found that the sensitivity of SPLiT-seq is comparable to droplet-based scRNA-seq methods (fig. S3).

Single-nuclei RNA-seq of developing mouse brain and spinal cord

We used SPLiT-seq to profile nuclei from the developing brain and spinal cord of postnatal day 2 and 11 (P2 and P11) mice. The first round of barcoding assigned identifiers for the P2 brain, P2 spinal cord, P11 brain, and P11 spinal cord samples (Fig. 2A and fig. S4). In total, four rounds of barcoding ($48 \times 96 \times 96 \times 14$) generated more than 6 million distinct barcode combinations, making it possible to process hundreds of thousands of nuclei in a single experiment with minimal barcode collisions (2.5% expected collisions for 150,000 nuclei).

To determine how many transcripts SPLiT-seq detects within nuclei from the central nervous system, we performed deep sequencing on a sublibrary containing only 131 nuclei. We detected 4943 UMIs and 2055 genes per nucleus (UMI duplication, 95%). We then sequenced the rest of the library at lower depth, resulting in a median of 677 genes and 1022 UMIs per nucleus (UMI duplication, 58%) (table S2). Low-quality transcriptomes were removed from analysis (*19*), yielding 156,049 single-nucleus transcriptomes (74,862 P2 brain; 7028 P2 spinal cord; 58,573 P11 brain; 15,586 P11 spinal cord).

Unsupervised clustering grouped transcriptomes into 73 distinct clusters (*19*) (tables S3 to S5), which were visualized by t-distributed stochastic neighbor embedding (t-SNE) (Fig. 2A). Each of these 73 clusters was assigned to a cell class on the basis of expression of established

¹Department of Electrical Engineering, University of Washington, Seattle, WA, USA. ²Department of Bioengineering, University of Washington, Seattle, WA, USA. ³Allen Institute for Brain Science, Seattle, WA, USA. ⁴Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA. ⁵Institute for Stem Cell and Regenerative Medicine, Seattle, WA, USA. ⁶Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: alex.b.rosenberg@gmail.com (A.B.R.); gseelig@uw.edu (G.S.)

marker genes (Fig. 2B). Neurons accounted for 83% of the profiled transcriptomes (54 clusters), with most clusters expressing *Meg3*.

The 27,096 non-neuronal transcriptomes spanned 19 different clusters, each assigned to a specific cell type. Four astrocyte types (Fig. 2C) accounted for 50% of all non-neuronal nuclei ($n = 13,481$). Oligodendrocytes (six types, $n = 4294$) and oligodendrocyte precursor cells (OPC) (one type, $n = 5793$) formed the second most abundant population. We further identified two vascular and leptomeningeal cell (VLMC) types (fig. S5A), endothelial cells, smooth muscle cells (fig. S5B), microglia, macrophages (fig. S5C) (20, 21), ependymal cells, and olfactory ensheathing cells (OEC).

Previous work has observed that t-SNE can order cells in two-dimensional space according to stages of differentiation (9). Moving through t-SNE space along the path of differentiation can then be viewed as moving through “pseudotime” (22). As oligogenesis spans the first two postnatal weeks of murine development (23), we asked whether the oligodendrocyte and OPC clusters might reflect a continuous developmental trajectory. When we examined the oligodendrocyte clusters, we found that they formed an overlapping elongated shape in the t-SNE visualization. OPCs and oligodendrocytes from the P2 mouse were enriched at one end of the structure, whereas oligodendrocytes from the P11 mouse were enriched at the opposite end (fig. S6), indicative of a lineage (19, 22).

We then performed a more thorough analysis of this putative lineage. To ensure that our ordering of oligodendrocytes was determined exclusively by their relationship to other oligodendrocytes, rather than all cells, we re-embedded only transcriptomes within these seven clusters with t-SNE (Fig. 2D and fig. S7A). We calculated the moving average of gene expression in the resulting pseudotime ordering (Fig. 2E and fig. S8). Analysis of these expression patterns confirmed that proliferating OPCs segregated to one end of the t-SNE, whereas mature oligodendrocytes segregated to the opposite end (fig. S7B). We also detected previously reported intermediate stages of oligodendrocyte development, with the order of gene expression across pseudotime nearly identical to the one defined previously (9) (fig. S7C) (Spearman r ; 0.94). When analyzing spinal cord- and brain-derived cells separately, we found more mature oligodendrocytes in the spinal cord than in the brain (fig. S7D), indicating that oligodendrocyte maturation occurs earlier in the spinal cord.

Neuronal cell types

Using known gene markers, we were able to assign most neuronal clusters to specific cell types (19). Although some clusters corresponded to abundant cell types, such as cerebellar granule cells (CGCs), others mapped to rare and often less-characterized cell types, such as mitral/tufted cells. Previously characterized regional markers were used to assign the majority of clusters to a

specific region of the brain (24) (Fig. 3A). Regional assignments were validated with RNA in situ hybridization (ISH) from the Allen Institute’s Developing Mouse Brain Atlas (Allen DMBA) (25). Specifically, we generated composite ISH maps by averaging across the five most highly enriched genes from each of our clusters (tables S6 and S7). For clusters primarily containing P2 or P11 nuclei, we used the P4 or P14 atlases, respectively. The resulting composite maps confirmed the high regional specificity of most types (Fig. 3B and figs. S9 and S10). Cortical pyramidal neuronal types could be further assigned to specific layers using marker genes (Fig. 3C) (7, 8).

Granule cell fate in the hippocampus

In the hippocampus, immature granule cells originating in the dentate gyrus give rise not only to mature granule cells but also to pyramidal neurons (26). This process is one of two instances of neurogenesis that continues into adulthood (27), but little is known about the underlying transcriptional program. We determined that three neuronal cell types from the hippocampus likely constituted a developmental trajectory (19). Analysis of only these transcriptomes with t-SNE revealed a clear branching structure (Fig. 3D and fig. S11A). The transcription factor *Prox1*, suspected to be necessary for granule cell identity (28), was exclusively expressed in one branch,

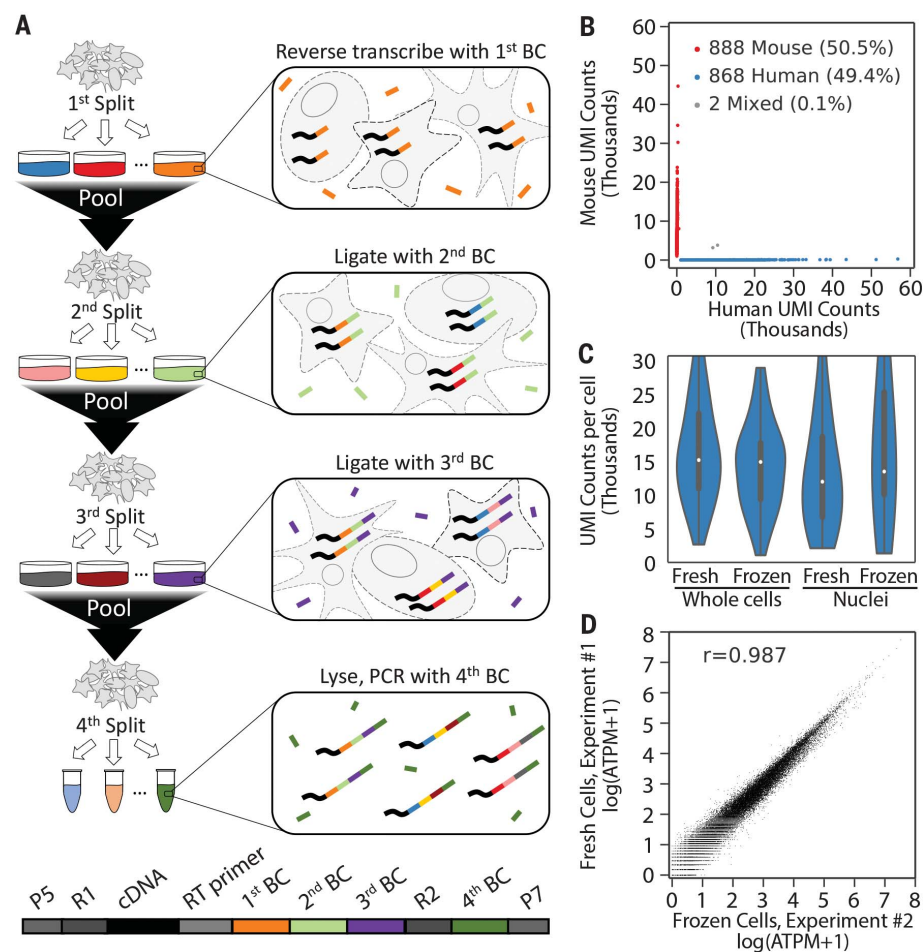


Fig. 1. Overview of SPLiT-seq. (A) Labeling transcriptomes with split-pool barcoding. In each split-pool round, fixed cells or nuclei are randomly distributed into wells, and transcripts are labeled with well-specific barcodes. Barcoded RT primers are used in the first round. Second- and third-round barcodes are appended to cDNA through ligation. A fourth barcode is added to cDNA molecules by PCR during sequencing library preparation. The bottom schematic shows the final barcoded cDNA molecule. (B) Species-mixing experiment with a library prepared from 1758 whole cells. Human UBCs are blue, mouse UBCs are red, and mixed-species UBCs are gray. The estimated barcode collision rate is 0.2%, whereas species purity is >99%. (C) UMI counts from mixing experiments performed with fresh and frozen (stored at -80°C for 2 weeks) cells and nuclei. Median human UMI counts for fresh cells: 15,365; frozen cells: 15,078; nuclei: 12,113; frozen nuclei: 13,636. (D) Measured gene expression by SPLiT-seq is highly correlated between frozen cells and cells processed immediately (Pearson r , 0.987). Frozen and fresh cells were processed in two different SPLiT-seq experiments.

whereas genes known to be specific to CA3 pyramidal neurons such as *Spock1* (29) were expressed exclusively in the other branch. Markers of dividing neuronal progenitors were expressed before the branching point, and genes in the Slit-Robo signaling pathway were differentially expressed between the two lineages (fig. S11B). We used these data to identify specific temporal dynamics of transcription factors across the two lineages, with *Meis2* as a candidate marker of early pyramidal cell differentiation (Fig. 3E and fig. S12).

Profiling cells in the developing cerebellum

The cerebellum accounts for only 9% of the brain mass in adult mice but contains nearly

85% of all neurons (30). Despite the wide range of functions performed by the cerebellum, many of the gene expression programs driving development of cerebellar cell types remain unknown. We identified the four main cerebellar neuronal types (Fig. 4A): Purkinje cells, Golgi cells, stellate/basket cells, and CGCs. Two types of Purkinje cells (Fig. 4B) were segregated primarily by age (P2 versus P11) and did not form a continuous trajectory in t-SNE but rather two clearly segregated clusters. The absence of cells at intermediate stages of maturation suggests that Purkinje cell development may be more synchronous than other processes of neurogenesis captured by our data set. CGCs, the most numerous type of neuron in the brain (31), drive the postnatal foliation of the cerebellar cortex by migrating from the external

granule layer (EGL) through the molecular layer (ML) and the Purkinje cell layer (PcL) to the internal granule layer (IGL) (32, 33). We created a pseudotime ordering of 15,360 CGCs (Fig. 4C and fig. S13) and measured gene expression across this lineage. We defined genes with specific expression at different points in pseudotime (fig. S14) and then used RNA ISH to map these genes to layers of the developing cerebellar cortex. Genes ordered from early to late in pseudotime were progressively expressed from outer to inner layers, consistent with the known direction of CGC migration (Fig. 4D). Our analysis revealed previously unknown pseudotime and layer-specific gene expression patterns within pathways related to axonal development and neuronal migration (fig. S15).

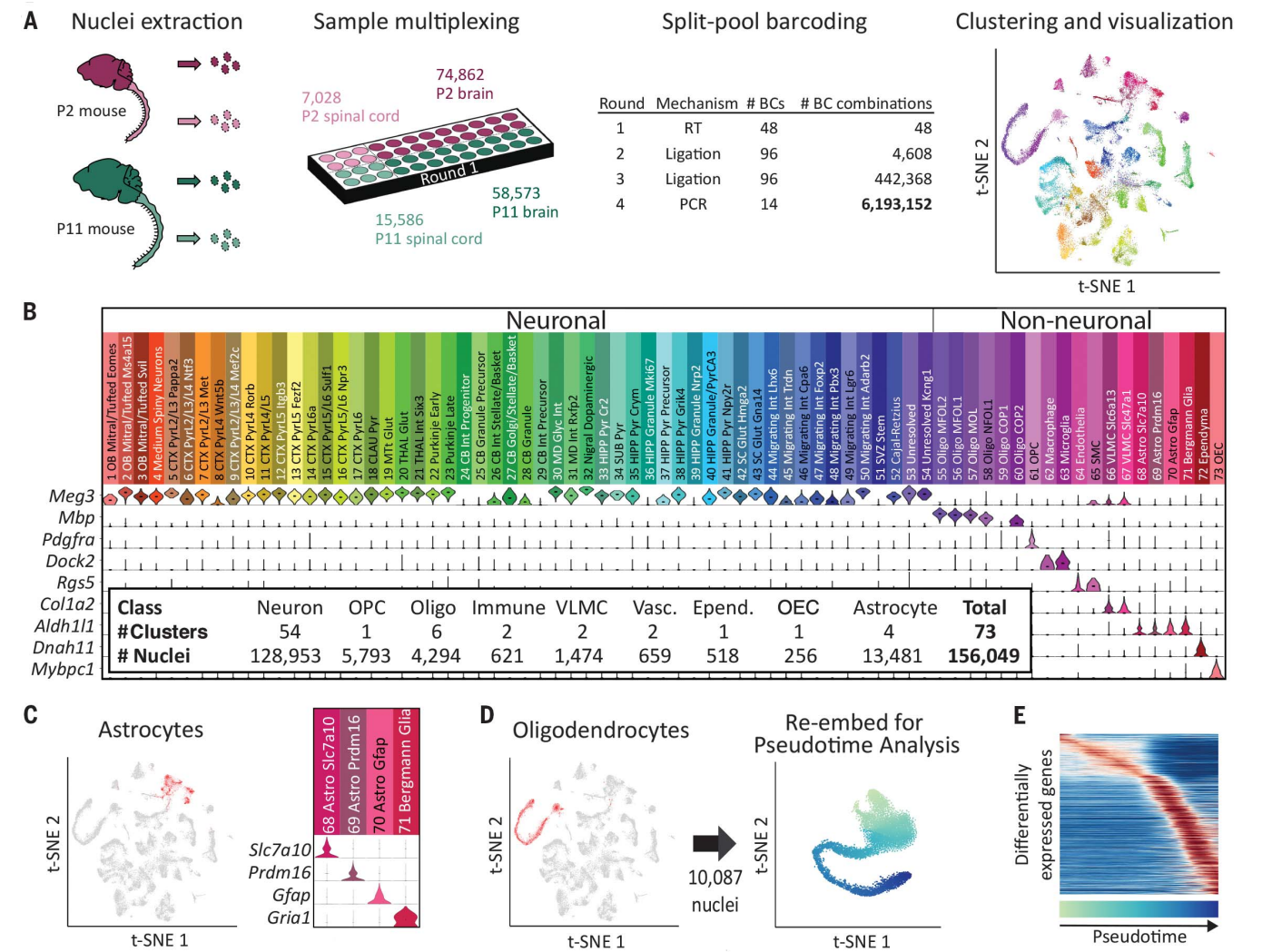


Fig. 2. Single-cell transcriptome landscape of postnatal brain and spinal cord development by SPLiT-seq. (A) More than 150,000 nuclei from P2 and P11 mouse brains and spinal cords were profiled in a single experiment employing more than 6 million barcode combinations. Transcriptomes were clustered and then visualized using t-SNE. Cells are colored according to cell type. Each cluster was downsampled to 1000 cells for visualization. (B) A total of 73 distinct clusters were assigned to nine cell classes based on expression of established markers.

The violin plots show marker gene expression in each cluster. (C) Astrocyte clusters are highlighted in red in the t-SNE. The violin plots show markers that are differentially expressed between astrocyte subtypes. (D) Seven OPC and oligodendrocyte clusters (containing 10,087 nuclei) colocalized in the original t-SNE (highlighted in red), forming a lineage. Cells from these clusters were re-embedded with t-SNE. (E) The heat map shows genes expressed differentially across pseudotime in the oligodendrocyte lineage.

Origins of cerebellar inhibitory interneurons

The question of whether all cerebellar inhibitory interneurons arise from the same progenitor population has been a point of contention (34). Early

hypotheses proposed that stellate/basket cells originated from precursors in the EGL, whereas Golgi cell precursors resided in the ventricular epithelium (35). Later evidence indicated that these two interneurons shared a common precursor in the

cerebellar white matter (36, 37). However, the molecular profile of the inhibitory neuron lineage in the cerebellum remains largely unknown.

We found a cerebellar inhibitory interneuron lineage (1517 cells) (Fig. 4E and fig. S16A) with a

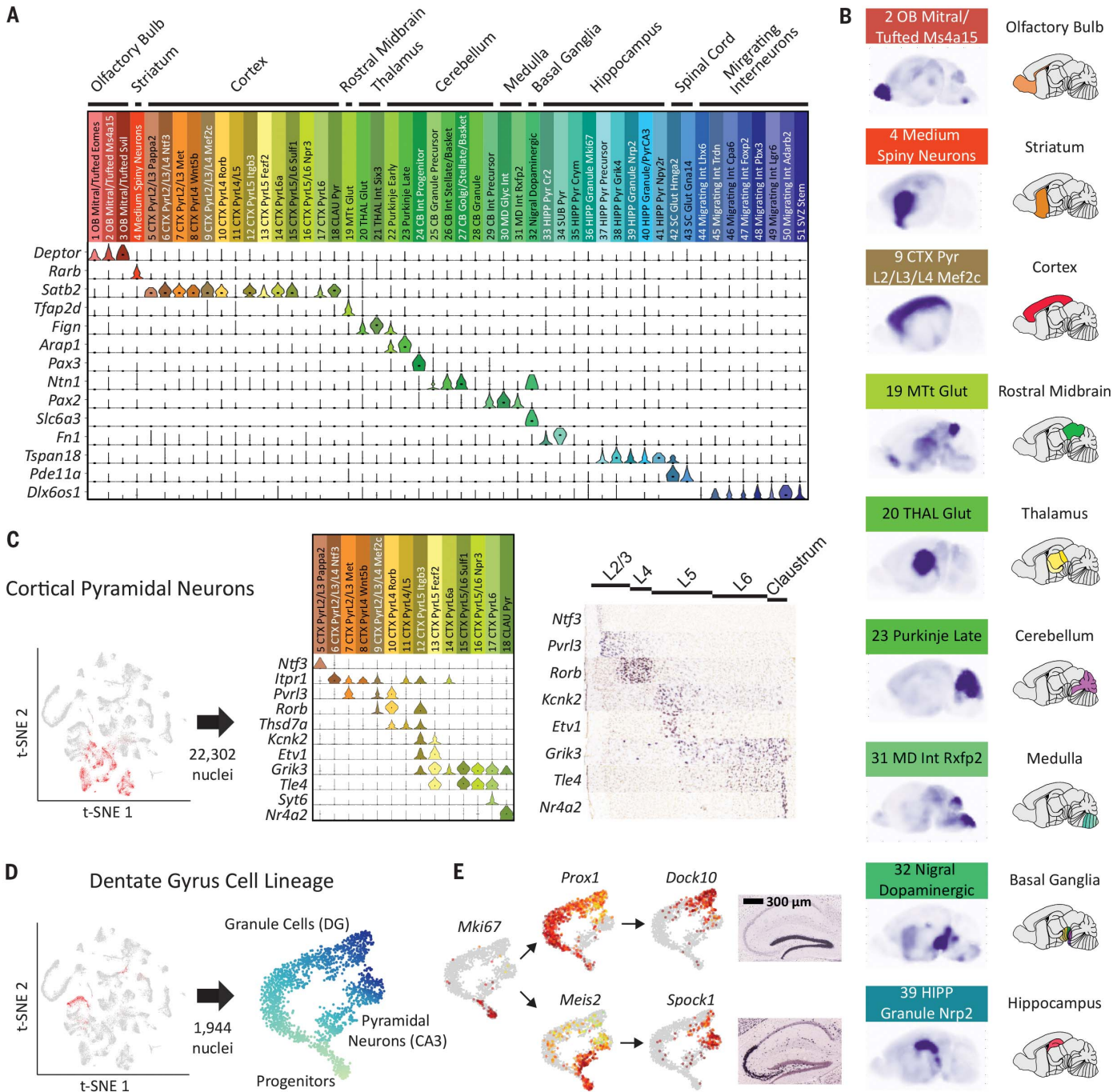


Fig. 3. Neuronal clusters exhibit regional specificity. (A) Marker gene expression was used to map neuronal clusters to specific brain regions. (B) Sagittal composite RNA ISH maps for nine representative clusters from distinct areas. For each cell type, we averaged ISH intensities from the Allen DMBA across the top five differentially expressed genes. (C) Types of pyramidal neurons in the cortex display layer-specific enrichments according to marker genes; cortical pyramidal neurons are highlighted in red in the t-SNE. Expression of example

marker genes in pyramidal clusters is shown in the middle, and corresponding available RNA ISH results are on the right. (D) Three clusters constitute a developmental trajectory in the hippocampus. Re-embedding these clusters highlights the branching of the two differentiation trajectories in pseudotime. (E) Expression of differentiation marker genes is overlaid on the t-SNE. RNA ISH maps (Allen DMBA) show the regional specificity of granule cell and pyramidal neuron markers.

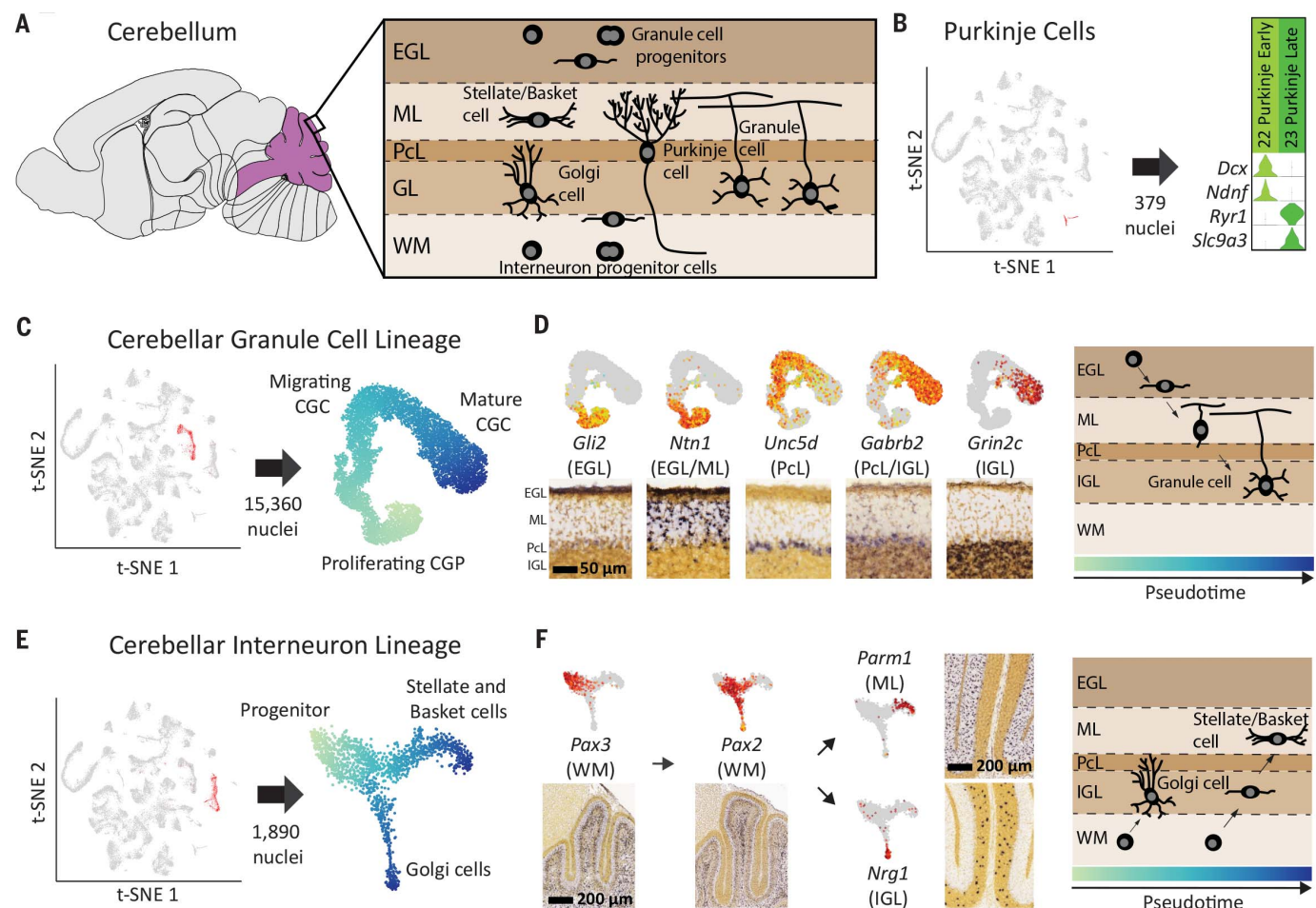


Fig. 4. Neuronal differentiation trajectories in the cerebellum

revealed by SPLiT-seq. (A) Major cell types and their locations in the cerebellum. (B) Two types of Purkinje cells with distinct gene expression programs were identified. Early Purkinje cells are primarily found in the P2 brain and late Purkinje cells in the P11 brain. (C) t-SNE re-embedding of 15,360 nuclei suggests a pseudotime ordering from proliferating, to migrating, to mature CGCs. (D) Expression of marker genes is overlaid on the t-SNE, and the corresponding RNA ISH from Allen DMBA is shown

below. Marker genes associated with different layers of the cerebellum are expressed at different points in pseudotime. Gene expression order is consistent with ordering of the physical layers. RNA ISH maps confirm regional specificity of marker genes. (E) t-SNE re-embedding of 1890 nuclei reveals a branching differentiation trajectory. Progenitors can either become Golgi cells or stellate/basket cells. (F) Markers for progenitors and mature cell types are expressed at different points in pseudotime and have layer specificity.

shared progenitor branching into either Golgi or stellate/basket cells (fig. S17). This lineage includes a known precursor cell type expressing *Pax2* (36) but also a previously unknown, earlier precursor expressing *Pax3* (Fig. 4F). RNA ISH analysis suggests that this *Pax3*⁺ precursor is located deep within the cerebellar white matter. Moreover, we found that stellate/basket cells expressed genes specific to the molecular layer, whereas Golgi cells expressed genes specific to the granule cell layer (Fig. 4F and fig. S18). The distribution of P2 and P11 nuclei within the lineage clearly demonstrated that the maturation of Golgi cells was well under way by P2 and complete by P11 (fig. S16B). In contrast, stellate/basket cells had not begun to differentiate at P2 and were still not fully mature by P11. These results indicate that the same molecularly defined precursor gives rise to two distinct interneurons at different stages of development.

Cell types in the developing spinal cord

The original clustering was dominated by cells in the brain, and many spinal cord cells did not segregate into well-defined clusters (fig. S19). To resolve more cell types in the spinal cord, we selected all the nuclei originating from the spinal cord and reclustered them (19), resulting in 44 clusters: 14 non-neuronal types (12 of which were also found in the brain) and 30 neuronal types (Fig. 5A and tables S8 to S10). We identified 11 different types of γ -aminobutyric acid-releasing (GABAergic) neurons, of which several were also glycinergic (Fig. 5B). One GABAergic type was identified as cerebrospinal fluid-contacting neurons (CSF-cNs) (38), with the other 10 types corresponding to inhibitory interneurons. Glutamatergic interneurons accounted for 15 additional types. We also identified two clusters of cholinergic motor neuron types (alpha and gamma) (39). To date, known markers exist only for gamma

motor neurons (e.g., *Esrrg*) (40); however, we identified specific markers for both alpha and gamma neurons (Fig. 5C).

To infer the spatial origin of neuronal types in the spinal cord, we identified the 10 most enriched genes in each type according to our snRNA-seq data and created composite ISH maps based on the Allen Mouse Spinal Cord Atlas (41) (Fig. 5D and fig. S20). Some interneuron subtypes appeared to originate primarily from laminae 1 to 3, with others originating from laminae 4 to 6. We found both inhibitory and excitatory neurons in each region. Motor neurons expressed genes found in laminae 9, whereas CSF-cNs were the only neuronal type expressing genes found in the central canal. These data allowed us to create an atlas of gene expression in the early spinal cord, providing a rich resource for further understanding development of the central nervous system.

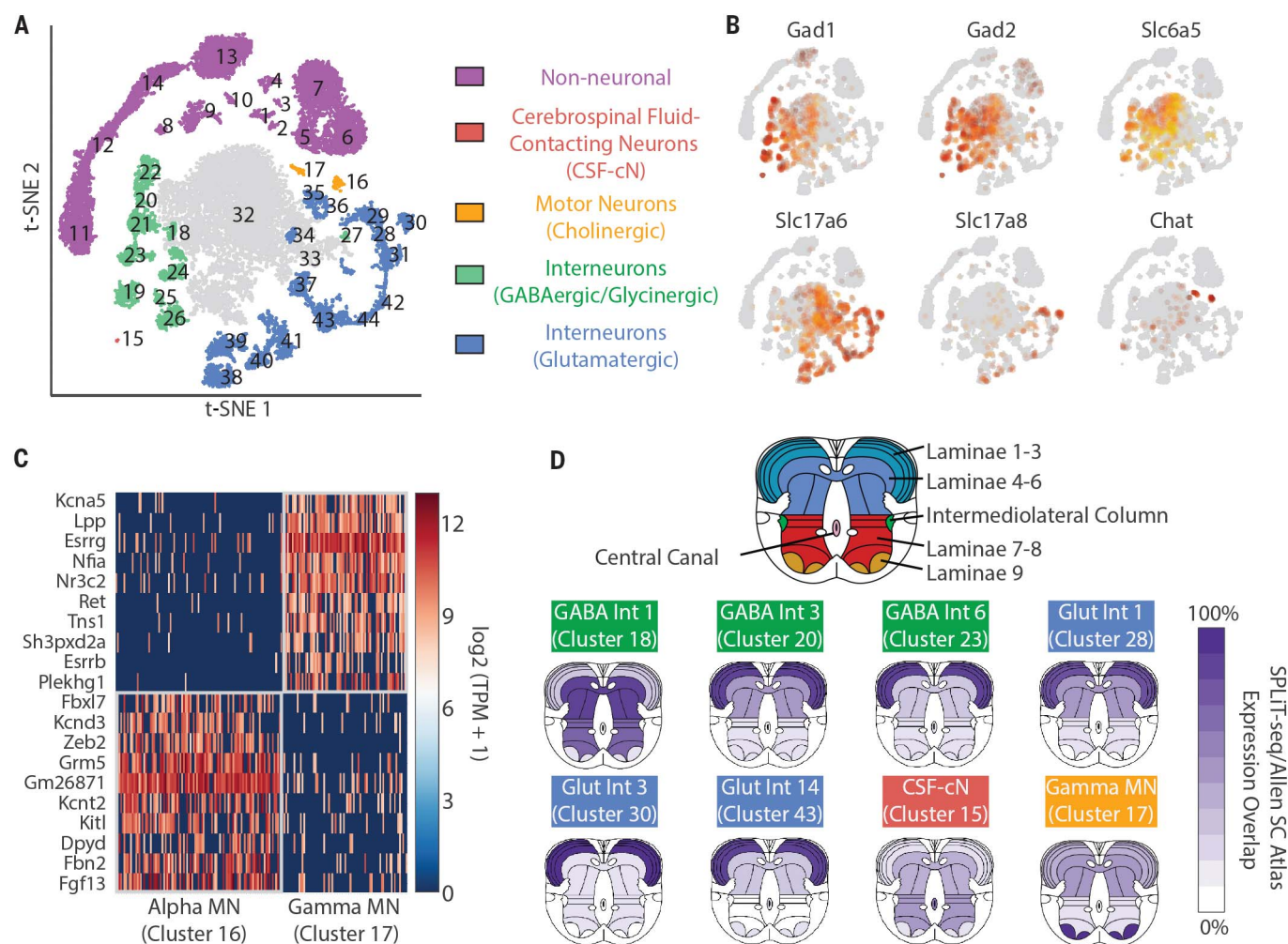


Fig. 5. Gene expression patterns and spatial origin of cell types in the spinal cord. (A) Reclustering spinal cord nuclei resulted in 30 neuronal and 14 non-neuronal clusters. (B) GABAergic neurons were defined by expression of *Gad1* and *Gad2*. A subset of GABAergic neurons are also glycinergic, based on expression of *Slc6a5*. Glutamatergic neurons were defined by expression of *VGLUT2* (*Slc17a6*), whereas cholinergic motor neurons express

Chat. (C) Novel gene markers distinguish gamma motor neurons from alpha motor neurons. (D) Inferred spatial origin of neuronal clusters within the spinal cord. We analyzed the Allen Spinal Cord Atlas expression patterns of the top 10 enriched genes in each cluster. Dark purple indicates expression of all 10 genes in the given region, whereas white indicates that none of the 10 genes were expressed in the given region.

Discussion

In this work, we profiled hundreds of thousands of cells using only basic laboratory equipment with a library preparation cost of ~\$0.01 per cell (fig. S21 and table S11). In our analysis of more than 150,000 single-nucleus transcriptomes from two early postnatal stages, we identified 69 types of cells in the brain and 44 types in the spinal cord. We defined many new molecular markers for specific cell types and explored gene expression in four different developmental lineages.

SPLiT-seq's compatibility with fixed cells and fixed nuclei overcomes challenges faced by other scRNA-seq methods. Fixation can reduce perturbations to endogenous gene expression during cell handling (42) and makes it possible to store cells for future experiments. Moreover, the use of nuclei bypasses the need to obtain intact single cells, which can be challenging for many complex tis-

sues. SPLiT-seq's compatibility with formaldehyde-fixed nuclei suggests that it may be used to profile single nuclei from formalin-fixed, paraffin-embedded tissue (43).

SPLiT-seq enables flexible and scalable cell and sample multiplexing. The use of the first-round barcode as a sample identifier makes it possible to profile a large number and variety of samples in parallel, thus minimizing batch effects. As the number of unique barcodes grows exponentially with the number of barcoding rounds, larger numbers of cells than presented here could be processed by adding a fifth barcoding round or by switching to a 384-well plate format. Although for such large cell numbers, sequencing cost may currently be forbidding, it is easy to imagine applications, such as targeted sequencing of gene panels, which would even now benefit from very large cell numbers and only require shallow sequencing depth.

Our hope is that the increased scale and accessibility provided by the low cost and minimal equipment requirements of SPLiT-seq will further accelerate the widespread adoption of scRNA-seq.

REFERENCES AND NOTES

1. S. Picelli et al., *Nat. Methods* **10**, 1096–1098 (2013).
2. T. Hashimshony, F. Wagner, N. Sher, I. Yanai, *Cell Reports* **2**, 666–673 (2012).
3. D. A. Jaitin et al., *Science* **343**, 776–779 (2014).
4. E. Z. Macosko et al., *Cell* **161**, 1202–1214 (2015).
5. A. M. Klein et al., *Cell* **161**, 1187–1201 (2015).
6. G. X. Y. Zheng et al., *Nat. Commun.* **8**, 14049 (2017).
7. B. Tasic et al., *Nat. Neurosci.* **19**, 335–346 (2016).
8. A. Zeisel et al., *Science* **347**, 1138–1142 (2015).
9. S. Marques et al., *Science* **352**, 1326–1329 (2016).
10. S. Darmanis et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7285–7290 (2015).
11. B. B. Lake et al., *Science* **352**, 1586–1590 (2016).
12. A. K. Shalek et al., *Nature* **498**, 236–240 (2013).
13. D. Grün et al., *Nature* **525**, 251–255 (2015).
14. V. Moignard et al., *Nat. Biotechnol.* **33**, 269–276 (2015).

15. A. S. Venteicher *et al.*, *Science* **355**, eaai8478 (2017).
16. I. Tirosh *et al.*, *Science* **352**, 189–196 (2016).
17. L. Sang, H. A. Collier, J. M. Roberts, *Science* **321**, 1095–1100 (2008).
18. C. Zheng *et al.*, *Cell* **169**, 1342–1356.e16 (2017).
19. Materials and methods are provided as supplementary materials.
20. S. E. Hickman *et al.*, *Nat. Neurosci.* **16**, 1896–1905 (2013).
21. O. Matcovitch-Natan *et al.*, *Science* **353**, aad8670 (2016).
22. C. Trapnell *et al.*, *Nat. Biotechnol.* **32**, 381–386 (2014).
23. S. W. Levison, J. E. Goldman, *Neuron* **10**, 201–212 (1993).
24. E. S. Lein *et al.*, *Nature* **445**, 168–176 (2007).
25. Allen Institute for Brain Science, Developing Mouse Brain Atlas (2008); available at <http://developingmouse.brain-map.org/>.
26. T. Iwano, A. Masuda, H. Kiyonari, H. Enomoto, F. Matsuzaki, *Development* **139**, 3051–3062 (2012).
27. C. Zhao, W. Deng, F. H. Gage, *Cell* **132**, 645–660 (2008).
28. A. Lavado, O. V. Lagutin, L. M. L. Chow, S. J. Baker, G. Oliver, *PLOS Biol.* **8**, e1000460 (2010).
29. F. Bonnet *et al.*, *J. Biol. Chem.* **271**, 4373–4380 (1996).
30. S. Herculanu-Houzel, *Front. Hum. Neurosci.* **3**, 31 (2009).
31. K. Nakashima, H. Umeshima, M. Kengaku, *Dev. Dyn.* **244**, 748–758 (2015).
32. A. Sudarov, A. L. Joyner, *Neural Dev.* **2**, 26 (2007).
33. J. C. Chang *et al.*, *J. Neuropathol. Exp. Neurol.* **74**, 261–272 (2015).
34. K. Schilling, J. Oberdick, F. Rossi, S. L. Baader, *Histochem. Cell Biol.* **130**, 601–615 (2008).
35. J. Altman, S. A. Bayer, *J. Comp. Neurol.* **257**, 477–489 (1987).
36. S. M. Maricich, K. Herrup, *J. Neurobiol.* **41**, 281–294 (1999).
37. G. Weisheit *et al.*, *Eur. J. Neurosci.* **24**, 466–478 (2006).
38. Y. L. Petracca *et al.*, *Development* **143**, 880–891 (2016).
39. A. Enjin *et al.*, *J. Comp. Neurol.* **518**, 2284–2304 (2010).
40. M. Lalancette-Hebert, A. Sharma, A. K. Lyashchenko, N. A. Shneider, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E8316–E8325 (2016).
41. Allen Institute for Brain Science, Allen Mouse Spinal Cord Atlas (2008); available at <http://mousespinal.brain-map.org/imageseries/showref.html>.
42. B. Lacar *et al.*, *Nat. Commun.* **7**, 11022 (2016).
43. E. R. Thomsen *et al.*, *Nat. Methods* **13**, 87–93 (2016).

ACKNOWLEDGMENTS

We thank T. N. Nguyen for help with cluster identity assignment. **Funding:** This work was supported by NIH R01CA207029 and NSF CCF-1317653 to G.S. and NIH R01NS064404 and R21NS086500 to S.H.P. Z.Y., L.T.G., and B.T. were supported by the Allen Institute for Brain Science. C.M.R. was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number TL1 TR002318. **Author contributions:** A.B.R., C.M.R., R.A.M., and G.S. were mainly

responsible for developing the method. A.B.R., C.M.R., R.A.M., A.K., P.S., S.M., W.C., and G.S. designed experiments. A.B.R., C.M.R., R.A.M., A.K., P.S., S.M., and W.C. performed experiments. Brain and spinal cord extraction was performed and supported by D.J.P., S.H.P., and D.L.S. A.B.R., C.M.R., Z.Y., and L.G. performed data analysis. Cluster annotation was performed by A.B.R., C.M.R., Z.Y., L.G., and B.T. A.B.R., C.M.R., B.T., and G.S. wrote the manuscript. **Competing interests:** A.B.R., R.M., and G.S. are inventors on a patent application (14/941,433) submitted by the University of Washington that covers the SPLiT-seq method. **Data availability:** All relevant sequencing files were deposited to the Gene Expression Omnibus under accession number GSE110823.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/176/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S12
References (44–92)

1 February 2017; resubmitted 30 September 2017
Accepted 26 February 2018
Published online 15 March 2018
10.1126/science.aam8999

TOPOLOGICAL MATTER

Observation of topological superconductivity on the surface of an iron-based superconductor

Peng Zhang,^{1*} Koichiro Yaji,¹ Takahiro Hashimoto,¹ Yuichi Ota,¹ Takeshi Kondo,¹ Kozo Okazaki,¹ Zhijun Wang,² Jinsheng Wen,³ G. D. Gu,⁴ Hong Ding,^{5,6*} Shik Shin^{1*}

Topological superconductors are predicted to host exotic Majorana states that obey non-Abelian statistics and can be used to implement a topological quantum computer. Most of the proposed topological superconductors are realized in difficult-to-fabricate heterostructures at very low temperatures. By using high-resolution spin-resolved and angle-resolved photoelectron spectroscopy, we find that the iron-based superconductor $\text{FeTe}_{1-x}\text{Se}_x$ ($x = 0.45$; superconducting transition temperature $T_c = 14.5$ kelvin) hosts Dirac-cone-type spin-helical surface states at the Fermi level; the surface states exhibit an s-wave superconducting gap below T_c . Our study shows that the surface states of $\text{FeTe}_{0.55}\text{Se}_{0.45}$ are topologically superconducting, providing a simple and possibly high-temperature platform for realizing Majorana states.

In a topological superconductor, the opening of the superconducting gap is associated with the emergence of zero-energy excitations that are their own antiparticles (1, 2). These zero-energy states, generally called Majorana zero modes or Majorana bound states (MBSs), have potential applications in quantum computing. One route to topological superconductivity is to realize a p-wave superconductor, which is an in-

trinsic topological superconductor; prominent candidates are Sr_2RuO_4 and $\text{Cu}_x\text{Bi}_2\text{Se}_3$. However, p-wave superconductivity is very sensitive to disorder, the experimental confirmation of the topological edge states is still elusive, and any application is highly challenging (3–5). Another way is to realize s-wave superconductivity on spin-helical states (6), such as in a topological insulator or a semiconductor with Rashba spin-split states

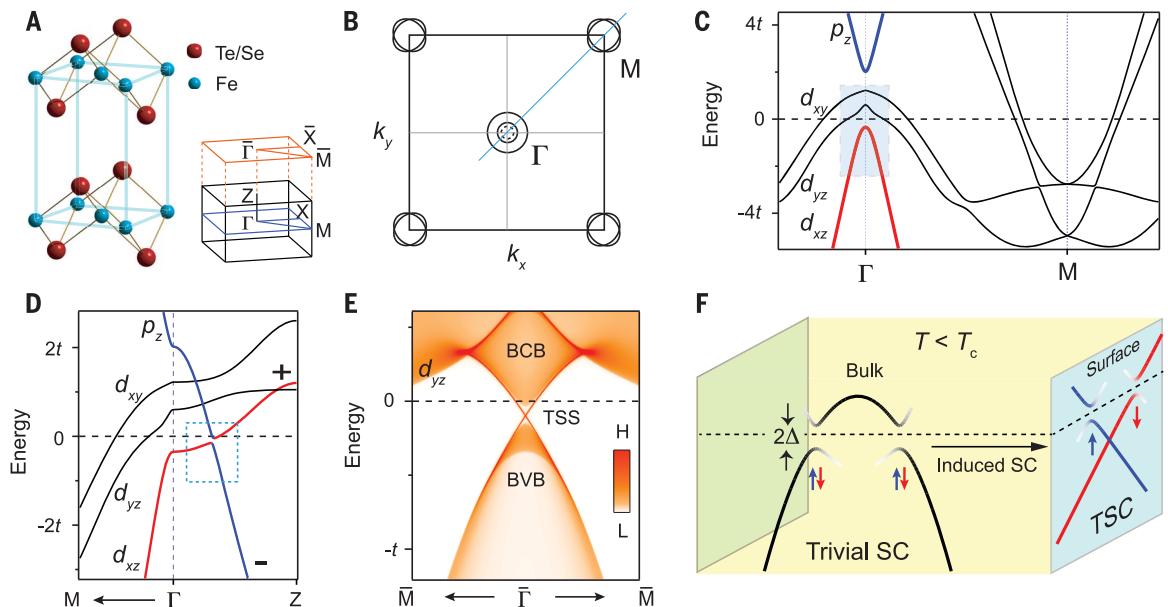
in proximity to a Bardeen-Cooper-Schrieffer (BCS) superconductor; some of the designs in this category have yielded strong experimental evidence of MBSs (7–11). However, this approach generally requires a long superconducting coherence length, which in principle prohibits the use of high-temperature superconductors. Additionally, the complicated heterostructures make further exploration and applications challenging. In this work, we show that the Fe-based superconductor $\text{FeTe}_{0.55}\text{Se}_{0.45}$, which can have a relatively high superconducting transition temperature T_c under certain conditions, hosts topological superconducting states on its surface, in accordance with theoretical predictions (12–14). This intrinsic topological superconductor, which takes advantage of the natural surface and interband superconducting coherence in the momentum space, can overcome the disadvantages of other implementations, paving a distinct route for realizing topological superconductivity and MBSs at higher temperatures.

First-principles calculations

$\text{Fe}(\text{Te},\text{Se})$ has the simplest crystal structure among Fe-based superconductors (Fig. 1A), making it easy to obtain high-quality single crystals and thin films. Its T_c can reach ~ 30 K under pressure (15) and exceeds 40 K in monolayer thin films (16). Its in-plane electronic structure is similar to that of most of the iron-based superconductors: There are two hole-like Fermi surfaces (FSs) at the Brillouin zone (BZ) center (Γ) and two electron-like FSs at the BZ corner (M) (Fig. 1B). For a cut along ΓM , there are three hole-like bands (two of

Fig. 1. Band structure and topological superconductivity of $\text{FeTe}_{0.5}\text{Se}_{0.5}$.

(A) Crystal structure of $\text{Fe}(\text{Te},\text{Se})$, together with the three-dimensional Brillouin zone (BZ) and projected-surface BZ. (B) Sketch of the in-plane BZ at $k_z = 0$ (\mathbf{k} is the wave vector in reciprocal space). There are two hole-like FSs at Γ and two electron-like FSs at M. The dashed circle at Γ indicates a hole-like band just below E_F . (C) First-principles calculations of band structure along the ΓM direction (20), indicated by the light blue line in (B). In the calculations, the energy scale $t = 100$ meV, whereas experiments yield $t \sim 12$ to 25 meV, depending on the bands (20). In this study, we focused on the small area around Γ shaded in light blue, where mainly the d_{xz} band is present. (D) First-principles calculations of band structure along ΓM and ΓZ . The dashed box shows the SOC gap of the inverted bands. (E) Band structure projected onto the (001) surface. The topological surface states (TSSs) between the bulk valence band (BVB) and bulk conduction band (BCB) are evident. H, high intensity; L, low intensity. (F) Superconducting (SC) states in the bulk and on the surface. The blue and red arrows illustrate the spin directions. The bulk states are spin-degenerated (black curves), whereas the TSSs are spin-polarized (blue and red curves). Below T_c , the bulk states open s-wave superconducting gaps, which are topologically trivial because of their spin degeneracy. Induced by the bulk-to-surface proximity, the TSSs open an s-wave gap and are topologically superconducting (TSC) as a consequence of the spin polarization (6). (The side surface is shown for convenience.)



them crossing the Fermi level (E_F) at Γ and two electron-like bands at M. Band calculations for the out-of-plane electronic structure predict that $\text{FeTe}_{0.5}\text{Se}_{0.5}$ has a nontrivial topology and hosts topological surface states near E_F (12–14).

Calculations show that the topological order originates from the Te substitution, which not only introduces large spin-orbit coupling (SOC) (17) but also shifts the p_z band downward to E_F (12), whereas the p_z band in FeSe or iron pnictides is generally above E_F (18, 19). Figure 1D shows the calculated band structure along ΓM and ΓZ (20). Along ΓZ , the p_z band has a large dispersion; near E_F , SOC causes an avoided crossing with the d_{xz} band, and a SOC gap opens. Further analysis shows that the p_z band has an odd parity (–) for

the inversion symmetry, whereas the d_{xz} band has an even parity (+). We note that the d_{xz} band consists of mixed d_{xz}/d_{yz} orbital characters along ΓZ . With these necessary ingredients, the calculated nontrivial topological invariance confirms that $\text{FeTe}_{0.5}\text{Se}_{0.5}$ hosts strong topological surface states near E_F (12). To show the predicted topological surface states clearly, we project the band structure onto the (001) surface in Fig. 1E. The Dirac-cone-type surface states are located near E_F , inside the SOC gap between the bulk valence band and the bulk conduction band. When $\text{FeTe}_{0.5}\text{Se}_{0.5}$ enters the superconducting state with s-wave gaps, superconductivity will be induced on the topological surface states, as shown in Fig. 1F. The spin polarization and s-wave superconductivity together would make the surface states topologically superconducting (6).

Dirac-cone-type spin-helical surface band and s-wave superconducting gap

To experimentally prove that $\text{FeTe}_{1-x}\text{Se}_x$ ($x \sim 0.5$) is a topological superconductor with intrinsic topological surface states and s-wave superconductivity on the surface, one needs to observe the following three phenomena in spectroscopic measurements: (i) Dirac-cone-type surface states; (ii) helical spin polarization of the surface states, which locks the spin direction perpendicular to the momentum

direction; and (iii) an s-wave superconducting gap of the surface states when $T < T_c$. Previously, we obtained some experimental evidence for the band inversion of the bulk p_z and d_{xz} bands (12, 21). However, the topological surface band was never directly observed, owing to the small energy and momentum scales. The SOC gap is estimated to be about 10 meV in the calculations, which makes it extremely difficult to resolve the Dirac-cone-type surface states in angle-resolved photoelectron spectroscopy (ARPES). In the previous ARPES experiments, only the three t_{2g} (d_{xy} , d_{yz} , and d_{xz}) and the p_z bulk bands were observed at Γ (12, 22, 23). In the experiment that we present below, by using ARPES with high energy and momentum resolution (HR-ARPES; energy resolution ~ 1.4 meV) (24) and spin-resolved ARPES (SARPES; energy resolution ~ 5.5 meV) (25), we were able to observe the three necessary phenomena required for the proof of topological superconductivity in high-quality single crystals of $\text{FeTe}_{0.55}\text{Se}_{0.45}$.

We first demonstrated the observation of the Dirac-cone-type surface states. High-resolution cuts of the band structure around Γ with p- and s-polarized photons are shown in Fig. 2, A and C, respectively. According to the matrix element effect [part I of (20)], both the surface and the bulk bands (p_z and d_{xz}) should be visible for p-polarized photons, whereas only the bulk valence

¹Institute for Solid State Physics, University of Tokyo, Kashiwa, Chiba 277-8581, Japan. ²Department of Physics, Princeton University, Princeton, NJ 08544, USA. ³National Laboratory of Solid State Microstructures and Department of Physics, Nanjing University, Nanjing 210093, China. ⁴Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973, USA. ⁵Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. ⁶CAS Center for Excellence in Topological Quantum Computation, University of Chinese Academy of Sciences, Beijing 100190, China. *Corresponding author. Email: zhangpeng@issp.u-tokyo.ac.jp (P.Z.); dingh@iphy.ac.cn (H.D.); shin@issp.u-tokyo.ac.jp (S.S.)

Fig. 2. Dirac-cone-type surface band.

(A) Band dispersion along ΓM , recorded with a p-polarized 7-eV laser. (B) MDC curvature plot of the data from (A), which enhances vertical bands (or the vertical part of one band) but suppresses horizontal bands (or the horizontal part of one band) (26). The red dots trace the points where the intensity of the MDC curvature exceeds the red bar in the color-scale indicator, and the blue lines are guides to the eye indicating the band dispersion. (C) Same as (A), but recorded with s-polarized light. The red line comes from the Lorentzian fitting of the EDC peaks. The red line is reproduced in (B) as a white line. (D and E) Zoomed-in view of the dashed box area in (A). The data are recorded at 2.4 K to reduce the thermal broadening. (D) EDCs of the zoomed-in area. The black and blue markers respectively trace the EDC peaks from two bands. arb., arbitrary. (E) EDC curvature plot of the zoomed-in area. The blue lines are the same as the ones in (B), and the red line is the same as the one in (C). (F) Summary of the overall band structure. The background image is a mix of raw intensity and EDC curvature (the area in the dashed box). The bottom hole-like band is the bulk valence band, whereas the Dirac-cone-type band is the surface band.

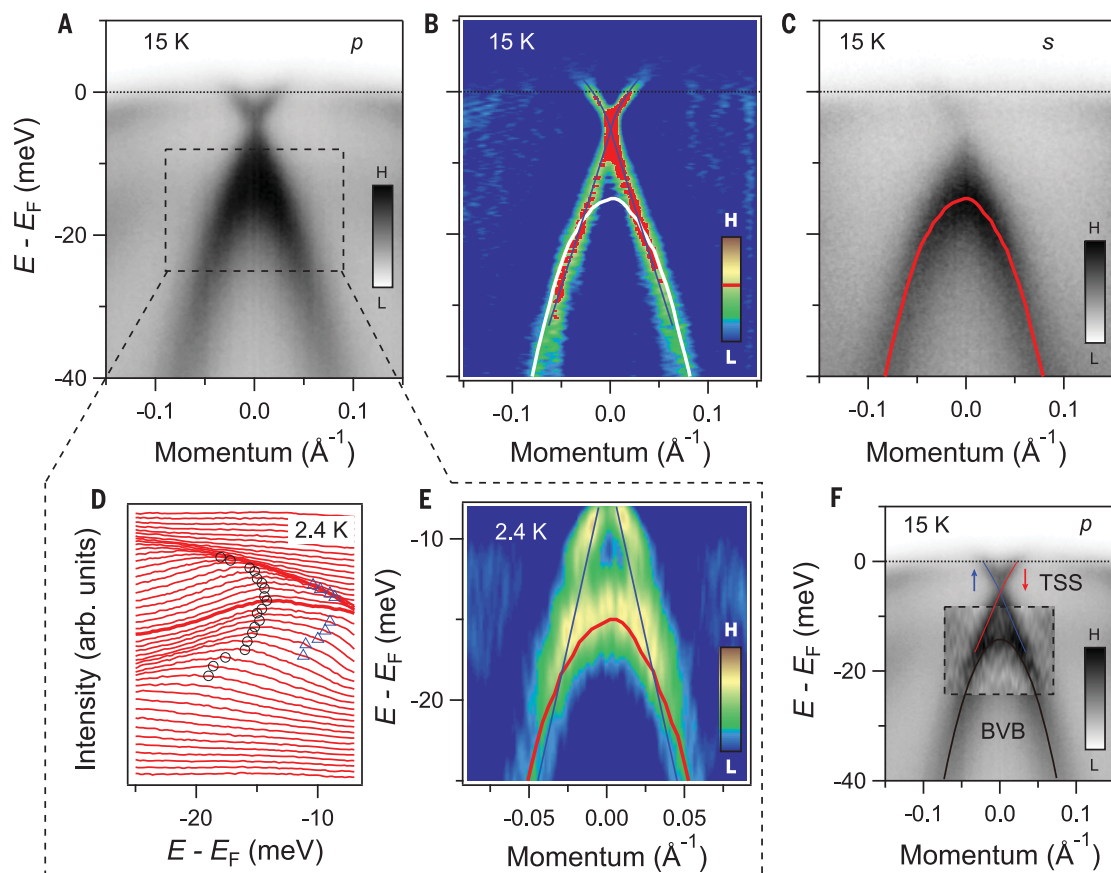
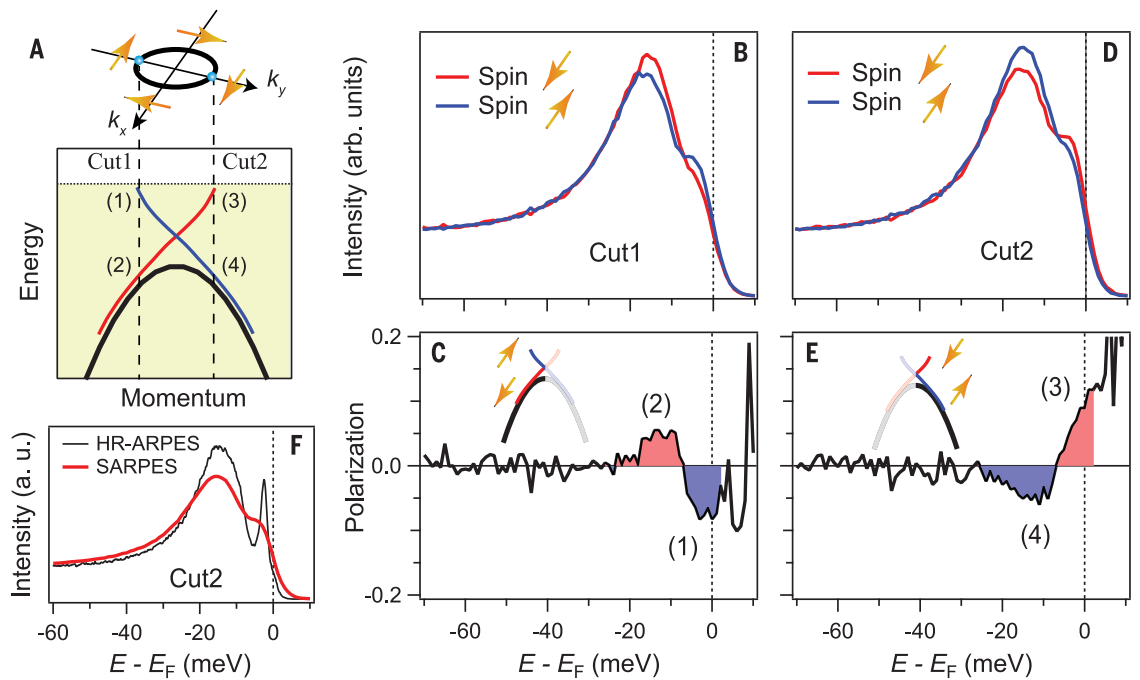


Fig. 3. Spin-helical texture of the surface band.

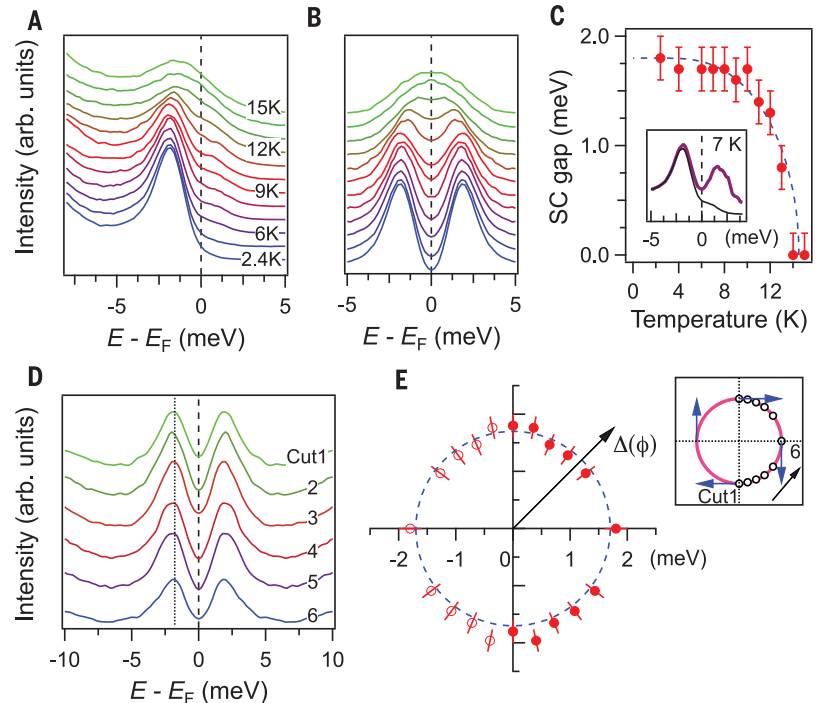
(A) Sketch of the spin-helical FS and the band structure along k_y , the sample ΓM direction. The EDCs at cuts 1 and 2 were measured with SARPES. The spin pattern shown in (12) comes from the bottom surface.

(B) Spin-resolved EDCs at cut 1. (C) Spin polarization curve at cut 1. (D and E) Same as (B) and (C), but for EDCs at cut 2.

(F) Comparison of the EDCs from SARPES and HR-ARPES measurements. The large broadening in the SARPES measurement could be partly responsible for the small spin polarization measured in (C) and (E). a. u., arbitrary units.

**Fig. 4. s-wave superconducting gap of the surface band.**

(A) Raw EDCs at different temperatures for a position on the surface FS. The shoulders above E_F are the indication of the superconducting Bogoliubov quasiparticles. (B) Symmetrized EDCs of the curves shown in (A). (C) Superconducting gap size as a function of temperature. Data points are extracted from the coherence peaks in (B); error bars come from the uncertainty of the extraction. The inset shows the raw EDC at 7 K (black) and the EDC divided by the Fermi function (purple), which shows the Bogoliubov quasiparticles above E_F . (D) Symmetrized EDCs at different Fermi wave vectors (k_F) recorded at 2.4 K. The k_F positions of the cuts are indicated in (E). (E) Polar representation of the superconducting gap size. The hollow markers mirror the solid markers. The panel on the right shows the positions of measurements on the surface FS.



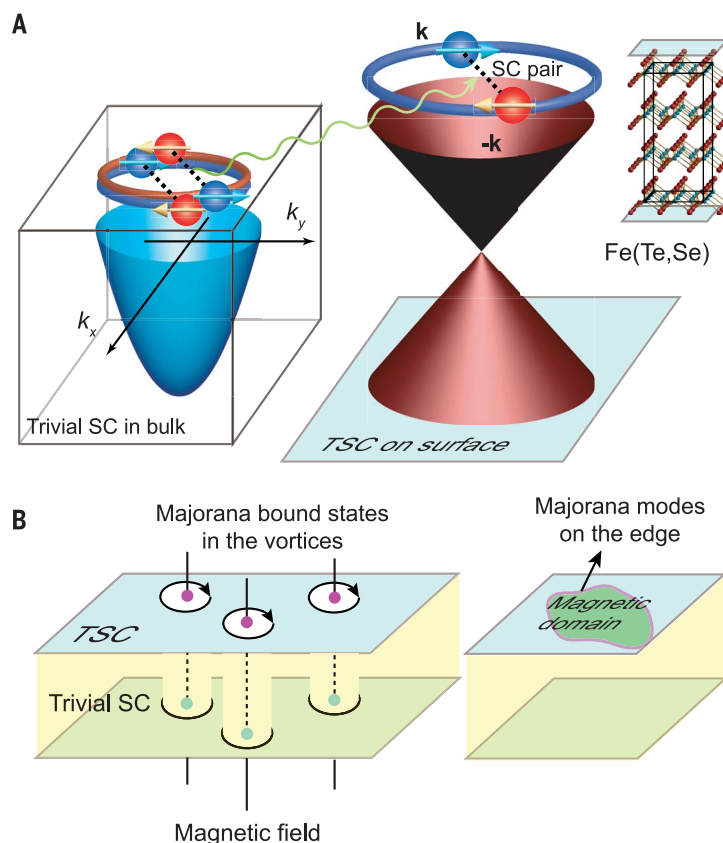
band (d_{xz}) is visible for s-polarized photons. The momentum distribution curve (MDC) curvature plot (an improved version of the second derivative method) (26) of the data with p-polarized photons shows a clear Dirac-cone-type band (Fig. 2B). We obtained a parabola-like band by extracting the energy distribution curve (EDC) peaks of the data with s-polarized photons (Fig. 2C). Combining the bands observed

in Fig. 2, A to C, we conclude that the Dirac-cone-type band (blue lines in Fig. 2B) is the topological surface band, and the parabolic band (white curve in Fig. 2B or red curve in Fig. 2C) is the bulk valence band. Further, we directly separated the bulk valence band from the Dirac-cone-type surface band with the data at very low temperature (2.4 K) when the spectral features were narrower (Fig. 2, D and E). We overlapped

the Dirac-cone-type surface band in Fig. 2B and the parabolic bulk band in Fig. 2C onto the EDC curvature plot in Fig. 2E. The extracted bands overlap well with the curvature intensity plot, confirming the existence of the parabolic bulk band and the Dirac-cone-type surface band. The overall band structure is summarized in Fig. 2F, demonstrating a Dirac surface band very close to E_F .

Fig. 5. Topological superconductivity and Majorana states on the surface.

(A) Topological superconductivity on the surface of $\text{FeTe}_{0.55}\text{Se}_{0.45}$. The electrons in the bulk are not spin-polarized, and the s-wave superconducting pairing is topologically trivial. The electrons on the surface are induced to form superconducting pairs by the bulk superconductivity. The superconductivity of the spin-helical surface states is topologically nontrivial. **(B)** A magnetic field creates vortices in $\text{FeTe}_{0.55}\text{Se}_{0.45}$, which behave as boundaries for the topological superconductivity on the surface. MBSs are expected to appear in the vortices. If there is a magnetic domain on the surface that destroys superconductivity within that domain, there will be itinerant Majorana modes along the boundary of the domain.



Next, we carried out high-resolution spin-resolved experiments to check the spin polarization of the Dirac-cone-type band. Two EDCs at the cuts indicated in Fig. 3A were measured. If the Dirac-cone-type band comes from the spin-polarized surface states, the EDCs at cuts 1 and 2 should show reversed spin polarizations. Indeed, the spin-resolved EDCs in Fig. 3, B and D, show that the spin polarizations are reversed for cuts 1 and 2, whereas the background shows no spin polarization (Fig. 3, C and E). These data are consistent with the spin-helical texture, which is the direct consequence of “spin-momentum locking” of topological surface states. We also measured an additional two EDCs at different positions on the FS [part III of (20)]. The spin polarizations of all four EDCs are consistent with the spin-helical texture predicted by theory (12). The small magnitude of the spin polarizations in Fig. 3, C and E, may partly be explained by the large broadening of the SARPES data, originating from the lower resolution of that technique (Fig. 3F).

As the final piece of evidence, we show the opening of an s-wave gap for the topological surface band. Figure 4A displays the evolution of one EDC from the surface band with temperature. The superconducting coherence peak gradually builds up with decreasing temperature; the symmetrized EDCs in Fig. 4B show the gap closing above T_c . The relation between the superconducting gap size and temperature (Fig. 4C) agrees well with BCS theory. The EDC divided by the corresponding Fermi function (Fig. 4C, inset) shows a clear peak at the symmetric position

above E_F , which comes from the particle-hole mixing of the Bogoliubov quasiparticles, thus proving the superconducting nature of the coherence peak. The momentum-dependent measurement of the superconducting gap size shows no anisotropy (Fig. 4, D and E), consistent with the s-wave superconducting nature of iron-based superconductors (27–29). The gap size of the surface band is about 1.8 meV, which is smaller than the bulk gap size of 2.5 meV for the hole band and 4.2 meV for the electron band, as reported in (27, 28). This result is consistent with induced superconductivity on the surface and may even suggest that the induced superconductivity mainly comes from interband scattering from the neighboring hole-like band.

Prospects for the observation of Majorana states

We summarize our results in Fig. 5A. A Dirac-cone-type topological surface band exists on the surface of $\text{FeTe}_{0.55}\text{Se}_{0.45}$. When the bulk bands open superconducting gaps, s-wave superconductivity is induced in the surface band through interband scattering. Because of its spin-helical texture, the surface band exhibits topological superconductivity, whereas the bulk superconductivity is topologically trivial. When an external magnetic field is applied, a pair of MBSs is expected to appear at the two ends of the vortices (Fig. 5B). This physical picture may explain the recent observations of zero-bias peaks in this material (30, 31). Furthermore, if a magnetic domain is deposited on the surface, destroying superconductivity within that

domain, there should be itinerant Majorana modes along the domain edge. As a result of the intrinsic topological superconductivity on the natural surface, it should be fairly easy to produce MBSs and Majorana edge modes. The relatively high T_c and facile growth of high-quality single crystals and thin films make $\text{Fe}(\text{Te},\text{Se})$ a promising platform for studying MBSs and may further advance research on quantum computing.

REFERENCES AND NOTES

1. X.-L. Qi, S.-C. Zhang, *Rev. Mod. Phys.* **83**, 1057–1110 (2011).
2. C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
3. L. Fu, E. Berg, *Phys. Rev. Lett.* **105**, 097001 (2010).
4. S. Sasaki et al., *Phys. Rev. Lett.* **107**, 217001 (2011).
5. N. Levy et al., *Phys. Rev. Lett.* **110**, 117001 (2013).
6. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
7. V. Mourik et al., *Science* **336**, 1003–1007 (2012).
8. S. Nadj-Perge et al., *Science* **346**, 602–607 (2014).
9. S. M. Albrecht et al., *Nature* **531**, 206–209 (2016).
10. S.-Y. Xu et al., *Nat. Phys.* **10**, 943–950 (2014).
11. H.-H. Sun et al., *Phys. Rev. Lett.* **116**, 257003 (2016).
12. Z. Wang et al., *Phys. Rev. B* **92**, 115119 (2015).
13. X. Wu, S. Qin, Y. Liang, H. Fan, J. Hu, *Phys. Rev. B* **93**, 115129 (2016).
14. G. Xu, B. Lian, P. Tang, X.-L. Qi, S.-C. Zhang, *Phys. Rev. Lett.* **117**, 047001 (2016).
15. K. Horigane, N. Takeshita, C.-H. Lee, H. Hiraka, K. Yamada, *J. Phys. Soc. Jpn.* **78**, 063705 (2009).
16. F. Li et al., *Phys. Rev. B* **91**, 220503 (2015).
17. P. D. Johnson et al., *Phys. Rev. Lett.* **114**, 167001 (2015).
18. S. Graser et al., *Phys. Rev. B* **81**, 214503 (2010).
19. H. Eschrig, A. Lankau, K. Koepf, *Phys. Rev. B* **81**, 155447 (2010).
20. Supplementary materials.
21. X. Shi et al., *Sci. Bull.* **62**, 503–507 (2017).
22. Y. Lubashevsky, E. Lahoud, K. Chashka, D. Podolsky, A. Kanigel, *Nat. Phys.* **8**, 309–312 (2012).

23. P. Zhang *et al.*, *Appl. Phys. Lett.* **105**, 172601 (2014).
24. K. Okazaki *et al.*, *Science* **337**, 1314–1317 (2012).
25. K. Yaji *et al.*, *Rev. Sci. Instrum.* **87**, 053111 (2016).
26. P. Zhang *et al.*, *Rev. Sci. Instrum.* **82**, 043712 (2011).
27. H. Miao *et al.*, *Phys. Rev. B* **85**, 094506 (2012).
28. K. Okazaki *et al.*, *Sci. Rep.* **4**, 4109 (2014).
29. D. C. Johnston, *Adv. Phys.* **59**, 803–1061 (2010).
30. F. Massee *et al.*, *Sci. Adv.* **1**, e1500033 (2015).
31. J.-X. Yin *et al.*, *Nat. Phys.* **11**, 543–546 (2015).

ACKNOWLEDGMENTS

We acknowledge C. Bareille, Y. Ishida, K. Kuroda, R. Noguchi, and A. Tsuzuki for assistance with the experiments and J. P. Hu, Z. Q. Wang, and X. X. Wu for useful discussions.

Funding: This work was supported by the Photon and Quantum Basic Research Coordinated Development Program from MEXT (Ministry of Education, Culture, Sports, Science

and Technology) of Japan, Japan Society for the Promotion of Science (KAKENHI grant nos. 25220707, 16H02209, and 16H06013); MEXT through the Innovative Area “Topological Materials Science” program (16H00979); the National Natural Science Foundation of China (11234014 and 11504117); the Ministry of Science and Technology of China (2016YFA0401000 and 2015CB921300); and the Chinese Academy of Sciences (XDPB08-1 and XDB07000000).

G.D.G. is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, under contract no. DE-SC0012704. **Author**

contributions: P.Z. performed the ARPES experiments with help from K.Y., T.H., Y.O., T.K., K.O., H.D., and S.S. Z.W. performed the first-principles calculations. J.W. and G.D.G. provided the samples. All authors discussed the manuscript. P.Z., H.D., and S.S. supervised the project. **Competing**

interests: The authors declare that they have no competing

interests. **Data and materials availability:** The data presented in this paper can be found in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/182/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S3

Table S1

References (32–41)

Data S1

16 April 2017; accepted 2 February 2018

Published online 8 March 2018

10.1126/science.aan4596

REPORT

ORGANIC CHEMISTRY

Predicting reaction performance in C–N cross-coupling using machine learning

Derek T. Ahneman,¹ Jesús G. Estrada,¹ Shishi Lin,²
Spencer D. Dreher,^{2*} Abigail G. Doyle^{1*}

Machine learning methods are becoming integral to scientific inquiry in numerous disciplines. We demonstrated that machine learning can be used to predict the performance of a synthetic reaction in multidimensional chemical space using data obtained via high-throughput experimentation. We created scripts to compute and extract atomic, molecular, and vibrational descriptors for the components of a palladium-catalyzed Buchwald–Hartwig cross-coupling of aryl halides with 4-methylaniline in the presence of various potentially inhibitory additives. Using these descriptors as inputs and reaction yield as output, we showed that a random forest algorithm provides significantly improved predictive performance over linear regression analysis. The random forest model was also successfully applied to sparse training sets and out-of-sample prediction, suggesting its value in facilitating adoption of synthetic methodology.

Machine learning (ML) is the study and construction of computer algorithms that can learn from data (1). The ability of these algorithms to detect meaningful patterns has led to their adoption across a wide range of applications in science and technology, from autonomous vehicle control to recommender systems (2). ML has also been successfully applied in the biomedical sciences to enhance the virtual screening of libraries of druglike molecules for biological function (3–5). However, its application to the chemical sciences, and synthetic organic chemistry in particular, has been limited (6, 7). Prior efforts have focused primarily on using ML to assist with synthetic planning via retrosynthetic pathways or to predict the products of chemical reactions given a set of reactants and conditions (8–11). Applications of ML to predict the performance of a given reaction, however, are rare. Studies in the area of heterogeneous catalysis have used ML to predict reaction performance when only a single component is varied (12, 13). Two recent studies have advanced the field by evaluating predictions in multidimensional chemical space, although these studies performed a binary classification of reaction success (14, 15). The use of regression-based ML to predict reaction yields in multidimensional chemical space could provide chemists with a powerful tool to navigate the adoption of synthetic methodology.

The many challenges in applying ML to reaction performance have previously hindered its use in the field of chemical synthesis. Implementation of these algorithms has historically been complicated for nonspecialists. Further, the amount of data required to obtain statistically meaningful results grows exponentially with the number of dimensions under study, a problem known as the “curse of dimensionality” (1). Given the multidimensionality of chemical structure and reactivity, it has been difficult to generate enough data or to get access to sufficiently complete and consistent data from databases to warrant implementation of these algorithms (14). Fortunately, over the past decade, high-throughput experimentation (HTE) has emerged as a powerful tool in industry and academia for reaction optimization and discovery (16, 17). We sought to evaluate whether ML could be applied to the scale of data available to modern HTE and enable yield prediction in multidimensional chemical space.

Linear regression is the traditional tool for reaction prediction and analysis in both industry and academia (18). In this approach, the user assumes a linear relationship between reaction input (e.g., catalyst descriptors) and output (e.g., product selectivity) and hand-selects input variables on the basis of specific mechanistic hypotheses (19, 20). A strength of linear regression is its interpretability: A good fit between reagent descriptors and output supports mechanistic inferences, such as in the seminal Hammett linear free-energy relationship (21).

The models obtained from linear regression analysis have also been used for prediction. Recently, Sigman and co-workers have applied multivariate linear and polynomial regression analyses

to optimize reaction selectivity by predicting catalyst, ligand, and substrate effects (22–24). Predicting yield tends to be more difficult; whereas product selectivity is determined by a small number of elementary steps, many on- and off-cycle events can substantially alter reaction yield. ML approaches accept numerous input descriptors without recourse to a mechanistic hypothesis and evaluate functions with greater flexibility to match patterns in data. We postulated that ML might outperform regression analysis for yield prediction and circumvent the challenge of selecting mechanistically relevant descriptors for large and multidimensional data sets. Here, we report that a random forest ML model trained on multidimensional chemical data can be used to predict the performance of a Buchwald–Hartwig amination reaction conducted in the presence of potentially inhibitory additives and to infer underlying reactivity. We have taken steps to automate reaction parameterization and modeling with the aim of making this tool accessible to the synthetic chemistry community.

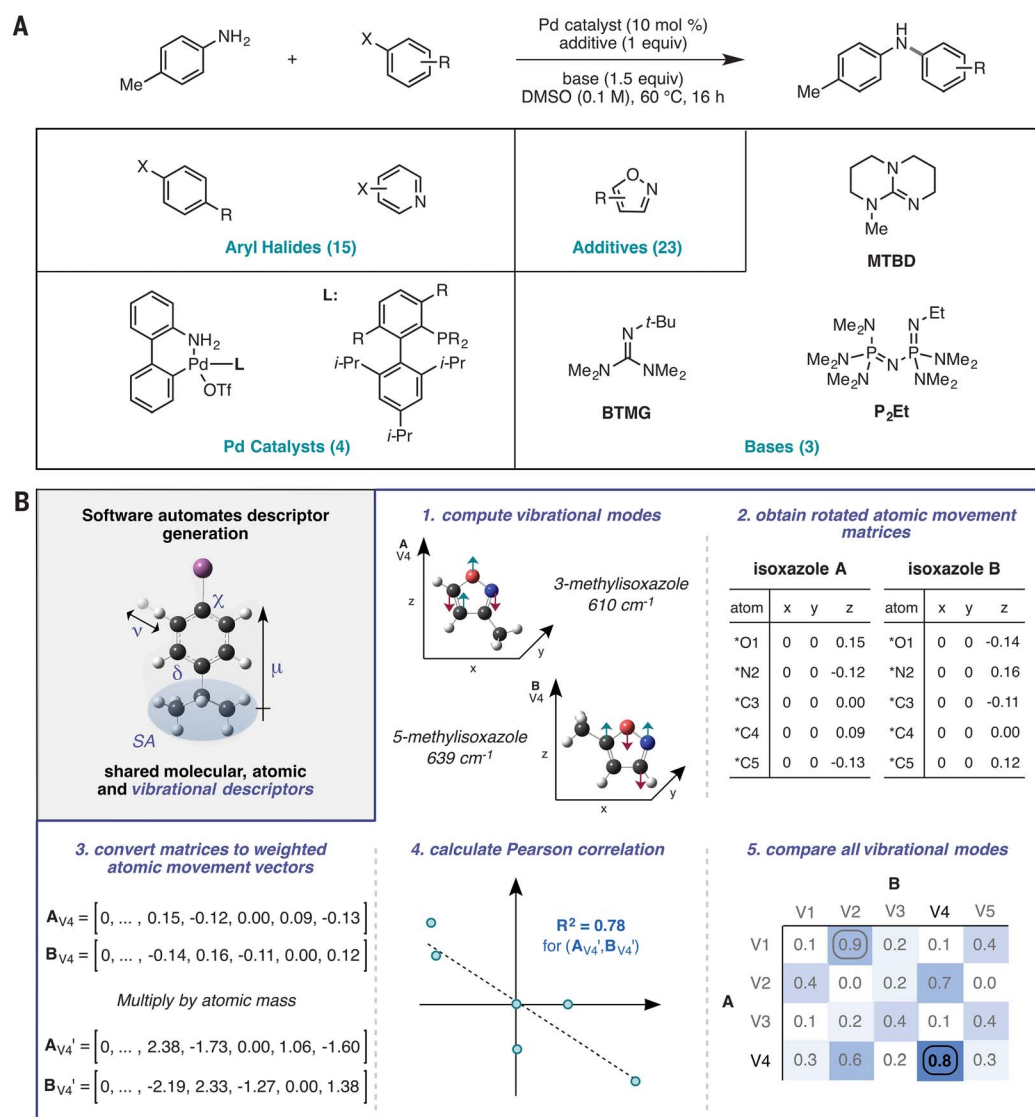
We selected the Pd-catalyzed Buchwald–Hartwig reaction as our test reaction for model development because of its broad value in pharmaceutical synthesis (Fig. 1A) (25). Nevertheless, the application of this reaction to complex drug-like molecules remains challenging (26). One limitation is the poor performance of substrates possessing five-membered heterocycles that contain heteroatom–heteroatom bonds, such as isoxazoles. These heterocycles have drug-like characteristics but are underrepresented in successful drug candidates (27). Thus, we sought to use ML to predict the performance of the Buchwald–Hartwig reaction in the presence of isoxazoles. Rather than evaluate the coupling of a collection of substrates directly bearing the heterocycle functionality, we pursued a Glorius fragment additive screening approach (28) wherein we evaluated the effects of isoxazole fragment additives on the amination of different aryl and heteroaryl halides. This method cannot always account for the full impact of a structural motif embedded within a substrate. However, the Glorius approach allowed us to test 345 diverse structural interactions between isoxazoles and aryl and heteroaryl halides. This large array would not be possible with whole molecules because of the necessity of synthesizing and isolating all possible products for quantification in this study. We conducted the coupling reactions using the ultra-high-throughput setup recently developed in the Merck Research Laboratories for nanomole-scale experimentation in 1536-well plates (16). Use of the Mosquito robot enabled simultaneous evaluation of more reaction dimensions than could previously be examined by classical statistical analysis. Three 1536-well plates consisting of a full matrix of 15 aryl and heteroaryl halides, 4 Buchwald ligands, 3 bases, and 23 isoxazole additives generated a total of 4608 reactions (including controls). The yields of these reactions were used as the model output. Approximately 30% of the reactions failed to deliver any product, with the

¹Department of Chemistry, Princeton University, Princeton, NJ 08544, USA. ²Chemistry Capabilities and Screening, Merck Sharp & Dohme Corporation, Kenilworth, NJ 07033, USA.
*Corresponding author. Email: spencer_dreher@merck.com (S.D.D.); agdoyle@princeton.edu (A.G.D.)

Fig. 1. Application of ML to reaction prediction. (A) A

Buchwald-Hartwig amination was used as a model reaction for data generation with simultaneous evaluation of four dimensions. The impact of 23 isoxazole additives on the amination reaction was investigated according to a Glorius fragment screening approach. Full structures are provided in fig. S1. Me, methyl; X, any halide; equiv, equivalent; DMSO, dimethyl sulfoxide; L, ligand; OTf, triflate; *i*-Pr, isopropyl; R, H or alkyl group; *t*-Bu, *tert*-butyl; BTMG, *t*-butyltetramethylguanidine; MTBD, methyltriazabicyclodecene; Et, ethyl. (B) Software was built to automate feature generation. Molecular, atomic, and vibrational property calculations were performed using Spartan (with density functional B3LYP and basis set 6-31G*), and these features were subsequently extracted from the resulting text files to generate a modeling data table filled with descriptors and yields. To include vibrational modes as descriptors, we compared molecular vibrations for all compounds in a class on the basis of atomic movements. To more appropriately include the movement of heavy atoms, we multiplied each atom's movement by its atomic mass. Vibrational mode vectors were compared using Pearson correlations. Only vibrational modes with $R^2 > 0.5$ and with values greater than any other entry in the same row and column

were treated as matching vibrations. If the first molecule in the set (chosen arbitrarily) shared a particular matched vibration with all others in the group, that vibrational mode was considered to be conserved. In this case, the vibration's frequency and intensity were included in the modeling data table. SA, surface area; V1 through V5, vibrational modes 1 through 5; *, shared atom.



remainder quite evenly spread over the range of yields (fig. S7).

Next we turned to the selection of appropriate descriptors. In linear regression analysis, this selection is typically done by hand according to a mechanistic hypothesis, with principal component analysis sometimes being used to reduce the parameter set to an uncorrelated and statistically tractable number (29). For the ML model, we sought a set of descriptors that adequately characterizes the differences among the reactions without recourse to a specific hypothesis. For reasons of internal consistency and descriptor availability, calculated properties were used. To avoid prohibitively time-consuming analysis and logging of computational data, we developed software to submit molecular, atomic, and vibrational property calculations to Spartan and subsequently extract these features from the resulting text files

for accessibility to a general user (Fig. 1B). The program requires only the input of reagent structures in the Spartan graphical user interface and specification of the reaction components in a Python script; it is applicable to any reaction type. The program then generates the data table that can be used for modeling. In total, 120 descriptors were extracted by the software to characterize each reaction (section III in the supplementary materials).

With these data in hand, we evaluated the predictive accuracies of linear regression and an array of ML methods using 70% of the data as a training set to predict the remaining 30% (test set) (Fig. 2A). For the linear regression models, we evaluated dimension reduction by removing correlated descriptors, as well as various regularization methods [such as LASSO (least absolute shrinkage and selection operator), ridge regression,

and elastic net], but none generated good predictive performance. Turning to supervised ML models, we found that *k*-nearest neighbors, support vector machines, and a Bayes generalized linear model provided no improvement over a linear regression model. However, a single-layer neural network delivered substantial improvement over these methods. Moreover, we found that the random forest algorithm provided even better predictive performance. The test-set root mean square error (RMSE) for the random forest model was 7.8%, with a coefficient of determination R^2 value of 0.92. A significant proportion of this variation is likely attributable to experimental and analytical error. Random forest algorithms operate by randomly sampling the data and constructing decision trees, which are then aggregated to generate an overall prediction (30). By combining a large

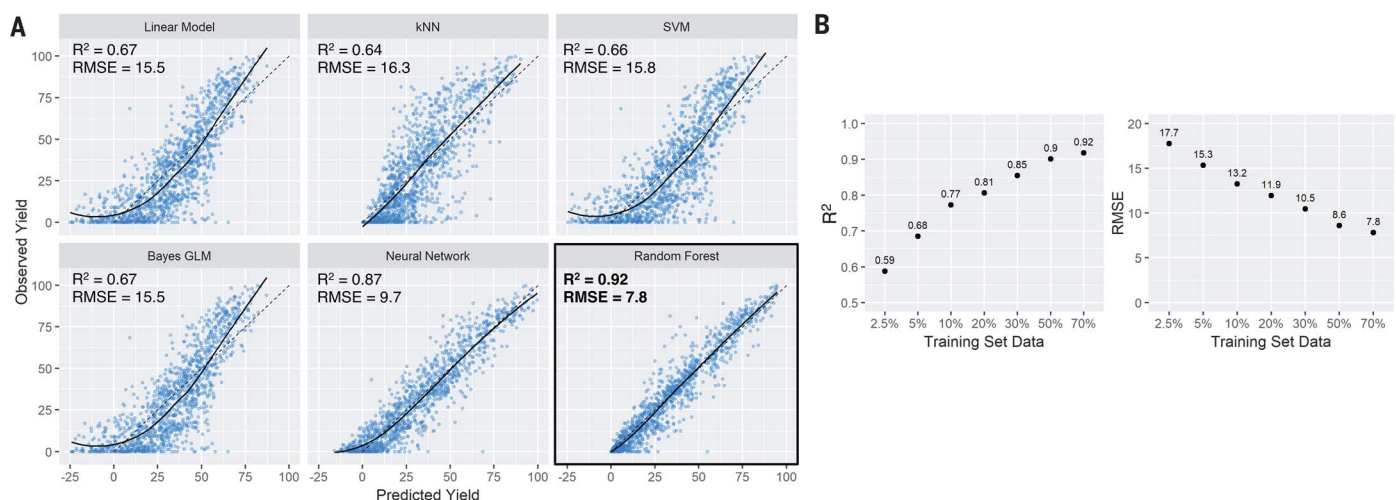


Fig. 2. Test set performance plots. (A) Observed versus predicted plots for various ML algorithms and linear regression analysis. For all the models, a 70/30 split of training and test data, with k -fold cross-validation on the training data, was performed to measure each model's generalizability to an independent data set. Only test set data are shown in plots. kNN, k -nearest neighbor; SVM,

support vector machine; GLM, generalized linear model; dashed line, $y = x$ line; solid line, Loess best-fit curve. (B) Test set performance of the random forest model with sparse data. A gradual erosion in predictive accuracy occurred from 70% of the data (the entire training set) down to 2.5% of the full data set. The smaller training sets were selected randomly from the original training data.

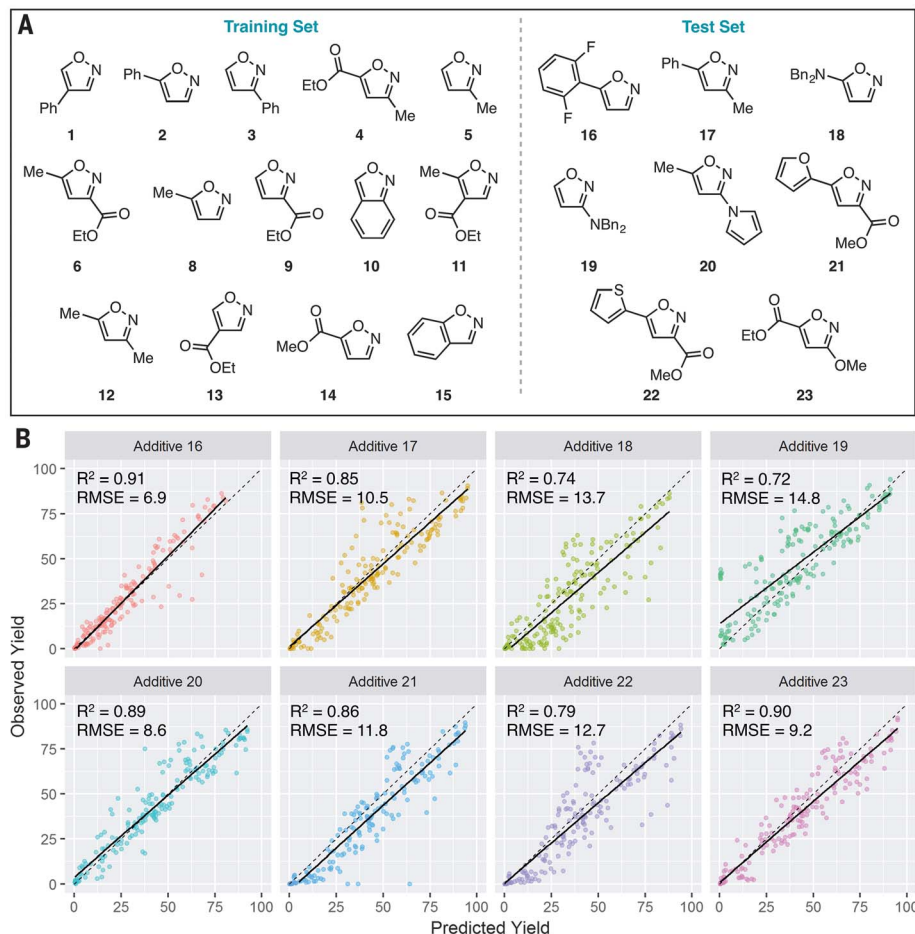


Fig. 3. Additive prediction. (A) Isoxazoles in the additive training set (1 to 6 and 8 to 15) were used to predict the performance of isoxazoles 16 to 23 in the test set. Ph, phenyl; Bn, benzyl. (B) Out-of-sample performance of the random forest model from (A). Test set data are shown.

number of low-precision models, the algorithm can deliver high predictive accuracy without succumbing to overfitting.

Nevertheless, ML tends to encounter predictive limitations when substantially different reaction conditions are used in the test set. This problem is exacerbated by the presence of activity cliffs, which are areas in reaction space where modest changes in chemical structure can lead to notable changes in reaction outcome (31). The tendency of ML algorithms to overfit and the presence of activity cliffs necessitate the collection of local reaction data (see fig. S30 for prediction of ArI and ArCl reaction outcomes from ArBr training data). One method for maximizing the extrapolative ability of a model is to use training data spread across the chemical space of interest. The ability to perform accurate prediction under sparsity effectively increases the reaction space that can be explored with the same number of experiments. For the random forest model, we were surprised to discover that enhanced predictive power over other methods could be achieved with a markedly smaller subset of the training data (Fig. 2B). With training on only 5% of the reaction data, the random forest algorithm outperformed linear regression using 70% of the same reaction data. Because 5% of the data set is only 230 experiments, these results indicate that ML can offer improvements in prediction on a scale routinely pursued in the course of reaction optimization and scope elucidation.

We next explored the ability of a random forest model to predict outcomes for reactions containing additives not included in the training data. If effective out-of-sample prediction was possible, ML could predict the effect of a new isoxazole or aryl halide structure on the outcome of a

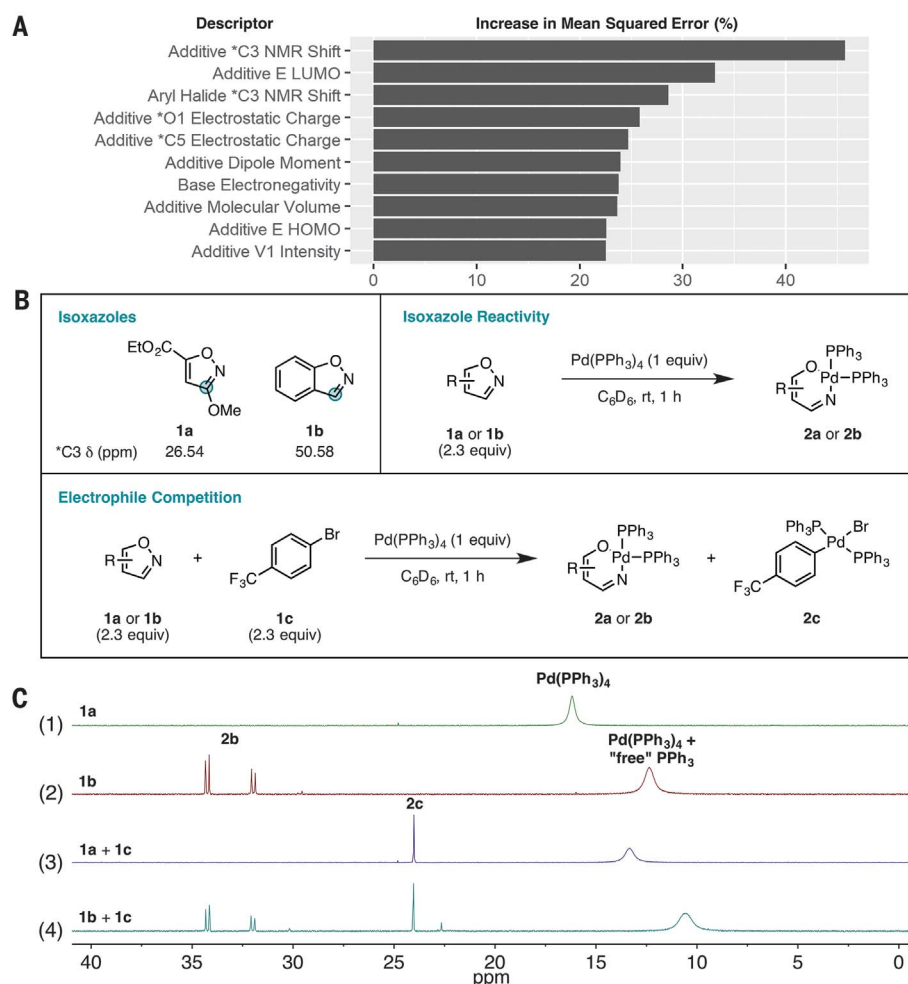


Fig. 4. Model analysis. (A) The 10 most important descriptors of the trained random forest model determined by measuring the percent increase in the MSE upon reshuffling of the values of a given descriptor and retraining of the model. * indicates a shared atom. E, energy; HOMO, highest occupied molecular orbital; V, vibration. (B) Isoxazoles and the set of reactions designed to test the hypothesis that Pd undergoes oxidative addition to certain additives, leading to diminished yield of the Buchwald-Hartwig amination. ppm, parts per million; rt, room temperature. (C) ³¹P-NMR spectra for the reactions depicted in (B). Spectrum 2 shows the generation of a new Pd species, designated **2b**, upon reaction of Pd(PPh₃)₄ with **1b**. Species **2b** is characterized by a pair of doublets with equal integration and a coupling constant (J) consistent with two *cis* phosphines ($^2J_{PP} = 37$ Hz, where $^2J_{PP}$ is the geminal phosphorus coupling constant). HRMS analysis of the reaction mixture indicates the presence of Pd(**1b**)(PPh₃)₂ (**2b**, [M + 1]⁺ = 750.13).

Buchwald-Hartwig amination and identify the combination of base and ligand that would deliver the highest yield. To this end, we evaluated whether the results for 15 additives could be used to predict the outcomes with 8 distinct additives (Fig. 3A). On average, the out-of-sample RMSE was 11.3%, with an *R*² value of 0.83 (Fig. 3B). None of the additives created significant systematic deviations from what was predicted by the model. The high predictive ability of the model suggests that the effects of these substituents on reaction outcome were captured well by the descriptors. However, as additive consumption was not included in the output, the algorithm is likely to encounter predictive limitations when applied to substrates with embedded isoxazoles.

Having obtained a predictive model, we sought to determine whether it could be used to guide mechanistic analysis. Unlike a linear regression model, the random forest model is challenging to interpret directly. We therefore evaluated the relative importance of descriptors used to construct the model. One such measure of a descriptor's importance is the percent increase in the model's mean square error (MSE) when values for that descriptor are randomly shuffled and the model is retrained (*I*). We found that four of the five most important descriptors in predicting reaction outcomes were the additive's *C-3 nuclear magnetic resonance (NMR) shift (where the asterisk indicates a shared atom), lowest unoccupied molecular orbital (LUMO) energy,

and *O-1 and *C-5 electrostatic charges (Fig. 4A). These features are not sufficient to obtain a predictive linear model (fig. S24). Taken together, the descriptors suggest that the propensity of the additive to act as an electrophile influences reaction outcomes (32–34). We hypothesized that competitive oxidative addition of the isoxazole could be a source of deleterious side reactivity. Although oxidative addition of Pd to isoxazoles is not known (35), such an elementary step has been reported previously for other transition metals (36).

To evaluate this proposal, we conducted a series of experiments with isoxazoles **1a** and **1b**, which possess the smallest and largest predicted *C-3 NMR chemical shifts of the additives in the test set, respectively (Fig. 4B). As shown in Fig. 4C, spectrum 1, isoxazole **1a** underwent no reaction with tetrakis(triphenylphosphine) palladium(0) [Pd(PPh₃)₄] in benzene at room temperature. On the other hand, with isoxazole **1b**, a new species was observed within 1 hour (Fig. 4C, spectrum 2). High-resolution mass spectrometry (HRMS) and spectroscopic (³¹P, ¹³C, and ¹H NMR) analyses provided strong evidence that isoxazole **1b** underwent oxidative addition at the N–O bond (section VI in the supplementary materials). Going further, we investigated how isoxazoles **1a** and **1b** performed in competition with an aryl halide. When **1a** was mixed with aryl bromide **1c**, formation of only the aryl bromide oxidative adduct (**2c**) was observed (Fig. 4C, spectrum 3). However, when isoxazole **1b** was subjected to the same competition experiment, the oxidative adducts of both the aryl bromide **1c** and isoxazole **1b** were observed in roughly equal amounts (Fig. 4C, spectrum 4). These data are consistent with the hypothesis that electrophilic isoxazole additives can undergo N–O oxidative addition to Pd(0) as a deleterious side reaction, causing diminished yields of the desired Buchwald-Hartwig aminations. Although such a hypothesis could have been obtained by alternate means, this study highlights how measuring the influence of a large collection of descriptors for their predictive ability in an ML algorithm can be used to generate hypotheses for further mechanistic inquiry. Although one should be hesitant to perform direct causal inference, this approach could be particularly enabling for larger and higher-dimensional data sets wherein it would be challenging or impossible to intuit a unified mechanism.

Vast resources and time are currently expended on the development of synthetic methods and their application to complex molecule synthesis, often in a largely ad hoc manner. Here we have shown that simple atomic, molecular, and vibrational descriptors that can be automatically extracted from the text files of Spartan calculations can be used as input for a random forest model to predict yields of multidimensional chemical data. We expect that this approach, coupled with advances in HTE and analysis with whole-molecule systems, will prove to be of broad utility in facilitating the adoption of synthetic methods by enabling prediction of a

new substrate's performance under given conditions or prediction of the optimal conditions for a new substrate.

REFERENCES AND NOTES

1. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).
2. M. I. Jordan, T. M. Mitchell, *Science* **349**, 255–260 (2015).
3. A. Lavecchia, *Drug Discov. Today* **20**, 318–331 (2015).
4. V. Svetnik *et al.*, *J. Chem. Inf. Comput. Sci.* **43**, 1947–1958 (2003).
5. J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, V. Svetnik, *J. Chem. Inf. Model.* **55**, 263–274 (2015).
6. M. H. Todd, *Chem. Soc. Rev.* **34**, 247–266 (2005).
7. S. Szymkuć *et al.*, *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016).
8. M. A. Kayala, C.-A. Azencott, J. H. Chen, P. Baldi, *J. Chem. Inf. Model.* **51**, 2209–2222 (2011).
9. J. N. Wei, D. Duvenaud, A. Aspuru-Guzik, *ACS Cent. Sci.* **2**, 725–732 (2016).
10. C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, K. F. Jensen, *ACS Cent. Sci.* **3**, 434–443 (2017).
11. B. Liu *et al.*, *ACS Cent. Sci.* **3**, 1103–1113 (2017).
12. S. Kite, T. Hattori, Y. Murakami, *Appl. Catal. A Gen.* **114**, L173–L178 (1994).
13. K. Omata, *Ind. Eng. Chem. Res.* **50**, 10948–10954 (2011).
14. P. Raccuglia *et al.*, *Nature* **533**, 73–76 (2016).
15. G. Skoraczynski *et al.*, *Sci. Rep.* **7**, 3582 (2017).
16. A. Buitrago Santanilla *et al.*, *Science* **347**, 49–53 (2015).
17. K. D. Collins, T. Gensch, F. Glorius, *Nat. Chem.* **6**, 859–871 (2014).
18. N. R. Draper, H. Smith, *Applied Regression Analysis* (Wiley, 1998).
19. M. S. Sigman, K. C. Harper, E. N. Bess, A. Milo, *Acc. Chem. Res.* **49**, 1292–1301 (2016).
20. S. E. Denmark, N. D. Gould, L. M. Wolf, *J. Org. Chem.* **76**, 4337–4357 (2011).
21. L. P. Hammett, *J. Am. Chem. Soc.* **59**, 96–103 (1937).
22. E. N. Bess, A. J. Bischoff, M. S. Sigman, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14698–14703 (2014).
23. A. Milo, E. N. Bess, M. S. Sigman, *Nature* **507**, 210–214 (2014).
24. A. Milo, A. J. Neel, F. D. Toste, M. S. Sigman, *Science* **347**, 737–743 (2015).
25. P. Ruiz-Castillo, S. L. Buchwald, *Chem. Rev.* **116**, 12564–12649 (2016).
26. P. S. Kutchukian *et al.*, *Chem. Sci.* **7**, 2604–2613 (2016).
27. E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **57**, 10257–10274 (2014).
28. K. D. Collins, F. Glorius, *Acc. Chem. Res.* **48**, 619–627 (2015).
29. M. Shahlalai, *Chem. Rev.* **113**, 8093–8103 (2013).
30. L. Breiman, *Mach. Learn.* **45**, 5–32 (2001).
31. M. Cruz-Monteagudo *et al.*, *Drug Discov. Today* **19**, 1069–1080 (2014).
32. Another possible source of incompatibility is Pd-catalyzed isoxazole C–H arylation. However, these reactions favor electron-rich isoxazoles and typically require more forcing conditions (>100°C) than the amination. See (33, 34).
33. Y. Fall, C. Reynaud, H. Doucet, M. Santelli, *Eur. J. Org. Chem.* **2009**, 4041–4050 (2009).
34. M. Shigenobu, K. Takenaka, H. Sasai, *Angew. Chem. Int. Ed.* **54**, 9572–9576 (2015).
35. Y. Tan, J. F. Hartwig, *J. Am. Chem. Soc.* **132**, 3676–3677 (2010).
36. S. Yu *et al.*, *Angew. Chem. Int. Ed.* **55**, 8696–8700 (2016).

ACKNOWLEDGMENTS

We thank K. Chuang and M. Keiser of the University of California, San Francisco, for help troubleshooting the neural network implementation and K. Wu of Princeton University and R. Sheridan and Z. Peng of Merck Research Laboratories for helpful discussions. **Funding:** Financial support was provided by Princeton University, an Amgen Young Investigator Award, and a Camille Dreyfus Teacher-Scholar Award. **Author contributions:** D.T.A., J.G.E., and S.L. performed the experiments. D.T.A. wrote the code. All authors designed the experiments, analyzed the data, and wrote the manuscript. **Competing interests:** None declared. **Data and materials availability:** All code and data used to produce the reported results can be found online at <https://github.com/doylelab/rxnpredict>. Additional HTE yields and model analysis data are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/186/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S34
Tables S1 to S4
References (37–43)

27 November 2017; accepted 1 February 2018
Published online 15 February 2018
10.1126/science.aar5169

METROLOGY

Measurement of the fine-structure constant as a test of the Standard Model

Richard H. Parker,^{1*} Chenghui Yu,^{1*} Weicheng Zhong,¹ Brian Estey,¹ Holger Müller^{1,2†}

Measurements of the fine-structure constant α require methods from across subfields and are thus powerful tests of the consistency of theory and experiment in physics. Using the recoil frequency of cesium-133 atoms in a matter-wave interferometer, we recorded the most accurate measurement of the fine-structure constant to date: $\alpha = 1/137.035999046(27)$ at 2.0×10^{-10} accuracy. Using multiphoton interactions (Bragg diffraction and Bloch oscillations), we demonstrate the largest phase (12 million radians) of any Ramsey-Bordé interferometer and control systematic effects at a level of 0.12 part per billion. Comparison with Penning trap measurements of the electron gyromagnetic anomaly $g_e - 2$ via the Standard Model of particle physics is now limited by the uncertainty in $g_e - 2$; a 2.5σ tension rejects dark photons as the reason for the unexplained part of the muon's magnetic moment at a 99% confidence level. Implications for dark-sector candidates and electron substructure may be a sign of physics beyond the Standard Model that warrants further investigation.

The fine-structure constant α characterizes the strength of the electromagnetic interaction between elementary charged particles. It has been measured by various methods from diverse fields of physics (Fig. 1), and the agreement of these results confirms the consistency of theory and experiment across fields. In particular, α can be obtained from measurements of the electron's gyromagnetic anomaly $g_e - 2$ by using the Standard Model of particle physics, including quantum electrodynamics to the fifth order (involving >10,000 Feynman diagrams) and muonic as well as ha-

dronic physics (1–3). This path leads to an accuracy of 0.24 part per billion (ppb) (4–6) and was until now the most accurate measurement of α .

An independent measurement of α at comparable accuracy creates an opportunity to test the Standard Model. The most accurate of previous such measurements have been based on the kinetic energy $\hbar^2 k^2 / (2m_{\text{At}})$ of an atom of mass m_{At} that recoils from scattering a photon of momentum $\hbar k$ (3), where \hbar is Planck's constant h divided by 2π , and $k = 2\pi/\lambda$ is the laser wave number (where λ is the laser wavelength). Experiments of

this type yield \hbar/m_{At} and have measured α to 0.62 ppb (7) via the relation

$$\alpha^2 = \frac{2R_\infty m_{\text{At}}}{c} \frac{\hbar}{m_e m_{\text{At}}}$$

The Rydberg constant R_∞ is known to 0.006-ppb accuracy (6), and the atom-to-electron mass ratio (m_{At}/m_e) is known to better than 0.1 ppb for many species. Here, c represents the speed of light in vacuum.

The fundamental tool of our experiment is a matter-wave interferometer (8, 9). Similar to an optical interferometer, this apparatus splits waves from a coherent source along different paths, recombines them, and measures the resulting interference to extract the phase difference accumulated between the waves on the paths. Sequences of laser pulses are used to direct and recombine the atomic matter waves along different trajectories, to form a closed interferometer (10). The phase evolution is governed by the Compton frequency of the atoms. The probability of detecting each atom at the output of the interferometers is a function of the phase accumulated between the different paths; measurement of the total atom population in each output enables an estimate of this phase. For the Ramsey-Bordé interferometer geometry used in this experiment, the phase is proportional to the photon recoil energy and can therefore be used to measure the ratio \hbar/m_{Cs} (m_{Cs} , mass of a cesium atom) and, from that, the fine-structure constant α .

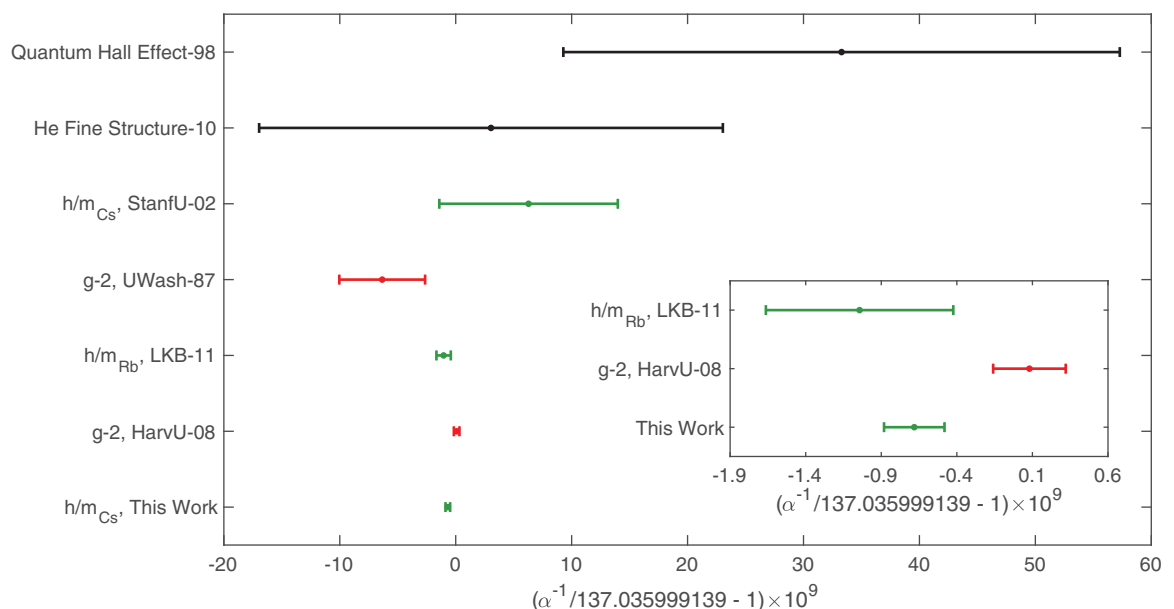
In our experiment, we used a number of methods to increase the signal and suppress systematic errors. We used 10-photon processes as beam

¹Department of Physics, 366 Le Conte Hall MC 7300, University of California, Berkeley, CA 94720, USA. ²Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: hm@berkeley.edu

Fig. 1. Precision measurements of the fine-structure constant. A comparison of measurements (1, 3–5, 7, 26–28). “0” on the plot is the CODATA 2014 recommended value (7). The green points are from photon recoil experiments; the red ones are from electron $g_e - 2$ measurements. The inset is a close-up view of the bottom three measurements. Error bars indicate 1σ uncertainty. StanfU, Stanford University; UWash, University of Washington; LKB, Laboratoire Kastler Brossel; HarvU, Harvard University.



splitters for the matter waves; these processes increase the recoil energy by a factor of 25 relative to standard two-photon Raman processes (12). To accelerate the atoms by up to another $800\hbar k$ ($400\hbar k$ up, $400\hbar k$ down), we applied a matter-wave accelerator: Atoms were loaded into an optical lattice, a standing wave generated by two laser beams, which was accelerated by ramping the frequency of the lasers (Bloch oscillations) (7, 12). Coriolis force compensation suppressed the effect of Earth's rotation. In addition, we ap-

plied ac Stark shift compensation (13, 14) and demonstrated a spatial-filtering technique to reduce sources of decoherence, further enhance the sensitivity, and suppress systematic phase shifts. An end-to-end simulation of the experiment was run (12) to help us identify and reduce systematic errors and confirm the error budget. To avoid possible bias, we adopted a blind measurement protocol, which was unblinded only at the end. Combining with precise measurements of the cesium (15) and electron (16) mass, we found

$$\alpha^{-1} = 137.035999046(27)$$

with a statistical uncertainty of 0.16 ppb and a systematic uncertainty of 0.12 ppb (0.20 ppb total). Our result is a more than threefold improvement over previous direct measurements of α (7). The measurement of $\hbar/m_{\text{Cs}} = 3.0023694721(12) \times 10^{-9} \text{ m}^2/\text{s}$ also provides an absolute mass standard in the context of the proposed new definition of the kilogram (10). This proposed definition will assign a fixed numerical value to Planck's constant, to which mass measurements could then be linked through measurements of \hbar/m_{At} , such as this one, via Avogadro spheres. Our result agrees with previous recoil measurements (7) within 1σ uncertainty and has a 2.5σ tension with measurements (4–6) based on the gyromagnetic moment.

Our matter-wave interferometer is based on the one described in (12), in which cesium atoms are loaded in a magneto-optical trap, launched upward in an atomic fountain, and detected as they fall back down—the interferometer sequence occurs during the parabolic flight. Figure 2 shows the trajectories of an atom wave packet in our experiment, formed by impulses from pairs of vertical, counterpropagating laser pulses on the atoms. Each pulse transfers the momentum of $2n = 10$ photons (where n is the order of Bragg diffraction) with near 50% probability by multiphoton Bragg diffraction, acting as a beam splitter for matter waves. Bragg diffraction allows for large momentum transfer at each beam splitter, creating a pair of atom wave packets that separate with a velocity of $\sim 35 \text{ mm/s}$. After a time interval T , a similar pulse splits the wave packets again, creating one pair that moves upward and one that moves down.

The third and fourth pulses recombine the respective paths to form two interferometers. Between the second and the third pulses, we accelerated the atom groups further from one another, using Bloch oscillations in accelerated optical lattices, to increase the sensitivity and suppress systematic effects. This transfers $+2N\hbar k$ of momentum to the upper interferometer and $-2N\hbar k$ to the lower interferometer (N , number of Bloch oscillations) (13).

The phase difference between the interferometer arms arises as a result of the kinetic energy $(\hbar k)^2/(2m_{\text{Cs}})$ that the atoms gain from the recoil momentum of the photon-atom interactions and from the phase transferred during the atoms' interaction with the laser beams. Taking the phase difference between the two interferometers cancels effects due to gravity and vibrations. In the absence of systematic effects, the overall phase Φ of the interferometer geometry shown in Fig. 2 is given by (12, 17)

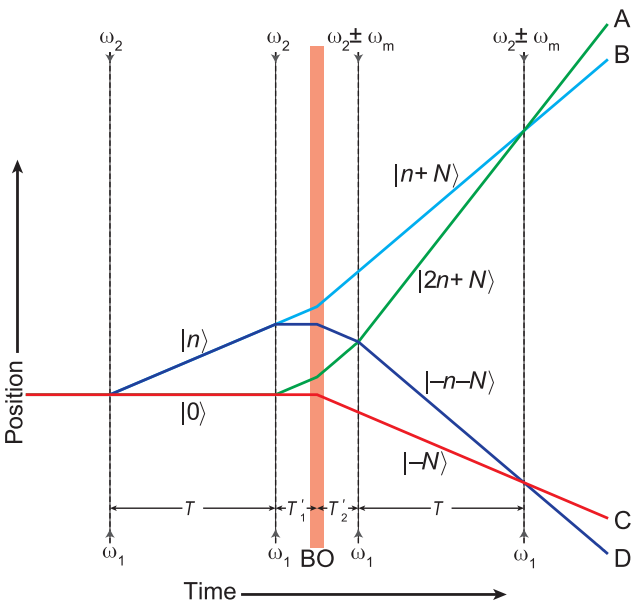
$$\Phi = \Delta\phi_1 - \Delta\phi_2 = 16n(n+N)\omega_r T - 2n\omega_m T$$

where $\Delta\phi_{1,2}$ are the measured phases of the two interferometers individually, $\omega_r = \hbar k^2/(2m_{\text{Cs}})$ is the photon recoil frequency, T is the time between the laser pulses, and ω_m is the laser frequency difference we choose to apply between

Table 1. Error budget. For each systematic effect, more discussion can be found in the listed section of the supplementary materials. N/A, not applicable.

Effect	Section	$\delta\alpha/\alpha$ (ppb)
<i>This study</i>		
Laser frequency	1	-0.24 ± 0.03
Acceleration gradient	4A	-1.79 ± 0.02
Gouy phase	3	-2.60 ± 0.03
Beam alignment	5	0.05 ± 0.03
Bloch oscillation light shift	6	0 ± 0.002
Density shift	7	0 ± 0.003
Index of refraction	8	0 ± 0.03
Speckle phase shift	4B	0 ± 0.04
Sagnac effect	9	0 ± 0.001
Modulation frequency wave number	10	0 ± 0.001
Thermal motion of atoms	11	0 ± 0.08
Non-Gaussian waveform	13	0 ± 0.03
Parasitic interferometers	14	0 ± 0.03
Total systematic error	All previous	-4.58 ± 0.12
Statistical error	N/A	± 0.16
<i>Other studies</i>		
Electron mass (16)	N/A	± 0.02
Cesium mass (6, 15)	N/A	± 0.03
Rydberg constant (6)	N/A	± 0.003
<i>Combined result</i>		
Total uncertainty in α	N/A	± 0.20

Fig. 2. Simultaneous conjugate atom interferometers. Solid lines denote the atoms' trajectories; dashed lines represent laser pulses with their frequencies indicated. $|n\rangle$ denotes a momentum eigenstate with momentum $2n\hbar k$. BO, Bloch oscillations. In this figure, gravity is neglected. A to D represent interferometer outputs.



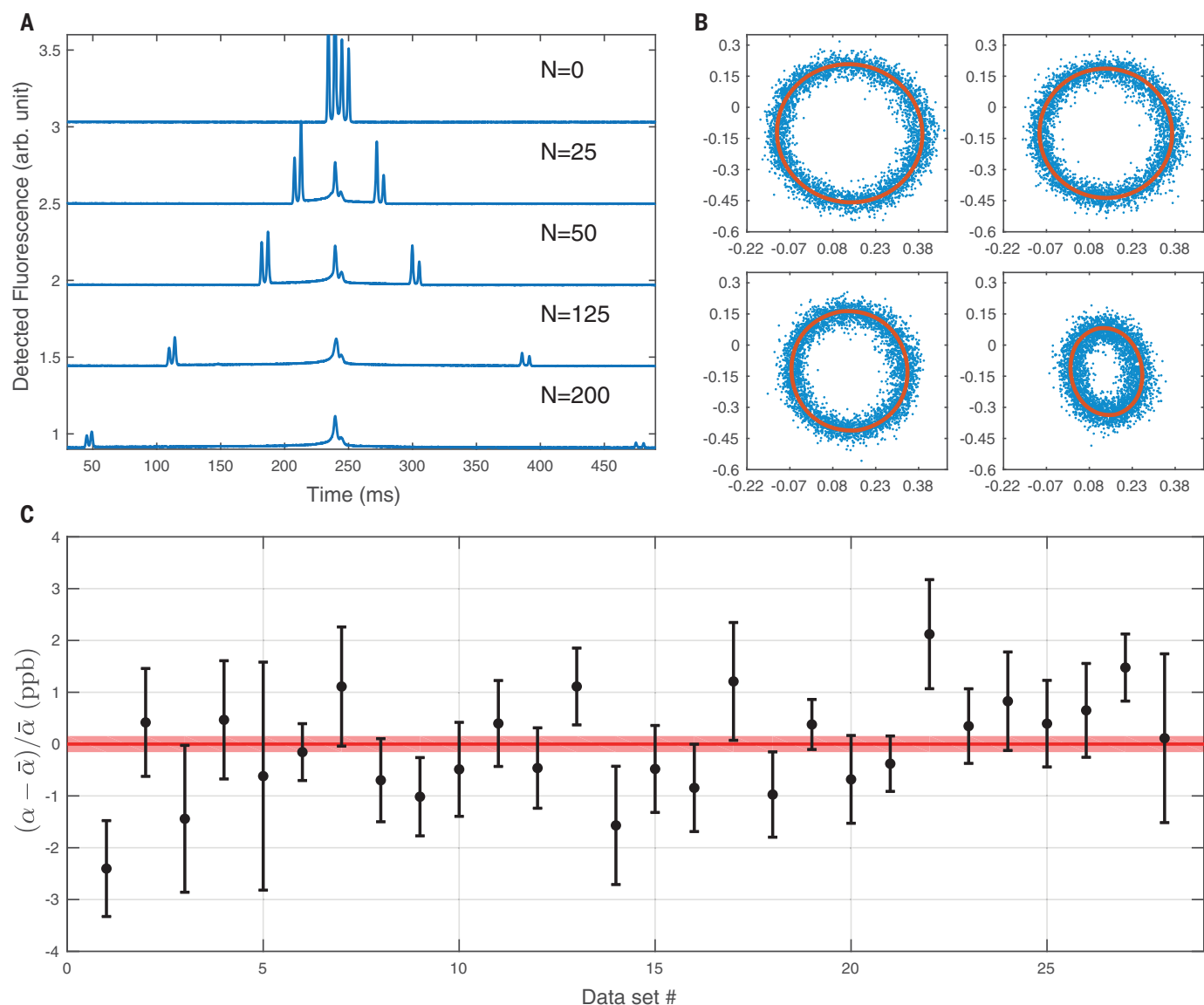


Fig. 3. Data analysis. (A) Fluorescence signals of the atom clouds as they fall through the detection region, after the interferometer sequence, for varying number N of Bloch oscillations, measured with fixed laser power and acceleration of the atoms during Bloch oscillations. For visibility, a vertical offset has been applied to each trace. The four outer peaks correspond to the four outputs A to D (Fig. 2) of the interferometers. Atoms left behind by the Bloch oscillations form the central peaks; they do not contribute to the measurement. $T = 5$ ms for these data sets. (B) Outputs of

each interferometer are normalized and plotted parametrically: the x axis is $(C - D)/(C + D)$ and the y axis is $(A - B)/(A + B)$ (A to D are defined in Fig. 2). This produces an ellipse, which is fitted to extract the differential phase. The ellipses shown are for $n = 5$, $N = 125$, and $T = 5$, 20, 40, and 80 ms (for a total interferometer phase of >10 Mrad), respectively. (C) Data sets used in the determination of α . The pink band represents the overall $\pm 1\sigma$ statistical error. The reduced χ^2 for the combined data is 1.2, with a P value of 0.2. $\bar{\alpha}$ is the weighted average of the measurements. Error bars indicate 1σ uncertainty.

the first and second pairs of pulses (Fig. 2). A measurement proceeds by adjusting ω_m to find the point where $\Phi = 0$ so that $\omega_m = 8(n + N)\omega_r$. Because the wave number k of the laser is related to the laser frequency, this yields h/m_{Cs} and, thus, α . In our measurement, $n = 5$, $N = 125$ to 200, and $T = 5$ to 80 ms, so that Φ is 10^6 to 10^7 rad and ω_m is 2 to 3 MHz.

Our error budget (Table 1) includes the systematic effects considered in the previous rubidium h/m_{Cs} measurement (7). These systematic effects are dominant, and several methods are used to reduce them (18). Our laser frequency is

monitored using a frequency comb generator. Effects caused by the finite radius of the laser beam are controlled by a retro-reflection geometry: delivering all components of the beam via the same single-mode optical fiber, using an apodizing filter to improve the Gaussian beam shape, selecting only atoms that stay close to the beam axis, and correcting for drift of the beam alignment in real time to further suppress such effects. The gravity gradient has been measured in situ for subtraction by configuring the atom interferometer as a gravity gradiometer (19–21). Keeping atoms in the same internal state while in all interfer-

ometer arms reduces the influence of the Zeeman effect to the one of an acceleration gradient, taken out by the gravity gradient measurement. The index of refraction and atom-atom interactions are reduced by the low density of our atomic sample (18).

New systematic effects arise from Bragg diffraction but can be suppressed to levels much smaller than the well-known systematics just mentioned. The potentially largest systematic is the diffraction phase Φ_0 , which we have studied in previous work (12, 13). It is caused primarily by off-resonant Bragg scattering in the third and fourth laser pulse,

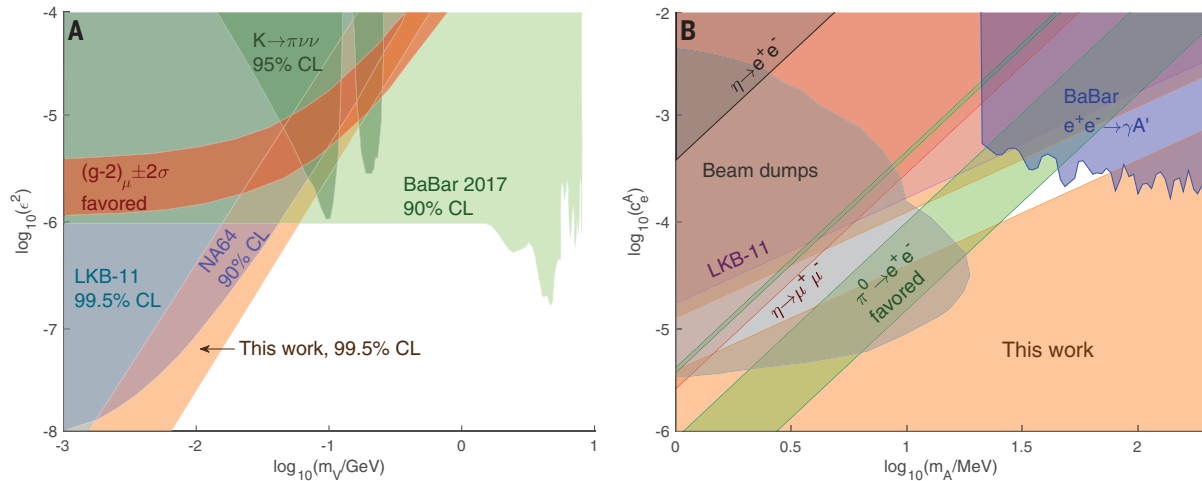


Fig. 4. Limits on dark bosons. (A) Excluded parameter space for dark photons (vector bosons), as a function of the dark-photon mass m_V and coupling suppressed by the factor ϵ . The shaded orange and blue regions are ruled out at the indicated CLs by comparing the measured a_e (4–6) with that predicted by our α measurement and the LKB-11 result, respectively (significance levels have been calculated for a one-tailed test). The red band denotes a 95% CL in which the muon $g_\mu - 2$ is

explained by a dark photon. Because our measured δa is negative, our measurement disfavors dark photons. Accelerator limits are adapted from (29). (B) Excluded parameter space for dark axial vector bosons, as a function of mass m_A and axial-vector coupling constant c_A , whose existence would produce a negative δa and is thus favored. Our work results in a two-sided bound. The region suggested by anomalous pion decay is shown in green (24) at 95% CL. Accelerator limits are adapted from (29).

where multiple frequencies for the Bragg beams are used to simultaneously address both interferometers (Fig. 2). We can therefore suppress it by using a large number N of Bloch oscillations; this increases the velocity of the atoms and thus the Doppler effect, moving the off-resonant component further off resonance. It also increases the total phase, further reducing the relative size of the systematic. The diffraction phase is nearly independent of the pulse-separation time T , so we alternate between two or more (usually six) pulse-separation times and extrapolate $T \rightarrow \infty$.

To determine the residual T -dependent diffraction phase, we employed a Monte Carlo simulation and numerically propagated atoms through the interferometer (13, 18). We ran the experiment at several different pulse-separation times, ensuring that there was no statistically significant signal for any unaccounted systematic variation. Overall, systematic errors contribute an uncertainty of 0.12 ppb to the measurement of α . As described in the supplementary materials, we corrected for systematic effects due to spatial intensity noise that have recently been pointed out (22) and for systematic effects due to deviations of the beam shape from a perfect Gaussian (18).

Figure 3C shows our data, which were collected over the course of 7 months. Each point represents roughly 1 day of data. The signal-to-noise ratio of our experiment would allow reaching a 0.2-ppb precision in less than 1 day, but extensive data were collected to suppress and control systematic effects. The measurement campaigns were interspersed with additional checks for systematic errors. Data sets typically include six different pulse-separation times, but nine data sets include only three different pulse-separation times and four data sets include four different pulse-separation times, repeated in ~ 15 -min bins; the

fit algorithm allows each bin of data to have a different diffraction phase (as the various experimental parameters may drift slowly over time) but assumes one value of h/m_{Cs} for the entire data set.

By combining our measurement with theory (5, 6), we calculated the Standard Model prediction for the anomalous magnetic moment of the electron as

$$a(\alpha) = \frac{g_e}{2} - 1 = 0.00115965218161(23)$$

Comparison with the value obtained through direct measurement (a_{meas}) (4) yielded a negative $\delta a = a_{\text{meas}} - a(\alpha) = -0.88(0.36) \times 10^{-12}$. Comparison of our result to previous measurements of α (Fig. 1) produced an error bar below the magnitude of the fifth-order quantum electrodynamics calculations used in the extraction of α from the electron $g_e - 2$ measurement and thus allows us to confront these calculations with experiment.

In addition, our measurement can be used to probe a possible substructure within the electron. An electron whose constituents have mass $m^* \gg m_e$ would result in a modification of the electron magnetic momentum by $\delta a \sim m_e/m^*$. In a chirally invariant model, the modification scales as $\delta a \sim (m_e/m^*)^2$. Following the treatment in (23), the comparison $|\delta a|$ of this measurement of α with the electron $g_e - 2$ result places a limit to a substructure at a scale of $m^* > 411,000 \text{ TeV}/c^2$ for the simple model and $m^* > 460 \text{ GeV}/c^2$ for the chirally invariant model (improvements over the previous limits of $m^* > 240,000 \text{ TeV}/c^2$ and $m^* > 350 \text{ GeV}/c^2$, respectively).

Precision measurements, such as ours, of α can also aid in the search for new dark-sector (or hidden-sector) particles (18). A hypothetical

dark photon, which is parameterized by a mixing strength ϵ and a nonzero mass m_V , for example, would lead to a nonzero δa that is a function of ϵ and m_V (24). We can test the existence of dark photons by comparing our data with the electron $g_e - 2$ measurement (4). The blue area in Fig. 4A shows the parameter space that is inconsistent with our data. We note that dark photons cause a $\delta a > 0$, opposite to the sign measured in both our experiment and the rubidium measurement (7). With the improved error of our measurement, this tension has grown. A model consisting of the Standard Model and dark photons of any m_V or ϵ is now incompatible with the data at up to a 99% confidence level (CL). Constraints on the theory obtained in this fashion (Fig. 4A) include regions not previously bounded by accelerator experiments and do not depend on the assumed decay branching ratios of the dark photon.

By contrast, a dark axial vector boson characterized by an axial vector coupling c_A and mass m_A is favored by the data because it would lead to a negative δa , but we emphasize that the 2.5σ tension in the data is insufficient to conclude the existence of a new particle (Fig. 4B). The discrepancy between the two methods of measuring α could be a hint of possible physics beyond the Standard Model that warrants further investigation. The calculated δa places limits on the axial vector parameter space from two sides. The allowed region is partially ruled out by other experiments. However, the region of parameter space consistent with our result and anomalous pion decay is also consistent with current accelerator limits, and thus the remaining region of parameter space warrants further study (24).

In particular, dark photons are one proposed explanation for the 3.4σ discrepancy in the muon

$g_\mu - 2$ with respect to the Standard Model prediction (25). As shown in Fig. 4, we rule out this explanation for nearly all values of m_V and ϵ , rejecting dark photons as an explanation for the discrepancy at the 99% CL for any dark-photon mass. The comparison of precision measurements of α and $g_e - 2$ embodies a broad probe for new physics and enables us to search for (or exclude) a plethora of other previously unidentified particles that have been proposed, such as B - L vector bosons, axial vector-coupled bosons, and scalar and pseudoscalar bosons including those that mix with the Higgs field, such as the relaxion.

REFERENCES AND NOTES

1. T. Aoyama, M. Hayakawa, T. Kinoshita, M. Nio, *Phys. Rev. Lett.* **109**, 111807 (2012).
2. S. Laporta, *Phys. Lett. B* **772**, 232–238 (2017).
3. A. Wicht, J. M. Hensley, E. Sarajlic, S. Chu, *Phys. Scr. T* **2002**, 82 (2002).
4. D. Hanneke, S. Fogwell, G. Gabrielse, *Phys. Rev. Lett.* **100**, 120801 (2008).
5. T. Aoyama, M. Hayakawa, T. Kinoshita, M. Nio, *Phys. Rev. D* **96**, 019901 (2017).
6. P. J. Mohr, D. B. Newell, B. N. Taylor, *Rev. Mod. Phys.* **88**, 035009 (2016).
7. R. Bouchendira, P. Cladé, S. Guellati-Khélifa, F. Nez, F. Biraben, *Phys. Rev. Lett.* **106**, 080801 (2011).
8. A. D. Cronin, J. Schmiedmeyer, D. E. Pritchard, *Rev. Mod. Phys.* **81**, 1051–1129 (2009).
9. G. M. Tino, M. A. Kasevich, Eds., *Atom Interferometry* (Proceedings of the International School of Physics “Enrico Fermi,” Course CLXXXVIII, Società Italiana di Fisica and IOS Press, 2014).
10. S.-Y. Lan et al., *Science* **339**, 554–557 (2013).
11. H. Müller, S. W. Chiow, Q. Long, S. Herrmann, S. Chu, *Phys. Rev. Lett.* **100**, 180405 (2008).
12. B. Estey, C. Yu, H. Müller, P.-C. Kuan, S.-Y. Lan, *Phys. Rev. Lett.* **115**, 083002 (2015).
13. R. H. Parker et al., *Phys. Rev. A* **94**, 053618 (2016).
14. T. Kovachy et al., *Nature* **528**, 530–533 (2015).
15. G. Audi et al., *Chin. Phys. C* **36**, 1157 (2012).
16. S. Sturm et al., *Nature* **506**, 467–470 (2014).
17. G. T. Foster, J. B. Fixler, J. M. McGuirk, M. A. Kasevich, *Opt. Lett.* **27**, 951–953 (2002).
18. See supplementary materials.
19. P. Asenbaum et al., *Phys. Rev. Lett.* **118**, 183602 (2017).
20. F. Sorrentino et al., *Phys. Rev. A* **89**, 023607 (2014).
21. G. Rosi et al., *Phys. Rev. Lett.* **114**, 013001 (2015).
22. S. Bade, L. Djadaoee, M. Andia, P. Cladé, S. Guellati-Khélifa, arXiv:1712.04023 [physics.atom-ph] (11 December 2017).
23. G. Gabrielse, S. F. Hoogerheide, J. Dorr, E. Novitski, in *Fundamental Physics in Particle Traps*, W. Quint, M. Vogel, Eds. (Springer Tracts in Modern Physics Series, Springer, 2014), vol. 256.
24. Y. Kahn, G. Krnjaic, S. Mishra-Sharma, T. M. P. Tait, *J. High Energy Phys.* **2017**, 2 (2017).
25. F. Terranova, G. M. Tino, *Phys. Rev. A* **89**, 052118 (2014).
26. A. Jeffery et al., *Metrologia* **35**, 83–96 (1998).
27. A.-M. Jeffery, R. E. Elmquist, L. H. Lee, J. Q. Shields, R. F. Dziuba, *IEEE Trans. Instrum. Meas.* **46**, 264–268 (1997).
28. M. Smiciklas, D. Shiner, *Phys. Rev. Lett.* **105**, 123001 (2010).
29. J. P. Lees et al., *Phys. Rev. Lett.* **119**, 131804 (2017).

ACKNOWLEDGMENTS

We acknowledge helpful discussions with O. Schwartz, P. Hamilton, M. Pospelov, S. Chu, B. Taylor, P.-C. Kuan, S.-Y. Lan, T. Tait, and E. Copenhaver. We thank R. Adhikari for being the “keeper” of the random number used in the blind analysis. We are particularly grateful to S. Guellati-Khélifa and P. Cladé for bringing to our attention the change in effective Gouy phase due to intensity variations on the beam.

Funding: This work was supported by the Alfred P. Sloan Foundation (grant BR-5044), the David and Lucile Packard Foundation (grant 2009-34712), Jet Propulsion Laboratory (grants 1458850, 1483242, 1531033, and 1553641), National Institute of Standards and Technology (award 60NANB9D9169), NSF (CAREER award PHY-1056620 and MRI award PHY-0923445), and the University of California Office of the President (award 040219). **Author contributions:** All authors contributed jointly to all aspects of this work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in this paper and the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/191/suppl/DC1
Supplementary Text
Figs. S1 to S10
Table S1
Reference (30)

24 August 2017; accepted 21 February 2018
10.1126/science.aap7706

QUANTUM INFORMATION

A blueprint for demonstrating quantum supremacy with superconducting qubits

C. Neill,^{1,*} P. Roushan,^{2,*} K. Kechedzhi,^{3,4} S. Boixo,² S. V. Isakov,² V. Smelyanskiy,² A. Megrant,² B. Chiaro,¹ A. Dunsworth,¹ K. Arya,² R. Barends,² B. Burkett,² Y. Chen,² Z. Chen,¹ A. Fowler,² B. Foxen,¹ M. Giustina,² R. Graff,² E. Jeffrey,² T. Huang,² J. Kelly,² P. Klimov,² E. Lucero,² J. Mutus,² M. Neeley,² C. Quintana,¹ D. Sank,² A. Vainsencher,² J. Wenner,¹ T. C. White,² H. Neven,² J. M. Martinis^{1,2,†}

A key step toward demonstrating a quantum system that can address difficult problems in physics and chemistry will be performing a computation beyond the capabilities of any classical computer, thus achieving so-called quantum supremacy. In this study, we used nine superconducting qubits to demonstrate a promising path toward quantum supremacy. By individually tuning the qubit parameters, we were able to generate thousands of distinct Hamiltonian evolutions and probe the output probabilities. The measured probabilities obey a universal distribution, consistent with uniformly sampling the full Hilbert space. As the number of qubits increases, the system continues to explore the exponentially growing number of states. Extending these results to a system of 50 qubits has the potential to address scientific questions that are beyond the capabilities of any classical computer.

A programmable quantum system consisting of merely 50 to 100 qubits could have a marked impact on scientific research. Although such a platform is naturally suited to address problems in quantum chemistry and materials science (1–4), applications extend to fields as diverse as classical dynamics (5) and computer science (6–9). An important milestone on the path toward realizing these applications will be the demonstration of an algorithm that exceeds the capabilities of any classical computer, thus achieving quantum supremacy (10). Sampling problems are an iconic example of algorithms designed specifically for this purpose (11–14). A successful demonstration of quantum supremacy would prove that engineered quantum systems, although still in their infancy, can outperform the most advanced classical computers.

Consider a system of coupled qubits whose dynamics uniformly explore all accessible states over time. The complexity of simulating this evolution on a classical computer is easy to understand and quantify. Because every state is equally important, it is not possible to simplify the problem by using a smaller truncated state-space. The complexity is then simply given by how much classical memory it takes to store the state vector. Storing the state of a 46-qubit system requires nearly a petabyte of memory and is at the limit of the most powerful computers (14, 15). Sampling

from the output probabilities of such a system would therefore constitute a clear demonstration of quantum supremacy. Note that this is an upper bound on only the number of qubits required—other constraints, such as computation time, may place practical limitations on even smaller system sizes.

In this study, we experimentally illustrate a blueprint for demonstrating quantum supremacy. We present data characterizing two basic ingredients required for any supremacy experiment: complexity and fidelity. First, we show that the qubits can quasi-uniformly explore the Hilbert space, providing an experimental indication of algorithm complexity [see (16) for a formal discussion of computational complexity]. Next, we compare the measurement results with the expected behavior and show that the algorithm can be implemented with high fidelity. Experiments probing complexity and fidelity provide a foundation for demonstrating quantum supremacy.

The more control a quantum platform offers, the easier it is to embed diverse applications. For this reason, we have developed superconducting gmon qubits, which are based on transmon qubits but have tunable frequencies and tunable interactions (Fig. 1A). The nine-qubit device consists of three distinct sections: control (bottom), qubits (center) and readout (top). A detailed circuit diagram is provided in (16).

Each of our gmon qubits can be thought of as a nonlinear oscillator. The Hamiltonian for the device is given by

$$\mathcal{H} = \sum_{i=1}^9 \delta_i \hat{n}_i + \frac{\eta_i}{2} \hat{n}_i (\hat{n}_i - 1) + \sum_{i=1}^8 g_i (\hat{a}_i^\dagger \hat{a}_{i+1} + \hat{a}_i \hat{a}_{i+1}^\dagger) \quad (1)$$

where \hat{n} is the number operator and \hat{a}^\dagger (\hat{a}) is the raising (lowering) operator. The qubit frequency sets the coefficient δ_i , the nonlinearity sets η_i , and the nearest-neighbor coupling sets g_i . The two lowest energy levels ($|0\rangle$ and $|1\rangle$) form the qubit subspace. The higher energy levels of the qubits, although only virtually occupied, substantially modify the dynamics. In the absence of higher levels, this model maps to free particles and can be simulated efficiently (16). The inclusion of higher levels effectively introduces an interaction and allows for the occurrence of complex dynamics.

In Fig. 1, B and C, we outline the experimental procedure and provide two instances of the raw output data. Figure 1B shows a five-qubit example of the pulses used to control the qubits. First, the system is initialized (red) by placing two of the qubits in the excited state; e.g., $|00101\rangle$. The dynamics result from fixing the qubit frequencies (orange) and simultaneously ramping all of the nearest-neighbor interactions on and then off (green). The shape of the coupling pulse is chosen to minimize leakage out of the qubit subspace (17). After the evolution, we simultaneously measure the state of every qubit. Each measurement results in a single output state, such as $|10010\rangle$; the experiment is repeated many times to estimate the probability of every possible output state. We then carry out this procedure for randomly chosen values of the qubit frequencies, the coupler pulse lengths, and the coupler pulse heights. The probabilities of the various output states are shown in Fig. 1C for two instances of the evolution after 10 coupler pulses (cycles). The height of each bar represent the probability with which that output state appeared in the experiments.

The Hamiltonian in Eq. 1 conserves the total number of excitations. This means that if we start in a state with half of the qubits excited, we should also end in a state with half of the qubits excited. However, most experimental errors do not obey this symmetry, allowing us to identify and remove erroneous outcomes. Although this symmetry helps to reduce the impact of errors, it slightly reduces the size of the Hilbert space. For N qubits, the number of states is given by the permutations of $N/2$ excitations in N qubits and is approximately $2^N / \sqrt{N}$. As an example, a 64-qubit system would access $\sim 2^{61}$ states under our protocol.

Although the measured probabilities appear largely random, they provide important insight into the quantum dynamics of the system. A key feature of these data sets are the rare, taller-than-average peaks, which are analogous to the high-intensity regions of a laser's speckle pattern. These highly likely states serve as a fingerprint of the underlying evolution and provide a means for verifying that the desired evolution was properly generated. The distribution of these probabilities provides evidence that the dynamics coherently and uniformly explore the Hilbert space.

In Fig. 2, we use the measured probabilities to show that the dynamics uniformly explore the Hilbert space for experiments carried out with

¹Department of Physics, University of California, Santa Barbara (UCSB), Santa Barbara, CA 93106, USA. ²Google, Santa Barbara, CA 93117, USA. ³Quantum Artificial Intelligence Laboratory (QuAIL), NASA Ames Research Center, Moffett Field, CA 94035, USA. ⁴Universities Space Research Association, Mountain View, CA 94043, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: cneill@google.com (C.N.); martinis@physics.ucsb.edu (J.M.M.)

five to nine qubits. We begin by measuring the output probabilities after five cycles for between 500 and 5000 distinct instances. To compare experiments with different numbers of qubits, the probabilities are weighted by the number of states in the Hilbert space. Figure 2A shows a histogram of the weighted probabilities where we find nearly universal behavior. Small probabilities ($<1/N_{\text{states}}$) appear most often, and probabilities as large as $4/N_{\text{states}}$ show up with a frequency of $\sim 1\%$. In stark contrast to this, we observe a tall, narrow peak centered at 1.0 for longer evolutions whose duration is comparable to the coherence time of the qubits.

A quantum system that uniformly explores all states is expected to have an exponential distribution of weighted probabilities. The solid line in Fig. 2A corresponds to such a distribution and is simply given by $e^{-\text{probability} \times N_{\text{states}}}$; this is also referred to as a Porter-Thomas distribution (14, 18). Although, in principle, approaching the universal form of the distribution takes exponential time, the nonuniversal deviations become small on a much shorter time scale that is linear in the number of qubits (14, 16). Note that the de-

viations from a purely exponential distribution are consistent with decoherence. The deviations scale with the number of qubits, and the histogram appears to be converging to the incoherent distribution shown in green.

A measure of algorithm complexity is a key ingredient for demonstrating quantum supremacy. We argue that evolution under the Hamiltonian in Eq. 1 cannot be efficiently simulated on a classical computer under plausible assumptions (16). The experimental results in Fig. 2A suggest that we can coherently evolve the system for long enough to realize this computationally notable regime.

In Fig. 2B, we illustrate the number of cycles necessary for the system to uniformly explore all states by comparing the measured probabilities to an exponential distribution. After each cycle, we compare the measured histogram to an exponential decay. The distance between these two distributions is measured using the Kullback-Leibler divergence D_{KL} .

$$D_{\text{KL}} = S(\rho_{\text{measured}} \cdot \rho_{\text{exponential}}) - S(\rho_{\text{measured}}) \quad (2)$$

where the first term is the cross-entropy between the measured distribution ρ_{measured} and an exponential distribution $\rho_{\text{exponential}}$, and the second term is the self-entropy of the measured distribution. The entropy of a set of probabilities is given by $S(P) = -\sum_i p_i \log(p_i)$ and the cross-entropy of two sets of probabilities is given by $S(P, Q) = -\sum_i p_i \log(q_i)$. Their difference, the Kullback-Leibler divergence, is zero if and only if the two distributions are equivalent.

We find that the experimental probabilities closely resemble an exponential distribution after just two cycles. For longer evolutions, decoherence reduces this overlap. These results suggest that we can generate very complex dynamics with only two pulses, a surprisingly small number. However, rather than breaking up the evolution into two-qubit gates, we allow the entire system to interact at once. Therefore, one of our pulses corresponds to roughly eight simultaneous two-qubit gates. Additionally, each of our pulses lasts long enough to effectively implement five square-root-of-swap gates. So, although the evolution is only two cycles, this translates to ~ 80 two-qubit gates.

In addition to demonstrating an exponential scaling of complexity, it is necessary to characterize the algorithm fidelity. Determining the fidelity requires a means for comparing the measured probabilities (P_{measured}) with the probabilities expected from the desired evolution (P_{expected}). On the basis of the proposal outlined in (14), we use the cross-entropy to quantify the fidelity

$$\frac{S(P_{\text{incoherent}}, P_{\text{expected}}) - S(P_{\text{measured}}, P_{\text{expected}})}{S(P_{\text{incoherent}}, P_{\text{expected}}) - S(P_{\text{expected}})} \quad (3)$$

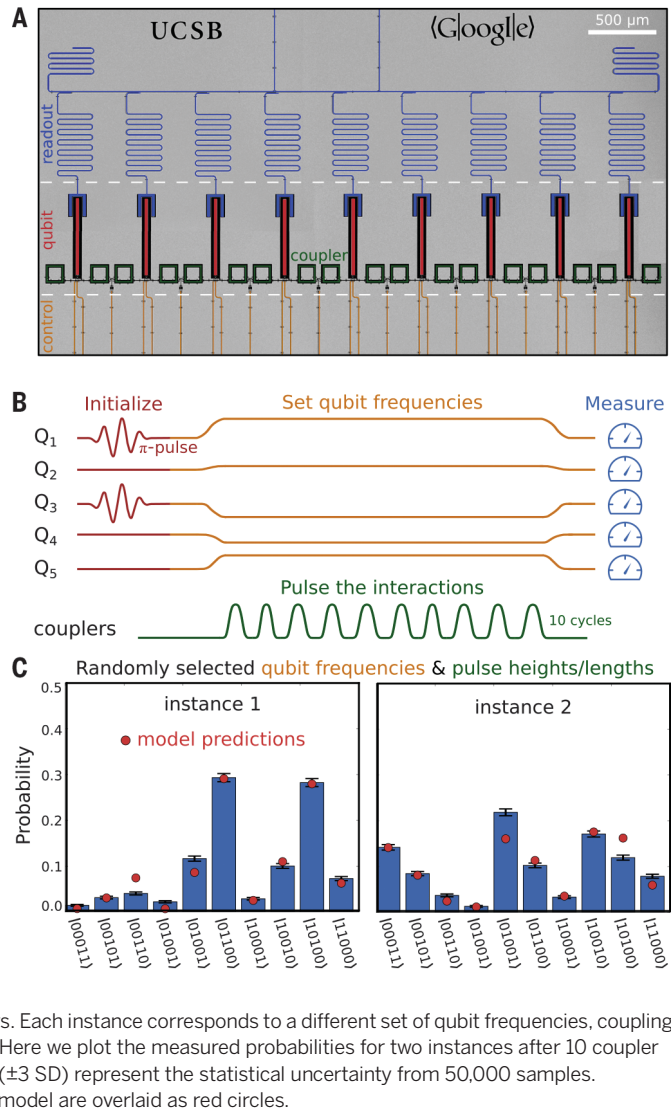
where $P_{\text{incoherent}}$ represents an incoherent mixture with each output state given equal likelihood—this is the behavior that we observe after many cycles. When the distances between the measured and expected probabilities are small, the fidelity approaches 1. When the measured probabilities approach an incoherent mixture, the fidelity approaches 0.

In Fig. 3A, we show that the desired evolution can be implemented with high fidelity. We find that at short times the fidelity decays linearly with an increasing number of cycles (fits to the data are shown as dashed lines). The slope of these lines measures the error per cycle; this slope is shown in the inset for each number of qubits. We find that the error scales with the number of qubits at a rate of $\sim 0.4\%$ error per qubit per cycle. If such an error rate extends to larger systems, we will be able to perform 60-qubit experiments of depth = 2 with a fidelity $>50\%$. These results provide promising evidence that quantum supremacy may be achievable with the use of existing technology.

Predicting the expected probabilities is a major challenge. First, substantial effort has been taken to accurately map the control currents to Hamiltonian parameters; the detailed procedure

Fig. 1. Device and experimental protocol.

(A) Optical micrograph of the nine-qubit array. Gray regions are aluminum; dark regions are where the aluminum has been etched away to define features. Colors have been added to distinguish readout circuitry, qubits, couplers, and control wiring. (B) Five-qubit example of the pulse sequences used in these experiments. First, the qubits are initialized using microwave pulses (red). Three of the qubits start in the ground state $|0\rangle$ and the rest start in the excited state $|1\rangle$. Next, the qubit frequencies are set using rectangular pulses (orange). During this time, all couplings are simultaneously pulsed (green); each pulse has a randomly selected duration. Last, we measure the state of every qubit. The measurement is repeated many times to estimate the probability of each output state. (C) We repeat this pulse sequence for randomly selected control parameters. Each instance corresponds to a different set of qubit frequencies, coupling pulse heights and lengths. Here we plot the measured probabilities for two instances after 10 coupler pulses (cycles). Error bars (± 3 SD) represent the statistical uncertainty from 50,000 samples. Predictions from a control model are overlaid as red circles.



for constructing this map is outlined in (16). Second, we model the Hamiltonian using only single-qubit calibrations, which we find to be accurate even when all of the couplers are used simultaneously. This is a scalable approach to calibration.

Third, when truncating the Hamiltonian to two levels, we find poor agreement with both an exact theoretical model and experimental results. We find that a three-level description must be used to account for virtual transitions to the second

excited state during the evolution. When including these states, truncating to a fixed number of excitations lowers the size of the computational Hilbert space from 3^N to approximately 0.15×2.42^N (table S1): Thus, a nine-qubit experiment requires accurately modeling a 414-dimensional unitary operation. Determining how many of these states are needed for sufficient accuracy depends on the magnitude of the coupling and is an open question, but the number should scale somewhere between 2.0^N and 2.5^N . The predictions are overlaid onto the data in Fig. 1C and show excellent agreement.

In Fig. 3B, we show how techniques from machine learning were used to achieve low error rates. To set the matrix elements of the Hamiltonian, we built a physical model for our gmon qubits. This model is parameterized in terms of capacitances, inductances, and control currents. The parameters in this model were calibrated using simple single-qubit experiments (16). We used a search algorithm to find offsets in the control model that minimize the error (1 - Fidelity). Figure 3B shows the error, averaged over cycles, versus the number of optimization steps. Before training the model, the data were split into two halves: a training set (red) and a verification set (black). The optimization algorithm was used only to access the training set, whereas the verification set was used only to verify the optimal parameters.

We find that the error in both the training set and the verification set fall considerably by the end of the optimization procedure. The high degree of correlation between the training and verification data suggests that we are genuinely learning a better physical model. Optimizing over more parameters does not further reduce the error. This suggests that the remaining error is not the product of an inaccurate control model but rather results from decoherence. Using the cross-entropy as a cost function for optimizing the parameters of a physical model was the key to achieving high-fidelity control in this experiment.

It is important to note how these experiments might change at the level of a few tens of qubits. At this level, it becomes exponentially unlikely that any state will appear twice, making it impractical to measure probabilities in an experiment. However, even for these large systems, sampling from the output states is sufficient to determine the fidelity (14). Therefore, the distribution of probabilities can be inferred from the classical computations, and a high-fidelity experimental result indicates that we are likely solving a difficult computational problem.

Ideally, in addition to exponential complexity and high fidelity, a quantum platform should offer valuable applications. In Fig. 4, we illustrate applications of our algorithms to many-body physics where the exponential growth in complexity is a substantial barrier to ongoing research (19–24). By varying the amount of disorder in the system, we are able to study disorder-induced localization. This is done using two-body correlations

$$|\langle \hat{n}_i \hat{n}_j \rangle - \langle \hat{n}_i \rangle \langle \hat{n}_j \rangle| \quad (4)$$

Fig. 2. Complexity: uniform sampling of an exponentially growing state-space.

(A) Histogram of the raw probabilities (see Fig. 1C) for five- to nine-qubit experiments after five cycles of evolution. Before making the histogram, probabilities were weighted by the number of states in the Hilbert space, with all curves placed on a universal axis. The width of the bars represents the size of the bins used to construct the histogram. The data are taken from more than 29.7 million experiments. For dynamics that uniformly explore all states, this histogram decays exponentially; an exponential decay is shown as a solid line for comparison. A histogram of the probabilities for seven qubits after 100 cycles is shown for contrast. In this plot, decoherence dominates and we observe a tall narrow peak around 1. (B) To measure convergence of the measured histograms to an exponential distribution, we compute their distance as a function of the number of cycles. Distance is measured using the Kullback-Leibler divergence (Eq. 2). We find that a maximum overlap occurs after just two cycles, and decoherence subsequently increases the distance between the distributions.

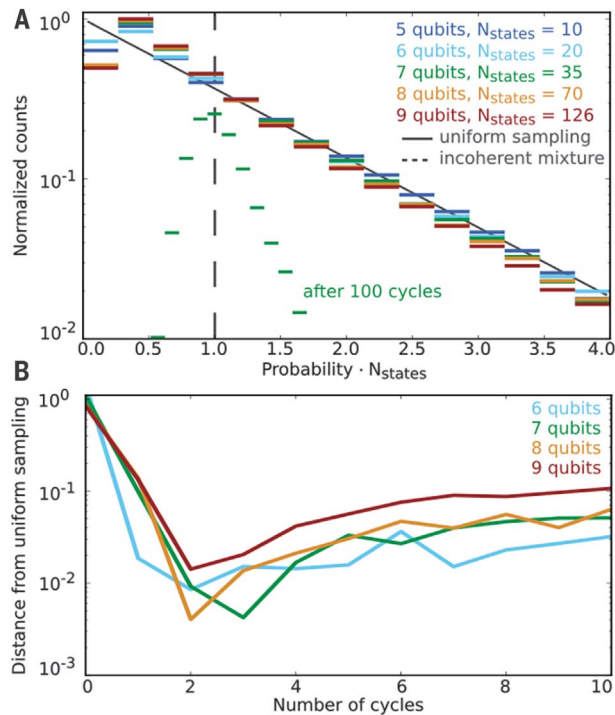


Fig. 3. Fidelity: learning a better control model.

(A) Average fidelity decay versus number of cycles for five- to nine-qubit experiments (circles). The fidelity is computed from Eq. 3. The error per cycle, presented in the inset, is the slope of the dashed line that best fits the data. (B) Using the fidelity as a cost function, we learn optimal parameters for our control model. We take half of the experimental data to train our model. The other half of the data is used to verify this new model; the optimizer does not have access to these data. The corresponding improvement in fidelity of the verification set provides evidence that we are indeed learning a better control model.

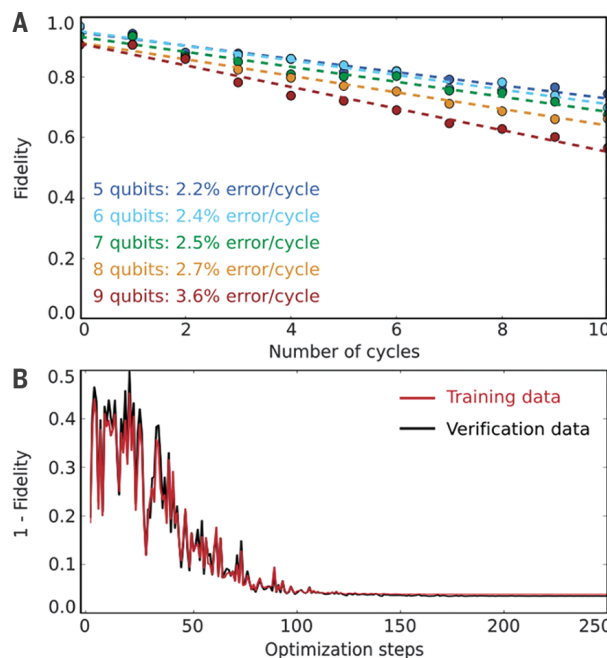
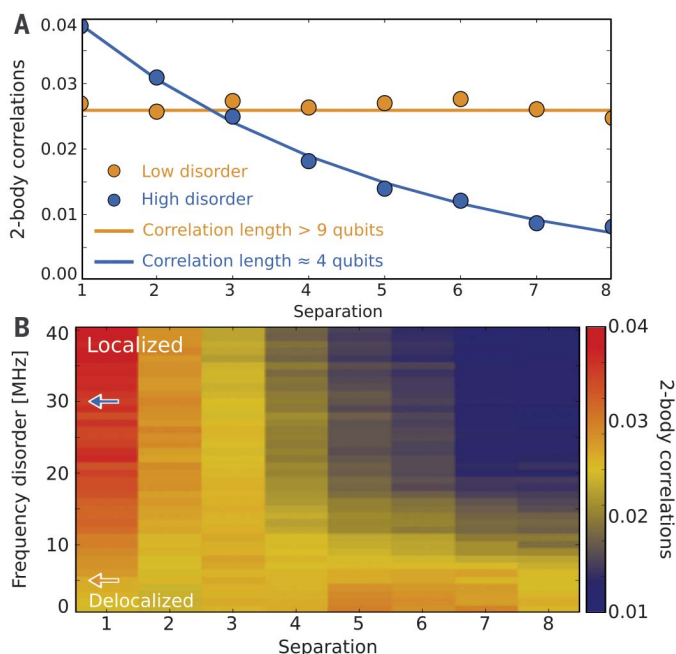


Fig. 4. Applications: localization and delocalization.

(A) Average two-body correlations (Eq. 4) as a function of the separation between qubits. Data are shown for two values of disorder strength. At low disorder, the qubit frequencies are set over a range of ± 5 MHz, and the two-body correlations are independent of separation (i.e., qubits at the ends of the chain are just as correlated as nearest neighbors). At high disorder, the qubit frequencies are set over a range of ± 30 MHz, and we find an exponential decay in correlations as a function of separation. (B) Map of correlations as a continuous function of frequency disorder. Arrows indicate the location of line cuts used in (A). We observe a clear transition from long-range to short-range correlations.



which we average over qubit pairs, cycles (number of coupler pulses), and instances (choice of randomly selected pulse parameters). In Fig. 4A, we plot the average two-body correlations against the separation between qubits. This experiment is performed for both low and high disorder in the qubit frequencies (shown in gold and blue, respectively). Figure 4B depicts the results of our experiment as we continuously vary the amount of disorder.

At low disorder, we find that the correlations are independent of separation: qubits at opposite ends of the chain are as correlated as nearest neighbors. At high disorder, the correlations fall off exponentially with separation. The rate at which this exponential decays allows us to determine the correlation length. A fit to the data is shown

in Fig. 4A as a solid blue line where we find a correlation length of roughly four qubits. The study of localization and delocalization in interacting systems provides a promising application of our algorithms.

REFERENCES AND NOTES

1. R. P. Feynman, *Int. J. Theor. Phys.* **21**, 467–488 (1982).
2. I. Buluta, F. Nori, *Science* **326**, 108–111 (2009).
3. B. P. Lanyon *et al.*, *Nat. Chem.* **2**, 106–111 (2010).
4. A. Peruzzo *et al.*, *Nat. Commun.* **5**, 4213 (2014).
5. A. Mezzacapo *et al.*, *Sci. Rep.* **5**, 13153 (2015).
6. T. Kadowaki, H. Nishimori, *Phys. Rev. E* **58**, 5355–5363 (1998).
7. S. Boixo *et al.*, *Nat. Phys.* **10**, 218–224 (2014).
8. S. Lloyd, M. Mohseni, P. Rebentrost, *arXiv:1307.0411* [quant-ph] (1 July 2013).
9. V. Denchev *et al.*, *Phys. Rev. X* **6**, 031015 (2016).
10. J. Preskill, *arXiv:1203.5813* [quant-ph] (26 March 2012).

11. S. Aaronson, A. Arkhipov, in *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, San Jose, CA, 6 to 8 June 2011 [Association for Computing Machinery (ACM), 2011], pp. 333–342.
12. B. Peropadre, G. G. Guerreschi, J. Huh, A. Aspuru-Guzik, *Phys. Rev. Lett.* **117**, 140505 (2016).
13. M. J. Bremner, A. Montanaro, D. J. Shepherd, *Phys. Rev. Lett.* **117**, 080501 (2016).
14. S. Boixo *et al.*, *arXiv:1608.00263* [quant-ph] (31 July 2016).
15. T. Häner, D. Steiger, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC17*, Denver, CO, 12 to 17 November 2017 (ACM, 2017), pp. 33:1–33:10.
16. See supplementary materials.
17. J. Martinis, M. Geller, *Phys. Rev. A* **90**, 022307 (2014).
18. C. Porter, R. Thomas, *Phys. Rev.* **104**, 483–491 (1956).
19. M. Schreiber *et al.*, *Science* **349**, 842–845 (2015).
20. D. Basko, I. Aleiner, B. Altshuler, *Ann. Phys.* **321**, 1126–1205 (2006).
21. A. Pal, D. Huse, *Phys. Rev. B* **82**, 174411 (2010).
22. L. F. Santos, A. Polkovnikov, M. Rigol, *Phys. Rev. Lett.* **107**, 040601 (2011).
23. M. Rigol, V. Dunjko, M. Olshanii, *Nature* **452**, 854–858 (2008).
24. A. Polkovnikov, K. Sengupta, A. Silva, M. Vengalattore, *Rev. Mod. Phys.* **83**, 863–883 (2011).

ACKNOWLEDGMENTS

We thank E. Kapit and J. Fitzsimons for discussions. **Funding:** This work was supported by Google. C.Q. and Z.C. acknowledge support from the NSF Graduate Research Fellowship under grant DGE-1144085. Devices were made at the UCSB Nanofabrication Facility, a part of the NSF-funded National Nanotechnology Infrastructure Network. K.K. acknowledges support from NASA Academic Mission Services, under contract number NNA16BD14C. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. government. The U.S. government is authorized to reproduce and distribute reprints for governmental purposes, notwithstanding any copyright annotation thereon. **Author contributions:** C.N. designed and fabricated the device. C.N. and P.R. designed the experiment. C.N. performed the experiment and analyzed the data. C.N., K.K., and V.S. developed the physical control model. S.B. and S.V.I. numerically validated the protocol for large qubit arrays. All members of the UCSB and Google teams contributed to the experimental setup and manuscript preparation. **Competing interests:** None declared. **Data and materials availability:** The data that support the plots presented in this paper and other findings of this study are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/195/suppl/DC1
Supplementary Text
Figs. S1 to S27
Tables S1 and S2
References (25–39)
Data S1

19 July 2017; accepted 14 February 2018
10.1126/science.aao4309

NANOMATERIALS

Bottom-up synthesis of multifunctional nanoporous graphene

César Moreno,^{1*} Manuel Vilas-Varela,^{2†} Bernhard Kretz,^{3†} Aran Garcia-Lekue,^{3,4} Marius V. Costache,¹ Markos Paradinas,¹ Mirko Panighel,¹ Gustavo Ceballos,¹ Sergio O. Valenzuela,^{1,5} Diego Peña,^{2*} Aitor Mugarza^{1,5*}

Nanosize pores can turn semimetallic graphene into a semiconductor and, from being impermeable, into the most efficient molecular-sieve membrane. However, scaling the pores down to the nanometer, while fulfilling the tight structural constraints imposed by applications, represents an enormous challenge for present top-down strategies. Here we report a bottom-up method to synthesize nanoporous graphene comprising an ordered array of pores separated by ribbons, which can be tuned down to the 1-nanometer range. The size, density, morphology, and chemical composition of the pores are defined with atomic precision by the design of the molecular precursors. Our electronic characterization further reveals a highly anisotropic electronic structure, where orthogonal one-dimensional electronic bands with an energy gap of ~1 electron volt coexist with confined pore states, making the nanoporous graphene a highly versatile semiconductor for simultaneous sieving and electrical sensing of molecular species.

Nanoporous graphene (NPG) has recently attracted great attention owing to its potential application as an active component of field-effect transistors (FET) (1, 2) and as an atom-thick selective nanosieve for sequencing (3, 4), ion transport (5, 6), gas separation (7–9), and water purification (10, 11). Selectivity in molecular sieving is achieved by reducing the pore size to the scale of single molecules, that is, in the nanometer range, for relevant greenhouse gases, amino acids, or single ions. This has been achieved in several studies at the single-pore level (6) or through the creation of randomly distributed pores (9, 11), where graphene remains semimetallic. Similarly, inducing semiconducting gaps for room-temperature gate actuation requires the generation of sub-10-nm ribbons between pores (1, 12). In this range, atomic-scale disorder and width fluctuations have substantial effect on gap uniformity. Hence, combining semiconducting and sieving functionalities in a single NPG material is a challenging task that requires the simultaneous generation of nanometer-sized pores and ribbons that have to be carved with atomic precision.

Inspired by successful on-surface routes to synthesize covalent carbon-based nanostructures

(13–20), we have devised a strategy that leads to the formation of NPG that exhibits both semiconducting and nanosieving functionalities. Our method relies on the hierarchical control of three thermally activated reaction steps, labeled T1 to T3 in Fig. 1. Nanoribbons and pores with nanometer size, atomic-scale uniformity, and long-range order are formed in separate steps. Graphene nanoribbons (GNRs) are first synthesized by following a previously used route (17, 20), consisting of the surface-assisted Ullmann coupling of aromatic dihalide monomers into polymer chains (T1) and the cyclodehydrogenative aromatization of the intermediate polymeric chains into GNRs (T2). The final step (T3) interconnects GNRs laterally in a reproducible manner by means of a highly selective dehydrogenative cross-coupling (21). This step requires a careful design of the monomer precursor, which defines the edge topology of the resulting GNR that is necessary for a high yield and selectivity of the cross-coupling reaction. The monomer precursor synthesized in this work, labeled DP-DBBA (diphenyl-10,10'-dibromo-9,9'-bianthracene), is a derivative of the DBBA used in the synthesis of seven-carbon atom-wide armchair GNRs (7-AGNR) (17), with phenyl substituents added at (2,2') sites. The latter is the key element for the promotion of the inter-GNR connections that lead to the NPG structure shown in Fig. 1D (see supplementary materials for details of the monomer synthesis) (22). The choice of catalytic surface is also relevant for the selection of the reaction paths that define the intermediates and for the separation of thermal windows that lead to their hierarchical control. Here we use the Au(111) surface, where each reaction step has a different thermal activation onset, as shown below. The NPG can then be transferred to suitable substrates in which its functionalities can be exploited (22).

The structures obtained in each step of the hierarchical synthetic route are characterized using scanning tunneling microscopy (STM). Representative topographic images are shown in Fig. 2. After deposition at room temperature and annealing to temperature $T = 200^\circ\text{C}$, monomers undergo debromination to form the corresponding aryl radicals, which are subsequently coupled by means of C–C bond formation (step T1) (17, 20). The resulting polymeric chains exhibit the characteristic protrusion pairs with a periodicity of 0.84 nm and an apparent height of 0.31 nm, which arises from probing the high ends of the staggered *bis*-anthracene units of the monomer with STM (Fig. 2, A and D) (17, 20). The chains, with lengths of up to 150 nm, predominantly align in close-packed ensembles along the zigzag orientation of the herringbone reconstruction of the Au(111) surface. Both the extraordinary length of the polymeric intermediates and their parallel alignment are crucial ingredients for the high yield and long-range order observed in the final step T3.

Annealing to $T = 400^\circ\text{C}$ triggers the intramolecular cyclodehydrogenation (step T2), giving rise to the aromatization of the chain and the corresponding reduction of the apparent height to 0.18 nm, which is characteristic of GNRs (Fig. 2, B and E) (17, 20). The nanoribbons appear dispersed as individual stripes, yet they maintain a predominantly parallel alignment along the zigzag orientation. As can be seen in the high-resolution image of Fig. 2E, the catafused benzene rings that arise from the cyclization of the phenyl substituent result in a periodic modulation of the width. Consecutive pairs of 7 and 13 carbon atoms define multibay regions made of three conjoined bays (yellow lines in Fig. 1, C and D). This particular edge structure of the nanoribbons, referred to as 7-13-AGNR hereafter, will define both the morphology and size of the corresponding pores in the NPG, as well as its electronic structure.

The aryl-aryl interribbon connection is induced by further annealing to $T = 450^\circ\text{C}$ (step T3). Figure 2C shows how GNRs tend to merge, connecting laterally from each of the fused benzene rings and forming a nanomesh (indicated with a green rectangle). The submolecular structure, observed in the high-resolution image of Fig. 2F, coincides with the NPG structure depicted in Fig. 1D, which reveals that the interribbon coupling occurs by means of a selective C–H bond activation. The activation of specific C–H bonds in polycyclic aromatic hydrocarbons is nontrivial because of the presence of multiple quasi-energetic bonds (three in the case of the 7-13-AGNR, labeled as H^1 to H^3 in Fig. 1). In step T3, the selectivity in the C–C bond formation between adjacent GNRs is driven by the easy accessibility to the radical formed after the C– H^3 bond cleavage, as opposed to the steric hindrance associated with the radicals formed after the C– H^1 or C– H^2 bond cleavage. Another remarkable milestone is the long-range order achieved. To date, the observation of selective intermolecular aryl-aryl coupling has been limited to small supramolecular structures (19, 23–28). The hierarchical

¹Catalan Institute of Nanoscience and Nanotechnology (ICN2), Consejo Superior de Investigaciones Científicas (CSIC) and The Barcelona Institute of Science and Technology, Campus UAB, Bellaterra, 08193 Barcelona, Spain. ²Centro Singular de Investigación en Química Biológica e Materiais Moleculares (CIQUS) and Departamento de Química Orgánica, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain. ³Donostia International Physics Center (DIPC), 20018 San Sebastian, Spain. ⁴Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Spain. ⁵ICREA–Institut de Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain.

*Corresponding author. Email: cesar.moreno@icn2.cat (C.M.); diego.pena@usc.es (D.P.); aitor.mugarza@icn2.cat (A.M.)

†These authors contributed equally to this work.

strategy of our method allows us to set the long-range order in step T2, where the length of pre-aligned GNRs represents the size limitation of the NPG. The high yield and remarkable selectivity of this coupling mechanism can be best appreciated when the surface is saturated with polymeric chains by depositing the precursor with the substrate at $T = 200^\circ\text{C}$ (Fig. 2, G and H). As a result, a coupling yield close to 100% is achieved, where every GNR is integrated in a NPG domain. Following this procedure, NPG sheets as large as 50 nm by 70 nm are easily obtained, with atomically reproducible pores of 0.4 nm by 0.9 nm, ultra-high densities of 480×10^3 pores per μm^2 , and a characteristic defect concentration of $\sim 2\%$.

The peculiar topology of the NPG imprints a band gap, one-dimensional (1D) anisotropy and different types of localization in the electronic

states, with potential implications in transport and sensing. These can be rationalized within the same hierarchical approach used in the synthesis, namely by considering the states of individual 7-13-AGNRs as building blocks that already contain the main features and following their evolution as the ribbons interconnect. By combining density functional theory (DFT) (Fig. 3, A and B) and scanning tunneling spectroscopy (STS) (Fig. 3, C to E), three types of bands are identified. Examples of each band and its corresponding wave function are highlighted in different colors in Fig. 3, A and B, respectively: longitudinal bands (L, yellow), transversal bands (T, purple), and bay and pore bands (P, green).

The L and T bands originate from the carbon s and p orbitals. The L bands are similar to the conventional bands in straight AGNRs and dis-

perse along the ribbon (along ΓZ). They appear confined within the seven-carbon atom-wide backbone of the GNR. On the contrary, the T bands are localized within the 13-carbon atom-wide periodic stripes, and thus they do not disperse in the longitudinal direction. They arise from the superlattice periodicity imprinted by the modulated width of the 7-13-AGNR and are therefore exclusively related to its edge topology. The semiconducting gap of the 7-13-AGNR is determined by the L bands. For the free-standing ribbon, DFT predicts a band gap of 0.74 eV, which is increased to 1.36 eV after including self-energy corrections within the GW approximation (22). Experimental STS spectra reveal a band gap of $\Delta_{\text{gap}} = 1.0$ eV (Fig. 3C), slightly lower than the GW band gap, as expected from the screening effect of the underlying substrate

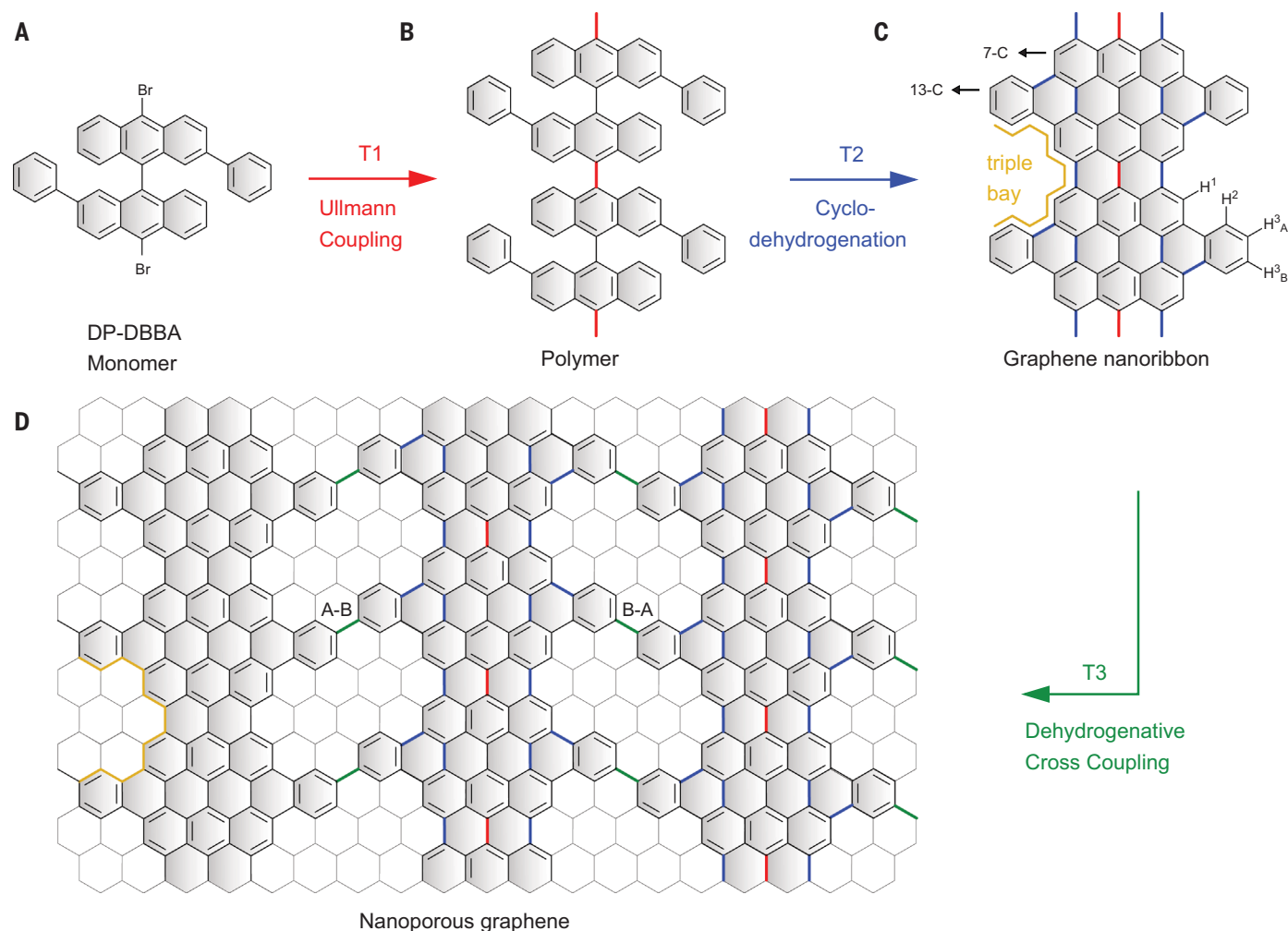


Fig. 1. Schematic illustration of the synthetic hierarchical path for the generation of NPG. (A) The DP-DBBA monomer used as precursor. (B) At step T1, DP-DBBA is debrominated, and the radical carbon atoms cross-couple to form polymer chains. (C) At step T2, an intramolecular cyclodehydrogenation leads to the planar graphene nanoribbon. The cyclization of the phenyl substituent modulates the width of the GNR with pairs of 7- and 13-carbon atom-wide sections, forming multibay regions that consist of three conjoined bay

regions (yellow lines) and leaving three types of C-H bonds at the edge (H^1 to H^3). Each type will have two equivalent positions, as represented for H^3 with A and B labels. (D) Finally, at step T3, the GNRs are interconnected by the H^3 bonds via dehydrogenative cross-coupling, giving rise to the NPG structure (the extended graphene structure is underlaid to highlight the structure of the nanopores). The A-B or B-A bonding combinations give rise to identical pores with different orientations.

(29, 30). Notably, this value is smaller than the 1.5 eV measured for the wider 13-AGNR (29), highlighting the role of edge topology in determining band gaps.

The effect of the interribbon connection is specific to the band type (Fig. 3A). Protected within the backbone, L bands remain unperturbed in the NPG, as indicated by the lack of dispersion in the transversal direction (along ΓX). The DFT band gap is only reduced by 0.12 eV when compared with GNRs, which agrees with a downshift of similar size measured by STS for the conduction band onset (22). By contrast, the extension of the T-band wave functions across the 13-carbon atom-wide section enables substantial interribbon coupling and the formation of 1D dispersing states with a similar mobility as the longitudinal ones. The resulting wave functions consist of noninteracting zigzag stripes that run across the GNRs. In the shown calculations, the structure consists of alternating pore orientations that correspond to the two equivalent C-H³ bonds

(labeled as C-H³_A and C-H³_B in Fig. 1C), but the same conclusions are obtained using NPGs formed exclusively by either of the two pore configurations (22).

The origin of the P bands is more exotic. Localized within the vacuum pocket defined by the multibay region, they are not related to atomic orbitals or their hybridization, as has been observed in other molecular pores (31). Instead, they originate from the free electron-like image potential states that are confined at the vacuum side along the GNR edge. They can be regarded as the 2D analog of the superatom states that develop when a graphene sheet bends into a fullerene (32). In our lower-dimension analog, the straight graphene edge “bends” into a periodic array of weakly coupled multibays that give rise to rather flat superatom bands (22). In the NPG, the P states of adjacent multibays interact when they pair to form pores, leading to bonding and antibonding bands, as observed in the DFT band structure (22). Experimentally, the

high-energy-lying antibonding band cannot be probed without affecting the integrity of the NPG. STS spectra, however, do reveal an energy reduction of $\Delta_{\text{bond}} = 0.30$ eV, which corresponds to the formation of the bonding band (Fig. 3D) and is in very close agreement with the shift of 0.28 eV obtained by DFT.

The band heterogeneity described above is reflected in the energy-dependent conductance across the two perpendicular directions (Fig. 4A). Three regions can be identified in the calculations, depending on the energy E : a region of true energy gap around the Fermi energy E_F , in which the conductance is fully suppressed; a region for $|E - E_F| < 1.2$ eV, in which transport is purely longitudinal; and a region for $|E - E_F| > 1.2$ eV, in which transport has both longitudinal and transversal components. To experimentally demonstrate the semiconducting properties of the NPG, the transport response was characterized by using FET structures. The NPG was first transferred onto a Si/SiO₂ substrate and then

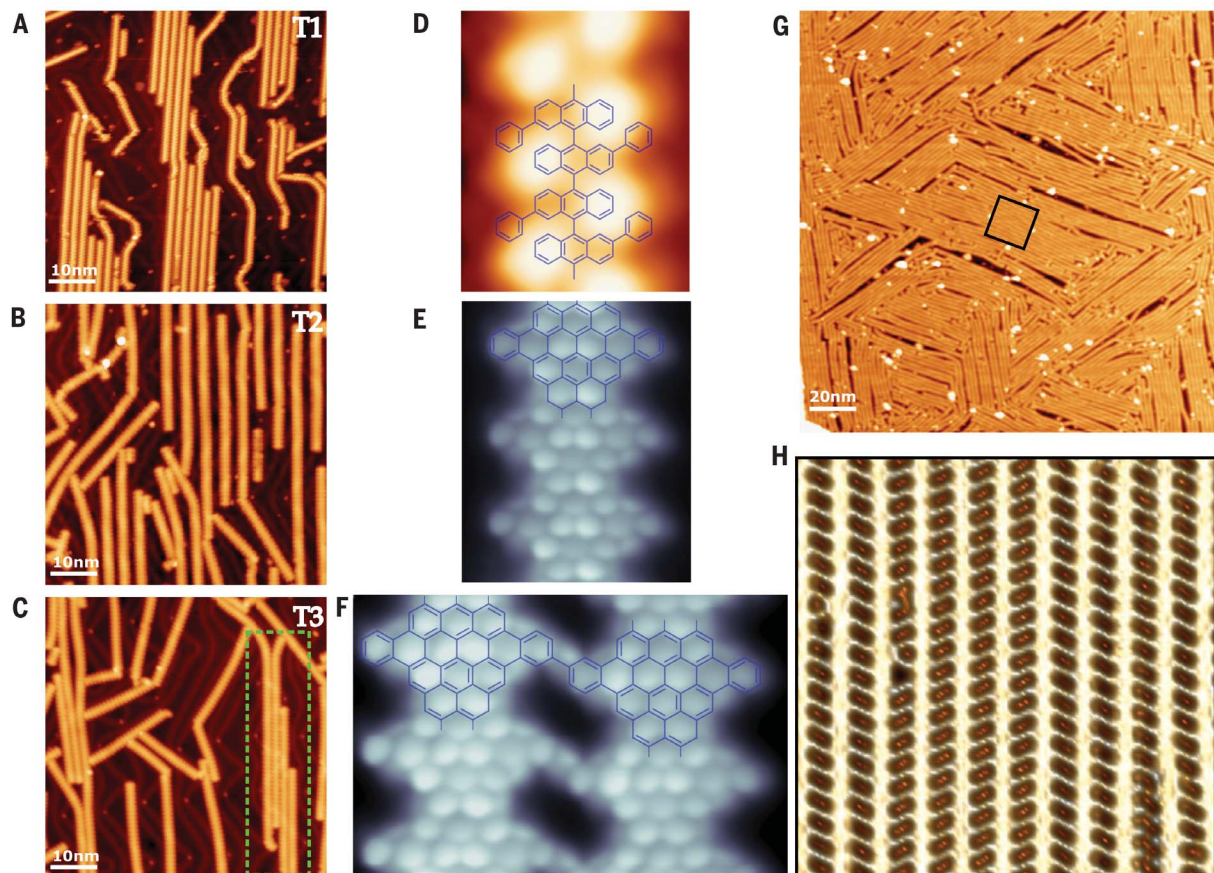


Fig. 2. Hierarchical synthesis of NPG. (A to C) Constant-current STM images showing the distribution and morphology of the different covalent structures obtained for a low coverage of precursors after each thermal annealing step T1, T2, and T3. At this coverage, NPG stripes form locally [indicated with a green rectangle in (C)]. (D to F) Magnified images revealing the internal structure for each case, (A) to (C). The high-resolution images in (E) and (F) are obtained by using a CO-functionalized tip in constant-height mode (22). The atomic models

depicted in Fig. 1 for steps T1 to T3 are overlaid in (D) to (F), respectively. (G) Constant-current STM image of a surface totally covered with NPG domains with sizes up to 50 nm by 70 nm, obtained by a saturated deposition of the precursor at $T = 200^\circ\text{C}$. (H) Laplacian-filtered topographic close-up image of the NPG region marked by the black rectangle in (G), showing a regular array of identical pores with low defect density. All imaging parameters are provided in the supplementary materials (22).

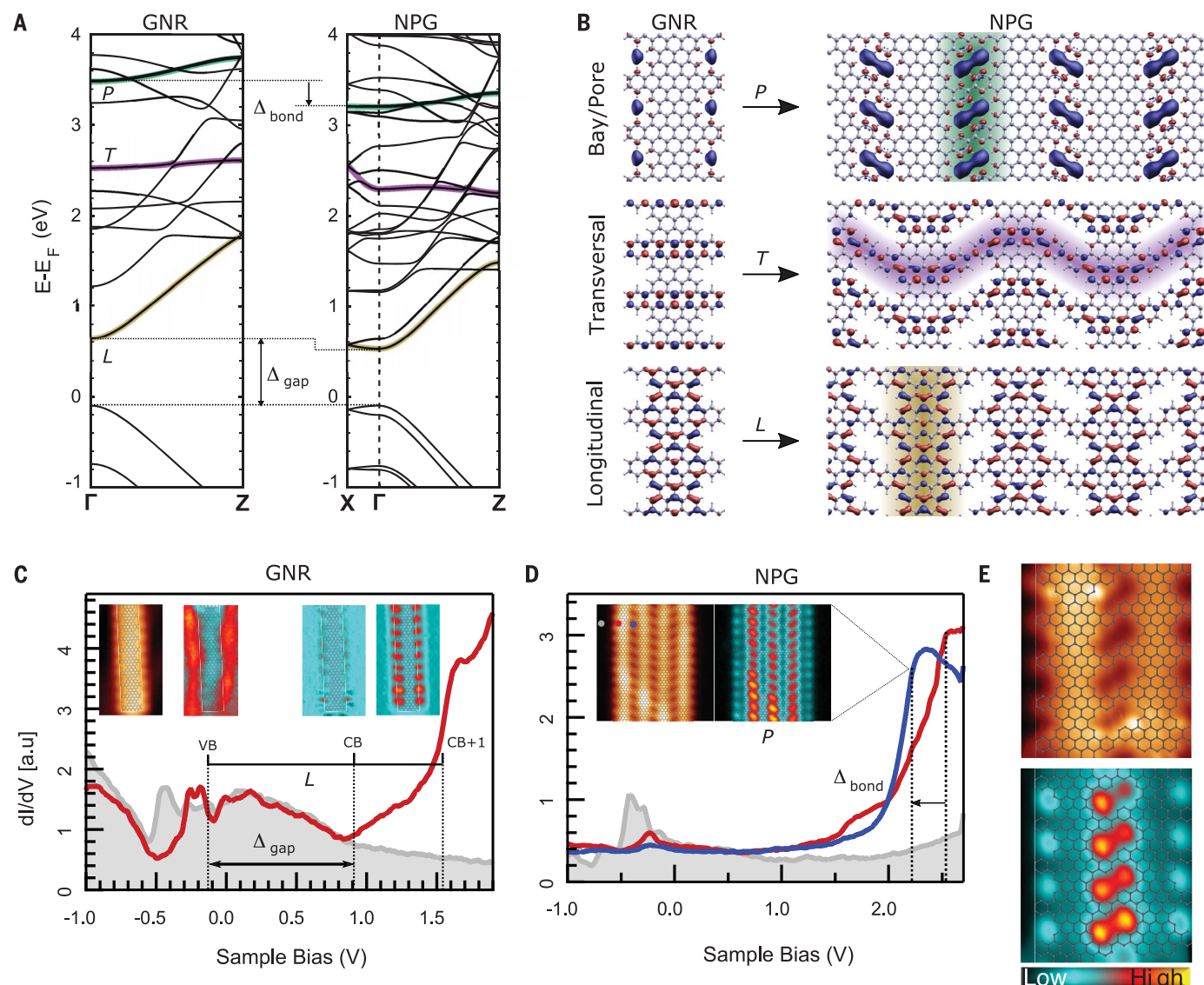


Fig. 3. Electronic properties of the 7-13-AGNR and the NPG. (A) Band structure calculated by DFT for individual 7-13-AGNRs (left) and the NPG (right). Examples of L, T, and P bands are highlighted with yellow, purple, and green guiding lines, respectively. The Fermi level is determined by using the experimental valence band (VB) energy as a reference (22). (B) Wave functions at Γ for each of the band examples in (A). Their dispersion directions are highlighted by guiding stripes. (C) dI/dV spectra acquired at the multibay edge of a 7-13-AGNR, where the onset of the CB, VB, and CB+1 bands can be identified. CB, conduction band; a.u., arbitrary units. (D) dI/dV spectra acquired at the peripheral

multibay (red) and a pore (blue) region of the NPG. The interaction between the two bay states within a pore results in an energy shift of Δ_{bond} owing to the formation of a bonding band. In (C) and (D), reference spectra acquired on Au(111) are added in shaded gray, and the insets show constant-height tunneling current (I_t) images (left) and dI/dV maps (right) acquired at the related energies. (E) High-resolution constant-height I_t image (top) and dI/dV map (bottom) of the pore states, acquired at +2.2 eV, where the double-lobed structure predicted by DFT is reproduced. The localization of this state within the pore is shown by overlaying an atomic model of the local NPG structure.

contacted with Pd electrodes by using electron-beam lithography and shadow evaporation (22). The Si substrate is highly doped to fulfill the role of a back-gate electrode across the 90-nm-thick SiO_2 gate dielectric. Remarkably, the large dimension of the NPG sheets enables a large device yield of ~75% for the designed channel length of 30 nm. Figure 4B shows typical room-temperature drain-source current-voltage gate ($I_{\text{ds}}-V_{\text{g}}$) behavior at fixed drain-source voltage bias (V_{ds}). The devices show good performance, presenting hole transport and an on-off ratio

of $\sim 10^4$, which is comparable to prior work on single GNRs (33). The transistor characteristics are highly nonlinear at low bias, indicating that the transport is limited by the presence of a Schottky barrier at the Pd-NPG interface (21, 33) and suggesting that larger on-off ratios could be obtained by lowering the contact work function (33).

The semiconducting functionality of the presented NPG architecture can be exploited in a new generation of graphene-based devices such as FET-sensors or gate-controlled sieves. At a

high doping level, the onset of the transversal bands provides an orthogonal, noninteracting 1D channel. The set of L and T bands brings to graphene the in-plane anisotropy that makes 2D materials, such as phosphorene or black phosphorous, appealing for FET, optical, and sensing applications (34, 35). Finally, the presence of confined states within the nanopores makes NPG very attractive for detection and electronic tracking in chemical and biological sensors and filters. Confined states could be shifted down to the Fermi level by the interaction with ions and

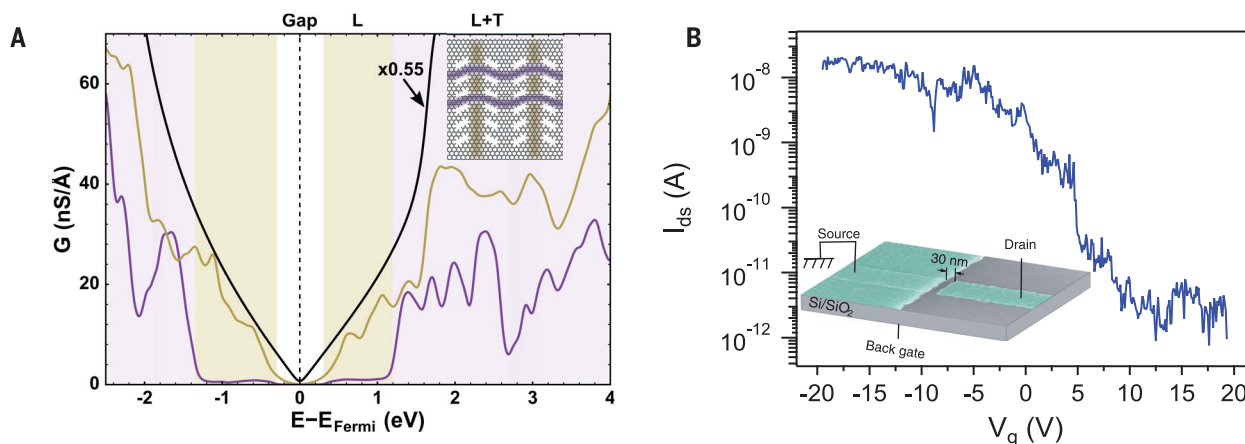


Fig. 4. Transport properties of the NPG. (A) Conductance (G) calculated in the longitudinal (yellow) and transversal (purple) directions of the NPG, as defined in Fig. 3, with that of pristine graphene added as a comparison (black line); the latter has been multiplied by 0.55. Colored

backgrounds delimit regions of total gap (white), only longitudinal (yellow), and combined longitudinal and transversal (purple) transport. (B) I_{ds} - V_g characteristics of a NPG device of 30-nm-channel length gated by a 90-nm-thick SiO_2 gate oxide.

molecules (36, 37), making them detectable in transport measurements.

REFERENCES AND NOTES

- J. Bai, X. Zhong, S. Jiang, Y. Huang, X. Duan, *Nat. Nanotechnol.* **5**, 190–194 (2010).
- X. Liang et al., *Nano Lett.* **10**, 2454–2460 (2010).
- B. M. Venkatesan, R. Bashir, *Nat. Nanotechnol.* **6**, 615–624 (2011).
- S. Garaj et al., *Nature* **467**, 190–193 (2010).
- K. Sint, B. Wang, P. Král, *J. Am. Chem. Soc.* **130**, 16448–16449 (2008).
- R. C. Rollings, A. T. Kuan, J. A. Golovchenko, *Nat. Commun.* **7**, 11408 (2016).
- D. E. Jiang, V. R. Cooper, S. Dai, *Nano Lett.* **9**, 4019–4024 (2009).
- S. Blankenburg et al., *Small* **6**, 2266–2271 (2010).
- S. P. Koenig, L. Wang, J. Pellegrino, J. S. Bunch, *Nat. Nanotechnol.* **7**, 728–732 (2012).
- D. Cohen-Tanugi, J. C. Grossman, *Nano Lett.* **12**, 3602–3608 (2012).
- S. P. Surwade et al., *Nat. Nanotechnol.* **10**, 459–464 (2015).
- M. Y. Han, B. Özyilmaz, Y. Zhang, P. Kim, *Phys. Rev. Lett.* **98**, 206805 (2007).
- G. Franc, A. Gourdon, *Phys. Chem. Chem. Phys.* **13**, 14283–14292 (2011).
- P. A. Held, H. Fuchs, A. Studer, *Chemistry* **23**, 5874–5892 (2017).
- M. Bieri et al., *Chem. Commun. (Camb.)* **2009** (45), 6919–6921 (2009).
- L. Lafferentz et al., *Nat. Chem.* **4**, 215–220 (2012).
- J. Cai et al., *Nature* **466**, 470–473 (2010).
- A. Basagni et al., *J. Am. Chem. Soc.* **137**, 1802–1808 (2015).
- A. Floris et al., *J. Am. Chem. Soc.* **138**, 5837–5847 (2016).
- Y. C. Chen et al., *ACS Nano* **7**, 6123–6128 (2013).
- The presence of two equivalent C-H³ bonds, labeled as A and B in Fig. 1C, leads to either A-B or B-A bonding configurations, resulting in upward- and downward-oriented pores of identical morphology.
- Materials and methods are available as supplementary materials.
- Q. Sun, C. Zhang, H. Kong, Q. Tan, W. Xu, *Chem. Commun. (Camb.)* **50**, 11825–11828 (2014).
- C. Sánchez-Sánchez et al., *ACS Nano* **10**, 8006–8011 (2016).
- Q. Li et al., *J. Am. Chem. Soc.* **138**, 2809–2814 (2016).
- N. Kocić et al., *J. Am. Chem. Soc.* **138**, 5585–5593 (2016).
- H. Huang et al., *Sci. Rep.* **2**, 983 (2012).
- J. D. Teeter et al., *Nanoscale* **9**, 18835–18844 (2017).
- Y.-C. Chen et al., *Nat. Nanotechnol.* **10**, 156–160 (2015).
- P. Ruffieux et al., *ACS Nano* **6**, 6930–6935 (2012).
- Y. Q. Zhang, J. Björk, J. V. Barth, F. Klappenberger, *Nano Lett.* **16**, 4274–4281 (2016).
- M. Feng, J. Zhao, H. Petek, *Science* **320**, 359–362 (2008).
- J. P. Llinas et al., *Nat. Commun.* **8**, 633 (2017).
- F. Xia, H. Wang, Y. Jia, *Nat. Commun.* **5**, 4458 (2014).
- L. Kou, T. Frauenheim, C. Chen, *J. Phys. Chem. Lett.* **5**, 2675–2681 (2014).
- M. Feng, J. Zhao, T. Huang, X. Zhu, H. Petek, *Acc. Chem. Res.* **44**, 360–368 (2011).
- G. Csányi, P. B. Littlewood, A. H. Nevidomskyy, C. J. Pickard, B. D. Simons, *Nat. Phys.* **1**, 42–45 (2005).

ACKNOWLEDGMENTS

We acknowledge thoughtful discussions with E. Guitián and D. Pérez for the synthesis of the monomer precursors. We thank P. Brandimarte and D. Sanchez-Portal for useful discussions related to the calculations. We thank E. Del Corro, C. J. J. Hebert, and J. A. Garrido for their support with the Raman measurements. **Funding:** This research was funded by the Centres de Recerca de Catalunya Programme-Generalitat de Catalunya, the Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2016–2019, ED431G/O9), and the European Regional Development Fund. This research was supported by the Spanish Ministry of Economy and Competitiveness (under contract no. MAT2016-

78293-C6-2-R, MAT2016-78293-C6-3-R, MAT2016-78293-C6-4-R, and MAT2016-75952-R and Severo Ochoa no. SEV-2013-0295); the Secretariat for Universities and Research, Knowledge Department of the Generalitat de Catalunya 2014 SGR 715, 2014 SGR 56, and 2017 SGR 827; the Basque Department of Education (contract no. PI-2016-1-0027); and the European Union's Horizon 2020 research and innovation program under grant agreement 696656. C.M. was supported by the Agency for Management of University and Research grants (AGAUR) of the Catalan government through the FP7 framework program of the European Commission under Marie Curie COFUND action 600385.

Author contributions: D.P. and M.V.-V. synthesized the monomer precursor. B.K. and A.G.-L. performed the ab initio calculations. C.M. and M.Pan. performed the STM measurements. C.M., G.C., and A.M. analyzed the STM data. M.Par. and C.M. carried out the sample transfer and Raman characterization. M.V.C. and S.O.V. fabricated the devices and carried out the transport measurements. All authors discussed the results and participated in writing the manuscript. D.P. and A.M. initiated and directed this research. **Competing interests:** C.M., M.V.-V., G.C., D.P., and A.M. are inventors on patent application EP18382088.5 submitted by the Catalan Institute of Nanoscience and Nanotechnology, which covers the reported NPG material.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/199/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (38–57)

19 October 2017; accepted 7 March 2018
10.1126/science.aar2009

NOROVIRUS

Tropism for tuft cells determines immune promotion of norovirus pathogenesis

Craig B. Wilen,¹ Sanghyun Lee,¹ Leon L. Hsieh,¹ Robert C. Orchard,¹ Chandni Desai,¹ Barry L. Hykes Jr.,¹ Michael R. McAllaster,¹ Dale R. Balce,¹ Taylor Feehley,¹ Jonathan R. Brestoff,¹ Christina A. Hickey,¹ Christine C. Yokoyama,¹ Ya-Ting Wang,¹ Donna A. MacDuff,² Darren Kreamalmayer,¹ Michael R. Howitt,³ Jessica A. Neil,⁴ Ken Cadwell,⁴ Paul M. Allen,¹ Scott A. Handley,¹ Menno van Lookeren Campagne,⁵ Megan T. Baldrige,⁶ Herbert W. Virgin^{1*}

Complex interactions between host immunity and the microbiome regulate norovirus infection. However, the mechanism of host immune promotion of enteric virus infection remains obscure. The cellular tropism of noroviruses is also unknown. Recently, we identified CD300lf as a murine norovirus (MNoV) receptor. In this study, we have shown that tuft cells, a rare type of intestinal epithelial cell, express CD300lf and are the target cell for MNoV in the mouse intestine. We found that type 2 cytokines, which induce tuft cell proliferation, promote MNoV infection in vivo. These cytokines can replace the effect of commensal microbiota in promoting virus infection. Our work thus provides insight into how the immune system and microbes can coordinately promote enteric viral infection.

Human noroviruses (HNoVs) are the leading cause of acute viral gastroenteritis worldwide, causing up to 700 million infections and 200,000 deaths annually (1). Despite this disease burden, it is unknown what cell type(s) mediate transmission and, in some individuals, chronic infection (2). Murine norovirus (MNoV) represents a model for HNoV pathogenesis and immunity. More broadly, MNoV serves as a tractable system to uncover previously unidentified virus-host interactions such as the capacity of MNoV infection to trigger human-relevant pathology in genetically susceptible animals and the role of intestinal bacteria in promoting enteric viral infection (3–8). Identifying the cell tropism of MNoV could provide mechanistic insight into such phenomena and thereby shed light on enteric immunity and the genotype-phenotype relationship.

Norovirus tropism is not fully understood in either immunocompetent mice or humans. Recently, we showed that a small population of epithelial cells is the reservoir for chronic MNoV infection and that this epithelial cell tropism is determined in part by the MNoV nonstructural protein NS1 (9). However, the reason for selective

intestinal epithelial cell infection, how infected cells differ from adjacent cells in the intestinal epithelium, and why we seldom observed adjacent infected epithelial cells are unknown. We recently identified CD300lf as a protein receptor for MNoV (10, 11). CD300lf is both necessary and sufficient for infection in vitro, and *Cd300lf*^{−/−} animals are resistant to fecal-oral transmission of persistent MNoV infection (10). In this study, we used this finding to identify the target cell of MNoV in vivo.

Because MNoV readily replicates in explanted macrophages and dendritic cells that express CD300lf (12), we first sought to determine whether bone marrow–derived myeloid cells were responsible for infection by performing bone marrow transplants between *Cd300lf*^{−/−} and wild-type (WT) littermates, followed by oral infection with MNoV strain CR6 (MNoV^{CR6}). MNoV^{CR6} infection is characterized by robust fecal-oral transmission, persistent enteric infection resistant to adaptive immune clearance, and prolonged fecal shedding (13). Such persistent MNoV strains replicate predominantly in the distal small intestine and colon and can be detected in mesenteric lymph nodes (MLNs), but evidence of infection in the spleen is scant (14).

WT mice receiving WT bone marrow remained susceptible to MNoV^{CR6}, but *Cd300lf*^{−/−} mice receiving *Cd300lf*^{−/−} bone marrow were resistant to MNoV infection, as measured by fecal shedding of MNoV (Fig. 1A) and tissue levels of viral genomes 21 days after infection (Fig. 1, B to E). Surprisingly, WT mice that received *Cd300lf*^{−/−} bone marrow were susceptible to MNoV^{CR6}, and *Cd300lf*^{−/−} animals receiving WT bone marrow were resistant to infection. Viral titers in the ileum and colon correlated with those in feces (Fig. 1, A to C). WT mice that received either WT

or *Cd300lf*^{−/−} bone marrow transplants remained susceptible to MNoV^{CR6}. Splenic infection was minimal in all groups examined, consistent with findings from prior studies with nonirradiated WT animals (Fig. 1D) (13). MNoV^{CR6} genomes were undetectable in the MLNs of *Cd300lf*^{−/−} mice receiving either WT or *Cd300lf*^{−/−} bone marrow, but viral genomes were detected at similar levels in the MLNs of WT mice receiving either WT or *Cd300lf*^{−/−} bone marrow transplants (Fig. 1E). Thus, a recipient *Cd300lf* genotype was the primary determinant of MNoV^{CR6} intestinal replication and shedding, indicating that radiation-resistant cells were responsible for MNoV^{CR6} enteric infection. In contrast, both radiation-sensitive and -resistant cells contributed to infection with MNoV strain CW3 (MNoV^{CW3}), which causes acute systemic infection (14–16). The inability of MNoV^{CW3} to infect epithelial cells, to be shed in the feces, and to establish chronic infection maps to the viral NS1 protein, which is required to counteract interferon-λ (IFN-λ) signaling (5, 9, 13). We focused further efforts on identifying the tropism responsible for MNoV^{CR6} enteric infection and shedding.

Consistent with our bone marrow transplant data, we recently determined that rare isolated intestinal epithelial cells were infected by MNoV^{CR6} during chronic infection, though the identity of the cells was not defined (9). Together with the bone marrow transplantation experiments described above, these findings indicate that a radiation-resistant epithelial cell must express the MNoV receptor (9). However, CD300lf is an immunoregulatory protein thought to be expressed on hematopoietic cells, particularly myeloid cells (17, 18). Expression of CD300lf on epithelial cells has not been described previously. We therefore performed immunofluorescence microscopy analysis of uninfected WT mice and observed a rare population of CD300lf-expressing cells throughout the ilea and colons (Fig. 2, A and B). Given the amphora-like morphology and the scarcity of CD300lf-expressing epithelial cells, we hypothesized that they were tuft cells, a rare chemosensory epithelial cell type in the hollow organs of mammals, including mice and humans (19). These cells, also known as brush, caveolated, multivesicular, or fibrillovesicular cells, contain a long apical “tuft” of microvilli, which protrudes into the intestinal lumen, and were recently discovered to be the primary source of interleukin-25 (IL-25), a cytokine that initiates a type 2 immune response upon intestinal helminth or parasite infection (20–22). Indeed, all observed CD300lf⁺ epithelial cells expressed the tuft cell markers doublecortin-like kinase 1 (DCLK1) and cytokeratin 18 (CK18) (Fig. 2, A and B) (23). We also confirmed tuft cell-specific expression of *Cd300lf* transcripts in previously reported single-cell RNA sequencing (RNA-seq) data from mouse intestinal enteroids (24, 25). Next, we assessed CD300lf expression on intestinal epithelial cells (EpCAM⁺ CD45[−]) in a mouse line expressing a tuft cell-specific fluorescent reporter (Gfi1b-GFP) (26). There was near perfect concordance between Gfi1b-GFP expression and CD300lf expression

¹Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA.

²Department of Microbiology and Immunology, University of Illinois at Chicago College of Medicine, Chicago, IL, USA.

³Department of Immunology and Infectious Disease, Harvard T. H. Chan School of Public Health, Boston, MA, USA.

⁴Kimmel Center for Biology and Medicine at the Skirball Institute and Department of Microbiology, New York University School of Medicine, New York, NY 10016, USA.

⁵Department of Immunology, Genentech, 1 DNA Way, South San Francisco, CA 94080, USA. ⁶Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA.

*Corresponding author. Email: virgin@wustl.edu

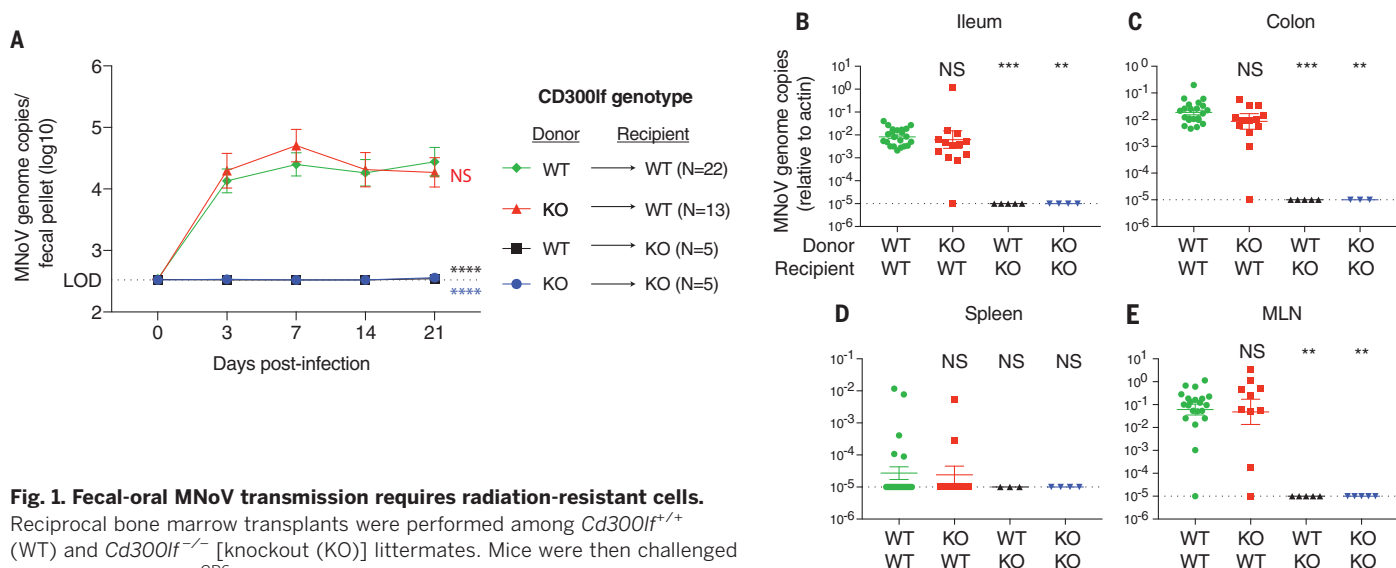


Fig. 1. Fecal-oral MNoV transmission requires radiation-resistant cells.

Reciprocal bone marrow transplants were performed among *Cd300lf^{+/+}* (WT) and *Cd300lf^{-/-}* [knockout (KO)] littermates. Mice were then challenged perorally with MNoV^{CR6}, which establishes persistent enteric infection in WT animals. **(A)** WT mice remained susceptible to MNoV, as measured by numbers of viral genome copies in feces at the indicated time points. In contrast, KO mice did not shed MNoV^{CR6} whether they received WT or KO bone marrow. **(B to E)** Twenty-one days postchallenge, MNoV viral genome loads were measured in the ileum (B), colon (C), spleen (D), and mesenteric lymph nodes (MLNs) (E). WT recipients had significantly more viral genomes than KO recipients. There was no significant difference between WT recipients of either WT or KO bone marrow. Fecal samples were analyzed by repeated-measures

analysis of variance (ANOVA). Tissue samples were analyzed by one-way ANOVA. Significant differences for both fecal and tissue samples were relative to data for the WT donor–WT recipient (WT→WT) control as indicated. Means \pm SEM are shown. NS, not significant; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$. LOD, limit of detection (represented by the dotted line in each graph). Data are pooled from three independent experiments. The numbers of mice per group are indicated in (A).

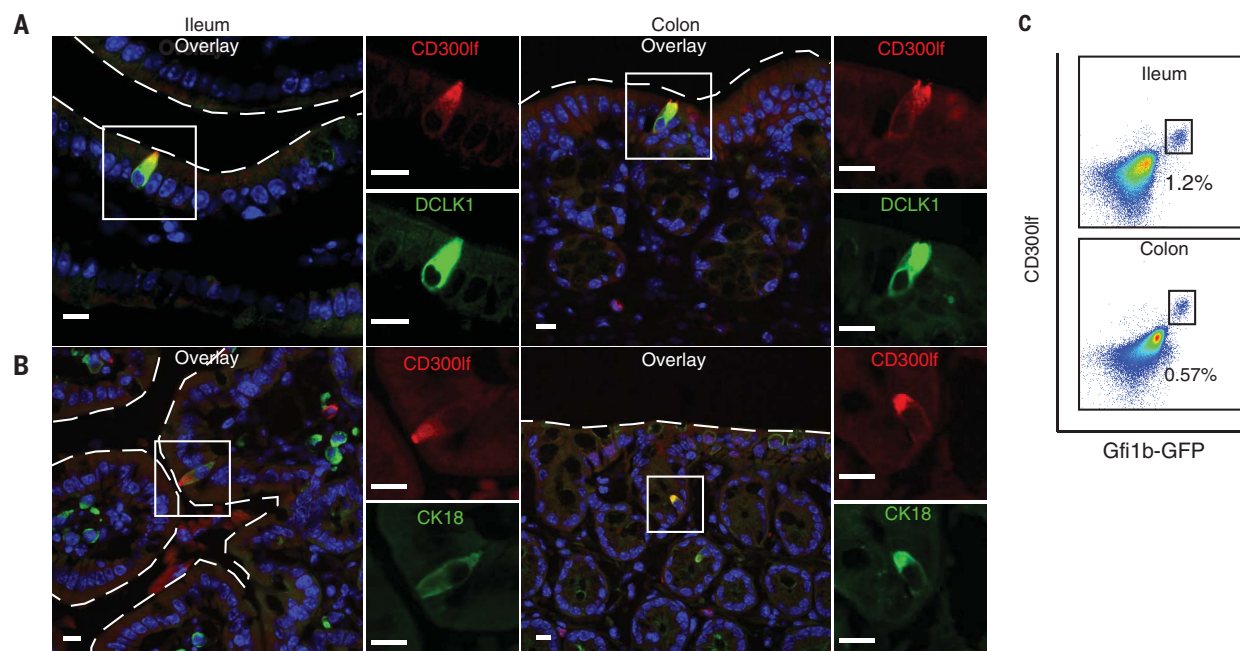
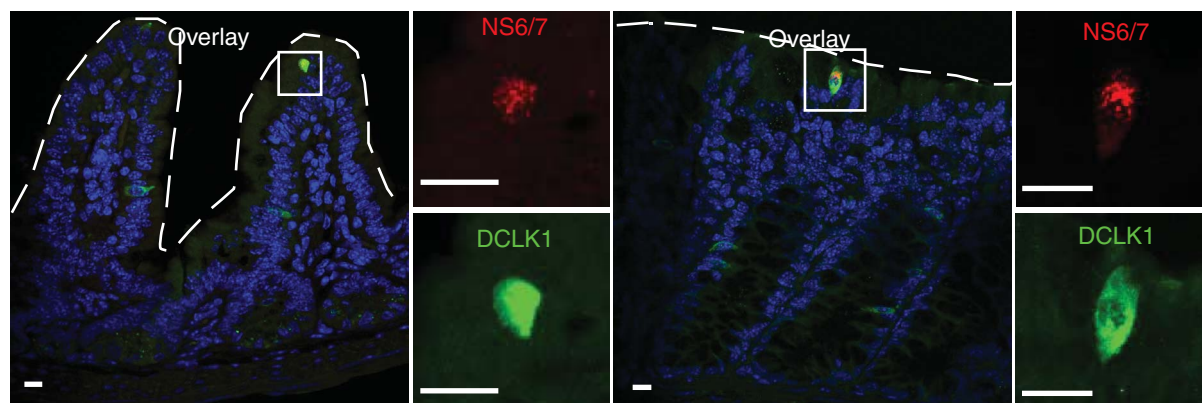


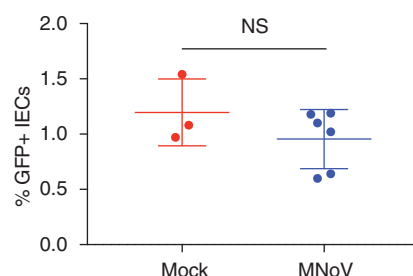
Fig. 2. CD300If is expressed on tuft cells but not on other intestinal epithelial cells. (A and B) The MNoV receptor CD300If is detectable on rare intestinal epithelial cells with morphology consistent with tuft cells. CD300If colocalizes with tuft cell markers (A) DCLK1 and (B) CK18 in mouse ileum and colon. CD300If is apically polarized toward the intestinal lumen. (C) CD300If is expressed on Gfi1b-GFP⁺ tuft cells but

not on other intestinal epithelial cells, as measured by flow cytometry. Events shown are Singlets⁺Live⁺CD45⁺EpCAM⁺. Images and fluorescence-activated cell sorting plots are representative of one of at least three independent experiments. Dashed lines represent the epithelial barrier. White boxes in the overlay images correspond to the magnified inset images. Scale bars, 10 μ m.

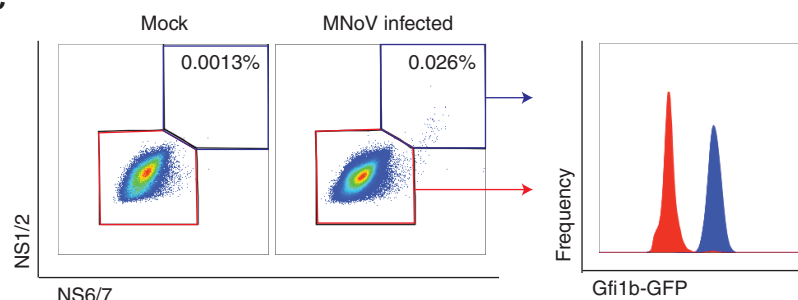
A



B



C



D

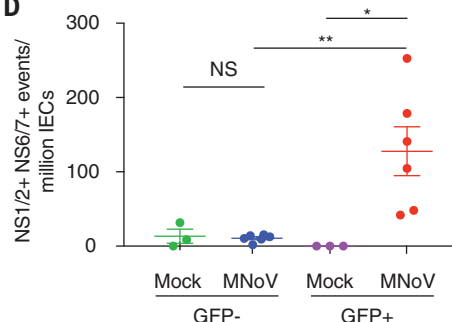


Fig. 3. MNoV^{CR6} specifically infects CD300lf-expressing intestinal tuft cells.

(A) MNoV nonstructural protein NS6/7 colocalizes with DCLK1 in the ilea and colons of WT mice infected with MNoV^{CR6} at 7 days postinfection. NS6/7 expression is punctate and cytoplasmic, consistent with the viral replication complex. (B) Flow cytometry analysis of intestinal epithelial cells (IECs) (Singlet⁺Live⁺CD45⁺EpCAM⁺) from Gfi1b-GFP⁺ tuft cell reporter mice revealed similar frequencies of tuft cells in infected and uninfected mice. (C) A rare population of cells that coexpress the MNoV nonstructural proteins NS1/2 and NS6/7 was observed. These MNoV-positive cells are Gfi1b-GFP⁺ cells, demonstrating that they are tuft cells. (D) NS1/2⁺ NS6/7⁺ events were significantly enriched among GFP⁺ cells. NS1/2⁺ NS6/7⁺ events were at background levels among non-tuft cells. Data are pooled from three independent experiments with one to two mice per group. Shown are means \pm SEM. NS, not significant; * P < 0.05; ** P < 0.01. Dashed lines represent the epithelial barrier. White boxes in the overlay images correspond to the magnified inset images. Scale bars, 10 μ m.

in both the ileum and colon, confirming that tuft cells are distinct among epithelial cells in their expression of CD300lf (Fig. 2C).

Given these findings, we assessed whether MNoV^{CR6} infects tuft cells. Immunofluorescence microscopy on intestines of WT mice infected with MNoV^{CR6} revealed rare cells expressing the MNoV nonstructural protein NS6/7 (Fig. 3A). These cells were in direct contact with the intestinal lumen and were observed in the surface epithelium of the colon and in both the villi and crypts of the ileum. All MNoV NS6/7-positive cells coexpressed the tuft cell marker DCLK1. No viral antigen-positive cells were observed in the lamina propria or immune cells. Similar histologic findings and viral tropism were identified in WT germ-free mice (fig. S1), indicat-

ing that intestinal bacteria are not required for either CD300lf expression by tuft cells or MNoV^{CR6} infection of tuft cells. To confirm and quantify MNoV^{CR6} infection of tuft cells, we performed flow cytometry analysis of colonic epithelial cells from infected Gfi1b-GFP mice. MNoV^{CR6} infection did not significantly reduce tuft cell frequency (Fig. 3B). Infected cells were defined as those expressing two independent viral nonstructural proteins (NS1/2 and NS6/7) (9). We observed 128 ± 33 (mean \pm SEM) infected tuft cells per million live epithelial cells (EpCAM⁺CD45⁺). We did not observe infection of non-tuft epithelial cells (Fig. 3, C and D). Overall, $1.4 \pm 0.37\%$ of Gfi1b-GFP⁺ tuft cells were MNoV infected. Together, our immunofluorescence and flow cytometric analyses in-

dicate that tuft cells are the physiologic target cell of MNoV in WT animals. This finding likely explains why we did not observe clusters of infected cells in the intestine, as tuft cells are isolated from one another, being surrounded by other intestinal epithelial cells (9).

Given the role of tuft cells in type 2 immunity, we hypothesized that there might be an intimate relationship between type 2 immunity and enteric norovirus infection. The type 2 cytokines IL-4 and IL-25 induce tuft cell hyperplasia (20–22). Therefore, we assessed whether these cytokines augmented MNoV transmission. WT mice were treated with IL-4, IL-25, or a phosphate-buffered saline (PBS) control prior to peroral challenge with a low dose [4.25×10^4 plaque-forming units (PFU) per mouse] of MNoV^{CR6} insufficient to

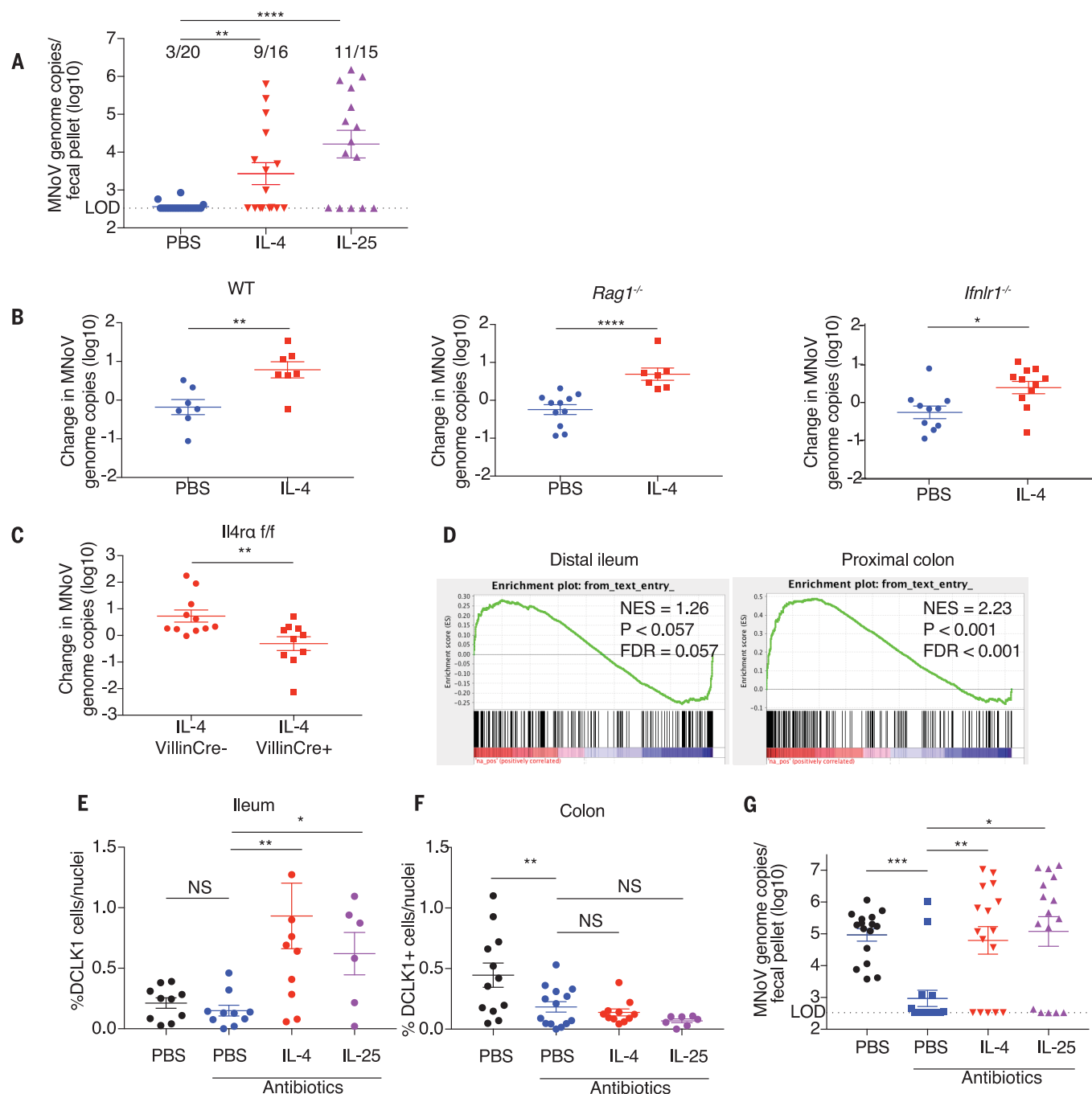


Fig. 4. Tuft cell tropism determines transkingdom interactions of MNoV. (A) WT mice were injected intraperitoneally with PBS, IL-4, or IL-25 prior to peroral challenge with a low dose (4.25×10^4 PFU) of MNoV^{CR6}. Both IL-4 and IL-25 increase MNoV transmission, as measured by detection of MNoV genomes in feces 7 days postinfection. The numbers above each column reflect the number of infected animals relative to the total number of animals per group (chi-square test, <0.0015). (B) WT, *Rag1*^{-/-}, and *Ifnlr1*^{-/-} mice chronically infected with a high dose (10^6 PFU) of MNoV^{CR6} for 21 days were administered PBS or IL-4. MNoV fecal shedding significantly increased in WT, *Rag1*^{-/-}, and *Ifnlr1*^{-/-} mice after IL-4 injection (24 days postinfection) compared with that after PBS administration. (C) IL-4 enhancement of MNoV fecal shedding during chronic infection requires *Il4ra* expression on VillinCre-expressing epithelial cells. (D) Broad-spectrum antibiotics (vancomycin, neomycin, ampicillin, and metronidazole), which prevent MNoV^{CR6} infection, significantly reduce

tuft cell-specific gene transcripts as measured by RNA-seq in the colon but not in the ileum. NES, normalized enrichment score; FDR, false discovery rate. (E and F) DCLK1⁺ tuft cells were quantified by immunofluorescence microscopy. Antibiotics reduce DCLK1⁺ tuft cells in the colon but not in the ileum. IL-4 and IL-25 increase DCLK1⁺ tuft cells in the ileum but not in the colon. (G) Antibiotic pretreatment prevents MNoV^{CR6} infection. This antiviral state can be reversed with IL-4 or IL-25 administration prior to MNoV^{CR6} challenge. Shown are means \pm SEM. NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$. Data in mouse experiments are pooled from at least three independent experiments with two to six mice per group, except for the *Ifnlr1*^{-/-} study, in which data are pooled from two independent experiments. Each dot in (E) and (F) represents the tuft cell frequency in one mouse. At least 10 independent low-power images were averaged per mouse. Data were analyzed by the Mann-Whitney *U* test unless otherwise indicated.

establish robust infection in the majority of control mice. Both IL-4- and IL-25-treated animals were significantly more likely to be productively infected than PBS-treated animals, as measured by numbers of viral genome copies in the feces 7 days postchallenge (Fig. 4A). These results show that type 2 immune responses can enhance enteric viral transmission. We therefore asked whether type 2 cytokines affect MNoV^{CR6} fecal shedding during persistent infection. WT mice were challenged perorally with a high dose (10^6 PFU/mouse) of MNoV^{CR6} that is sufficient to infect all animals. After at least 21 days of infection, IL-4 or a PBS control was injected intraperitoneally and fecal shedding of virus was monitored. IL-4 significantly increased MNoV^{CR6} fecal shedding as detected 1 day after the second and final IL-4 injection (Fig. 4B). As in WT animals, IL-4 increased MNoV shedding in *Rag1*^{-/-} and *Ifn λ 1*^{-/-} mice persistently infected with MNoV (Fig. 4B), indicating that cytokine promotion of infection was not caused by effects on T cells or B cells and is independent of IFN- λ -induced innate immune signaling, a potent regulator of intestinal norovirus infection (5). In addition, we demonstrated that IFN- λ treatment did not alter tuft cell abundance in the intestine (fig. S3).

The murine parasite *Trichinella spiralis* induces type 2 inflammation and augments MNoV^{CR6} infection (7). The mechanism of action was previously hypothesized to be increased viral replication in alternatively activated macrophages exposed to type 2 cytokines such as IL-4 and IL-13 (7). However, in this study we showed that tuft cells and not macrophages are the target cell for MNoV^{CR6}. Thus, we tested whether the enhanced MNoV^{CR6} infection resulting from IL-4 treatment was mediated by effects of this cytokine on epithelial cells. To test this hypothesis, we generated epithelial cell-specific IL-4 receptor α (IL4 α) conditional knockout mice (IL4 α ^{eff} \times VillinCre) (27). Mice were infected with MNoV for at least 21 days, after which IL-4 was administered. IL-4 enhanced MNoV^{CR6} shedding in IL4 α ^{eff} \times VillinCre⁻ animals but not IL4 α ^{eff} \times VillinCre⁺ animals, demonstrating that IL-4 signals through its receptor on epithelial cells (Fig. 4C) (22). These data suggest that IL-4 promotes norovirus infection via effects on tuft cells, the only epithelial cells infected with the virus.

Prior work showed that the bacterial microbiome is required for efficient establishment of enteric MNoV^{CR6} infection (6). Specifically, broad-spectrum antibiotics that deplete intestinal bacteria prevent MNoV^{CR6} transmission and persistent infection (6). The mechanism for this effect is incompletely understood. We therefore asked whether antibiotic treatment affected expression of tuft cell-specific transcripts. RNA-seq was performed on control and antibiotic-treated mice, and the expression of a list of curated tuft cell genes was used to assess differences in tuft cell-specific genes (24). Antibiotic treatment resulted in a decrease in tuft cell-specific gene expression in the colon (normalized enrichment score, 2.23; $P < 0.001$; false discovery

rate, <0.001) (Fig. 4D); changes in tuft cell genes did not reach statistical significance in the ileum. Consistent with the RNA-seq gene set enrichment analysis, antibiotics decreased DCLK1⁺ cells in the colon but not the ileum (Fig. 4, E and F). IL-4 and IL-25 induced tuft cell hyperplasia in the ilea of antibiotic-treated mice, whereas colonic tuft cells were not increased in number by IL-4 or IL-25 (Fig. 4, E and F). These findings indicate that both type 2 cytokines and intestinal bacteria regulate tuft cells, albeit in a tissue-specific manner (22). The observation that intestinal bacteria contribute to tuft cell regulation in vivo raised the question of whether the antiviral role of antibiotics could be overcome with administration of type 2 cytokines that act on epithelial cells to control MNoV infection. WT mice were pretreated with antibiotics for 2 weeks prior to challenge with a high dose (10^6 PFU) of MNoV^{CR6}. Consistent with prior findings, antibiotics significantly reduced MNoV^{CR6} infection (Fig. 4G) (6). IL-4 or IL-25 administration, prior to MNoV^{CR6} challenge, rescues viral infection in antibiotic-treated mice. Both IL-4 and IL-25 significantly increased both the proportion of mice infected with virus and the magnitude of fecal shedding (Fig. 4G). The differential regulation of tuft cells by type 2 cytokines and antibiotics in the ileum and colon, respectively, suggests that a threshold number of tuft cells may matter more than the anatomic location of tuft cells within the intestine.

Here, we have identified intestinal tuft cells as the physiologic target cell of MNoV. This discovery has important implications for our understanding of transkingdom interactions and the pathogenesis of persistent intestinal infection. Norovirus infection triggers inflammatory bowel disease-like phenotypes in genetically susceptible hosts (3, 4). Now that we have identified the tropism of norovirus for tuft cells, a question to consider is whether tuft cells regulate inflammatory bowel disease-like phenotypes. Tuft cell tropism also links the proviral effects of helminths and commensal bacteria, which increase tuft cells in the ileum and colon, respectively (20–22). Noroviruses can persist in the intestine for months in both mice and humans (2, 28–30). This persistent infection is resistant to both antibody and CD8⁺ T cell-mediated clearance, yet the mechanism of immune evasion is unknown (31). Our identification of MNoV tropism for tuft cells suggests that tuft cells represent an immune-privileged site for enteric viral infection in mice. It is possible that other viruses also infect tuft cells, enabling these viruses to take advantage of type 2 immune responses to promote infection.

REFERENCES AND NOTES

1. M. M. Patel et al., *Emerg. Infect. Dis.* **14**, 1224–1231 (2008).
2. S. M. Karst, C. E. Wobus, I. G. Goodfellow, K. Y. Green, H. W. Virgin, *Cell Host Microbe* **15**, 668–680 (2014).
3. K. Cadwell et al., *Cell* **141**, 1135–1145 (2010).
4. M. Basic et al., *Inflamm. Bowel Dis.* **20**, 431–443 (2014).
5. T. J. Nice et al., *Science* **347**, 269–273 (2015).
6. M. T. Baldrige et al., *Science* **347**, 266–269 (2015).
7. L. C. Osborne et al., *Science* **345**, 578–582 (2014).

8. E. Kernbauer, Y. Ding, K. Cadwell, *Nature* **516**, 94–98 (2014).
9. S. Lee et al., *Cell Host Microbe* **22**, 449–459.e4 (2017).
10. R. C. Orchard et al., *Science* **353**, 933–936 (2016).
11. K. Haga et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, E6248–E6255 (2016).
12. C. E. Wobus et al., *PLOS Biol.* **2**, e432 (2004).
13. T. J. Nice, D. W. Strong, B. T. McCune, C. S. Pohl, H. W. Virgin, *J. Virol.* **87**, 327–334 (2013).
14. D. W. Strong, L. B. Thackray, T. J. Smith, H. W. Virgin, *J. Virol.* **86**, 2950–2958 (2012).
15. S. M. Karst, C. E. Wobus, M. Lay, J. Davidson, H. W. Virgin IV, *Science* **299**, 1575–1578 (2003).
16. K. R. Grau et al., *Nat. Microbiol.* **2**, 1586–1591 (2017).
17. D. H. Chung et al., *J. Immunol.* **171**, 6541–6548 (2003).
18. H. Xi et al., *J. Exp. Med.* **207**, 7–16 (2010).
19. F. Gerbe, C. Legraverend, P. Jay, *Cell. Mol. Life Sci.* **69**, 2907–2917 (2012).
20. F. Gerbe et al., *Nature* **529**, 226–230 (2016).
21. M. R. Howitt et al., *Science* **351**, 1329–1333 (2016).
22. J. von Moltke, M. Ji, H. E. Liang, R. M. Locksley, *Nature* **529**, 221–225 (2016).
23. F. Gerbe, B. Brulin, L. Makrini, C. Legraverend, P. Jay, *Gastroenterology* **137**, 2179–2180 (2009).
24. D. Grün et al., *Nature* **525**, 251–255 (2015).
25. K. S. Yan et al., *Nature* **545**, 238–242 (2017).
26. M. Bjerknes et al., *Dev. Biol.* **362**, 194–218 (2012).
27. M. Mohrs et al., *J. Immunol.* **162**, 7302–7308 (1999).
28. G. Phillips, C. C. Tam, L. C. Rodrigues, B. Lopman, *Epidemiol. Infect.* **138**, 1454–1458 (2010).
29. J. Ayukekbong et al., *J. Med. Virol.* **83**, 2135–2142 (2011).
30. F. H. Sukhrie, J. J. Siebenga, M. F. Beersma, M. Koopmans, *J. Clin. Microbiol.* **48**, 4303–4305 (2010).
31. V. T. Tomov et al., *Immunity* **47**, 723–738.e5 (2017).

ACKNOWLEDGMENTS

We acknowledge J. Guo, X. Zhang, A. Orvedahl, A. Mayer, T. Schaff, M. Artymov, G. Broze, and N. Lasky for helpful discussions and technical support. We also thank W. Beatty at the Molecular Microbiology Imaging Facility and E. Lantelme and D. Brinja at the Flow Cytometry Core at Washington University School of Medicine. **Funding:** This work was supported by NIH grants K08 AI128043 (C.B.W.); K22 AI127846 (M.T.B.); K99 DK116666 (R.C.O.); T32 AI007163 (C.C.Y.); R01 AI127552 and U19 AI109725 (H.W.V.); and DK093668, DK103788, HL123340, and R01 AI121244 (J.A.N. and K.C.). In addition, S.L. was supported by the Basic Sciences Research Program through the National Research Foundation of Korea, funded by the Ministry of Education (grant NRF-2016R1A6A3A03012352). M.T.B. was supported by Digestive Diseases Research Core Centers grant P30 DK052574. J.A.N. was supported by a Vilcek Fellowship and a Sir Keith Murdoch Fellowship. **Author contributions:** C.B.W., S.L., L.L.H., M.R.M., D.R.B., C.C.Y., Y.-T.W., D.A.M., D.K., J.A.N., K.C., and M.T.B. performed experiments. C.B.W., S.L., R.C.O., C.A.H., K.C., P.M.A., M.R.H., S.A.H., M.V.L.C., M.T.B., and H.W.V. designed the project. C.B.W., S.L., L.L.H., R.C.O., C.D., B.L.H., T.F., J.R.B., H.W.V., and S.A.H. analyzed data. C.B.W. wrote the paper. All authors read and edited the manuscript. **Competing interests:** C.B.W., R.C.O., and H.W.V. are inventors on a patent application submitted by Washington University entitled “Receptor for norovirus and uses thereof” (U.S. provisional application 62/301,965). M.V.L.C. is an employee of Genentech, a for-profit institution. **Data and materials availability:** The data from this study are tabulated in the main paper and supplementary materials. All reagents are available from H.W.V. under a material transfer agreement with Washington University. All data and code to understand and assess the conclusions of this research are available in the main text and supplementary materials and via the European Nucleotide Archive under accession no. PRJEB23132.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/204/suppl/DC1
Materials and Methods
Figs. S1 to S3
Table S1
References (32–39)

3 November 2017; accepted 5 March 2018
10.1126/science.aar3799

CARBON CYCLE

Microbial oxidation of lithospheric organic carbon in rapidly eroding tropical mountain soils

Jordon D. Hemingway,^{1,2*} Robert G. Hilton,³ Niels Hovius,^{4,5} Timothy I. Eglinton,⁶ Negar Haghipour,⁶ Lukas Wacker,⁷ Meng-Chiang Chen,⁸ Valier V. Galy¹

Lithospheric organic carbon (“petrogenic”; OC_{petro}) is oxidized during exhumation and subsequent erosion of mountain ranges. This process is a considerable source of carbon dioxide (CO_2) to the atmosphere over geologic time scales, but the mechanisms that govern oxidation rates in mountain landscapes are poorly constrained. We demonstrate that, on average, $67 \pm 11\%$ of the OC_{petro} initially present in bedrock exhumed from the tropical, rapidly eroding Central Range of Taiwan is oxidized in soils, leading to CO_2 emissions of 6.1 to 18.6 metric tons of carbon per square kilometer per year. The molecular and isotopic evolution of bulk OC and lipid biomarkers during soil formation reveals that OC_{petro} remineralization is microbially mediated. Rapid oxidation in mountain soils drives CO_2 emission fluxes that increase with erosion rate, thereby counteracting CO_2 drawdown by silicate weathering and biospheric OC burial.

Erosion-induced weathering in collisional mountain belts is a major carbon-cycle regulator over million-year time scales and provides a link between tectonics and climate (1, 2). Atmospheric CO_2 is consumed by the export and burial in marine sediments of biospheric organic carbon (OC_{bio}) and carbonate minerals precipitated after silicate rock weathering (1). The CO_2 drawdown flux associated with both processes increases with erosion rate (3, 4), highlighting the importance of steep, erosive orogens in driving CO_2 drawdown. By comparison, CO_2 release during exhumation and erosion has received considerably less attention, despite its potential to partially or fully negate the effects of geological CO_2 consumption (1, 5, 6). Oxidative weathering of sulfide minerals (coupled with carbonate dissolution) and lithospheric, or “petrogenic,” organic carbon (OC_{petro}) contained in exhumed rocks can increase atmospheric CO_2 and decrease O_2 concentrations over geologic time scales (1, 7–9). However, the mechanisms that govern oxidation rates and efficiencies in mountain belts are underconstrained (5, 8, 9).

To better constrain orogenic CO_2 emissions, we assessed the controls on OC_{petro} oxidation and export in the Central Range of Taiwan, one of the fastest-exhuming and -eroding mountain belts on Earth (10). Steep relief (11), frequent typhoon landfalls (10), and high bedrock landslide rates (11) lead to long-term erosion rates

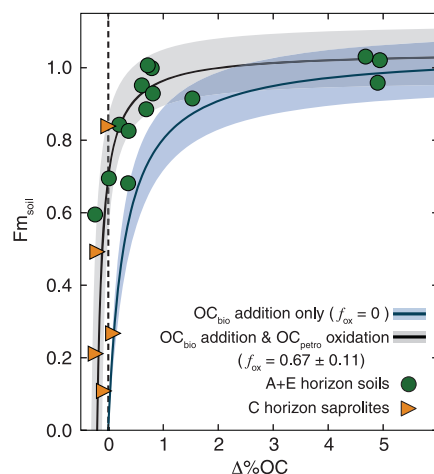


Fig. 1. Evidence for bedrock OC oxidation. The blue line is the solution to Eq. 2, assuming no OC_{petro} oxidation during soil formation (OC_{bio} addition only; $f_{\text{ox}} = 0$). The black line is the orthogonal distance regression best-fit solution that minimizes the residual error between measured (green circles and orange triangles) and predicted Fm_{soil} values. The shaded region around each line is the propagated $\pm 1\sigma$ uncertainty (14). Best-fit results indicate that $67 \pm 11\%$ of bedrock OC is lost during oxidative weathering. $\Delta\%OC = 0$ is shown as a vertical dashed line. Measurement error bars ($\pm 1\sigma$) are smaller than marker sizes.

of 3 to 6 mm year^{−1} across the range (10). Although supplemental contributions from deeper in the exhumation path are likely, weathering in such mountain landscapes occurs primarily on hillslopes and in colluvial deposits (12, 13). We therefore assessed OC molecular and isotopic evolution in multiple hillslope soil profiles located in the LiWu and WuLu River basins (fig. S1) and verified these observations at the catchment scale by using LiWu River suspended sediments (14). Soils at our study sites are ≤ 1 m thick, including mineral (A and E) and saprolite (C) layers (15); experience residence times on the order of centuries (14); and overlay bedrock ranging from Mesozoic greenschist and amphibolite at low elevations (Tananao schists) to Cenozoic slate and phyllite near the Lishan fault (Pilushan and Lushan Formations) (16). All lithologies are carbonaceous, with bedrock outcrops containing 0.2 to 0.7% OC_{petro} (table S1) (17).

We observed substantial OC_{petro} loss in all soil profiles, as evidenced by the relationship between soil OC content ($\%OC_{\text{soil}}$) and ^{14}C activity (expressed as “fraction modern” or Fm) (14). To account for differences in $\%OC$ between bedrock lithologies (17), $\%OC_{\text{soil}}$ is expressed as

$$\Delta\%OC = \%OC_{\text{soil}} - \%OC_{\text{bedrock}} \quad (1)$$

where $\%OC_{\text{bedrock}}$ is the OC content of bedrock immediately underlying each soil sample. The average fraction of bedrock OC that is oxidized during soil formation, f_{ox} , can then be quantified by utilizing the fact that OC_{petro} is inherently ^{14}C -free ($Fm_{\text{petro}} = 0.0$) and setting $Fm_{\text{bio}} = 1.045 \pm 0.079$, the measured ^{14}C activity of vascular plant-wax fatty acids extracted from A- and E-horizon soils (table S2) (14). Soil OC is treated as a mixture of OC_{bio} and residual OC_{petro} , leading to the equation (14)

$$Fm_{\text{soil}} = Fm_{\text{bio}} \left[\frac{\Delta\%OC + (f_{\text{ox}})(\%OC_{\text{bedrock}})}{\Delta\%OC + \%OC_{\text{bedrock}}} \right] \quad (2)$$

Fm_{soil} is a hyperbolic function of $\Delta\%OC$ with curvature that is defined by both $\%OC_{\text{bedrock}}$ and f_{ox} , as shown in Fig. 1. We simultaneously solved Eq. 2 for the best-fit $\%OC_{\text{bedrock}}$ and f_{ox} values using orthogonal distance regression and accounted for uncertainty using Monte Carlo resampling (14).

On average, $67 \pm 11\%$ ($\pm 1\sigma$) of bedrock OC is lost during soil formation, a minimum estimate because deep weathering has likely already removed OC from the initial bedrock (18). To test whether observed $\%OC$ trends simply reflect mobile element losses during weathering and not oxidation per se, we solved Eq. 2 for a subset of samples after normalizing OC content to the immobile element titanium (table S1) (14). Calculated f_{ox} values using normalized and unnormalized data are identical within uncertainty, indicating no appreciable mobility effect on our results (fig. S2).

Assuming that all OC lost is oxidized to CO_2 (8), f_{ox} can be used to estimate the steady-state

¹Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution (WHOI), 266 Woods Hole Road, Woods Hole, MA 02543, USA. ²Massachusetts Institute of Technology (MIT)–WHOI Joint Program in Oceanography and Applied Ocean Science and Engineering, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. ³Department of Geography, Durham University, South Road, Durham DH1 3LE, UK. ⁴GFZ German Research Center for Geoscience, Telegrafenberg, Potsdam 14473, Germany. ⁵Department of Earth and Environmental Sciences, University of Potsdam, Karl-Liebknechtstraße 24, Golm 14476, Germany. ⁶Geological Institute, Department of Earth Sciences, ETH Zürich, Sonneggstraße 5, Zürich 8092, Switzerland. ⁷Laboratory of Ion Beam Physics, Department of Physics, ETH Zürich, Otto-Stern-Weg 5, Zürich 8092, Switzerland. ⁸Taroko National Park Headquarters, Fu-Su Village, Hualien 972, Taiwan. *Corresponding author. Email: jordan_hemingway@fas.harvard.edu †Present address: Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, MA 02138, USA.

CO₂ emission flux from soils owing to OC_{petro} oxidation, termed Φ_{ox} according to

$$\Phi_{ox} = \frac{(f_{ox})(\% \text{ OC}_{\text{bedrock}})(\rho_{\text{soil}})(z_{\text{soil}})}{\tau_{\text{soil}}} \quad (3)$$

where ρ_{soil} is the soil density, z_{soil} is the soil thickness (15), and τ_{soil} is the soil residence time on hillslopes. We estimated τ_{soil} using three independent methods (landslide rates, catchment-average denudation rates, and OC_{bio} erosion rates) and incorporated uncertainty for each variable in Eq. 3 using Monte Carlo resampling across the range of observed values (14), resulting in a median Φ_{ox} range of 6.1 to 18.6 metric tons C km⁻² year⁻¹ for conditions that are prevalent across the Central Range (fig. S3A) (14). We emphasize that Φ_{ox} is a minimum estimate of total CO₂ emissions by OC_{petro} oxidation owing to the potential for OC losses occurring during deep weathering (18). Still, this flux is statistically identical to two independent, catchment-integrated OC_{petro} oxidation estimates for Taiwanese rivers—based on fluvial OC_{petro} export (≤ 12 metric tons C km⁻² year⁻¹) (19) and dissolved rhenium yield (7 to 13 metric tons C km⁻² year⁻¹; fig. S3B) (5)—and is two to six times as high as estimates of CO₂ drawdown by silicate weathering in the LiWu catchment (fig. S3C) (18). The observation that Φ_{ox} matches catchment-integrated emissions implies that OC_{petro} oxidation in Taiwan occurs predominantly in rapidly eroding hillslope soils.

A saprolite depth profile collected from the WuLu catchment indicates that bedrock OC can be oxidized and replaced with OC_{bio} before A and E horizons have fully developed. Two samples collected at 0.5 and 0.2 m depth contain similar OC concentrations (0.20 and 0.28%, respectively) but have drastically different Fm values (0.108 and 0.839, respectively; table S1). Rapid OC_{petro} oxidation can occur (i) abiotically without chemical alteration, (ii) abiotically with chemical alteration, (iii) biotically without chemical alteration, or (iv) biotically with chemical alteration and ¹⁴C-depleted biomass production (20–22). To assess alteration and track multiple OC sources within a single sample, we used Ramped PyrOx (RPO) serial combustion (23). This technique heats each sample at a constant ramp rate to separate OC on the basis of thermal lability and determines Fm values for specific temperature intervals (termed RPO fractions) (14). To quantitatively compare OC chemical structure, we determined the underlying thermal activation energy (E) distribution for each sample, termed $p(0, E)$, because this is an intrinsic property of carbon bond strength and thus a proxy for chemical composition (23). Unlike ¹⁴C activity, end-member mixing does not shift OC in E space. Mixing OC_{bio} with unaltered OC_{petro} will thus result in a bimodal $p(0, E)$ distribution, whereas chemical alteration is required to explain the presence of intermediate E values (14, 23).

We constrained bedrock E by using particulate OC (POC) from 27 suspended sediment samples, including isolated ≥ 2 -mm clasts, collected from the LiWu River during four typhoon

events (14). Because sediment exported during typhoons is dominated by material sourced from bedrock incision, distributed runoff erosion, and landsliding throughout the basin (11, 12), we expect this sample set to integrate outcropped

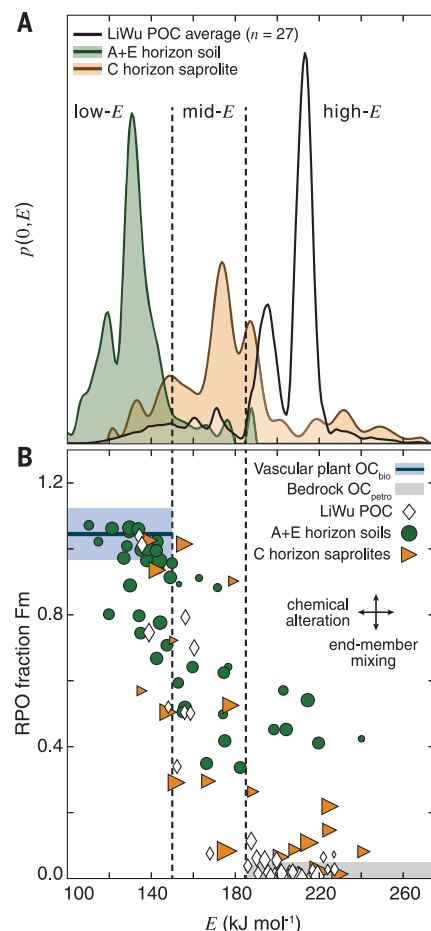


Fig. 2. Evidence for OC_{petro} chemical alteration.

(A) Representative $p(0, E)$ distributions highlighting the differences between OC end-members: LiWu POC exported during typhoon events (average, $n = 27$; black), organic-rich A- and E-horizon topsoil (green), and C-horizon saprolite (orange). Each $p(0, E)$ distribution integrates to unity (y -axis values not shown) (14, 23). (B) E versus Fm relationships for all soils (green circles and orange triangles) and LiWu POC (white diamonds) in which RPO-fraction ¹⁴C activity was measured. Marker sizes represent the relative amount of total OC contained in each RPO fraction. Constraints on end-member E and Fm ranges are described in the text (blue, vascular plant OC_{bio}; gray, OC_{petro}). Black arrows represent theoretical trends for end-member mixing (vertical) and chemical alteration (horizontal) (23) and indicate that alteration is necessary to explain the presence of mid- E OC. In both panels, dashed lines separate OC into low- E (< 150 kJ mol⁻¹), mid- E ($150 \leq E < 185$ kJ mol⁻¹), and high- E (≥ 185 kJ mol⁻¹) regions. Fm error bars ($\pm 1\sigma$) are smaller than marker sizes.

bedrock lithologies that contain relatively unweathered OC_{petro}. This is supported by bulk POC ¹³C content (expressed as $\delta^{13}\text{C}$ values) and total nitrogen to POC ratios (table S3), which span the range of Tananao schist, Lushan Formation, and Pilushan Formation values (17). Figure 2A shows that bedrock OC is exclusively associated with $E \geq 185$ kJ mol⁻¹ (termed high- E ; fig. S3A) (14), consistent with the observed partial graphitization of this material (16). We additionally constrained vascular-plant OC $p(0, E)$ by using two organic-rich ($\geq 5\%$) surface soils characterized by bulk Fm values similar to those of plant-wax fatty acids (14). For both samples, $\geq 90\%$ of OC is associated with $E < 150$ kJ mol⁻¹ (termed low- E), indicating that OC_{bio} and OC_{petro} are effectively separated in E space.

Energy distributions and ¹⁴C activity in soil and saprolite materials provide strong evidence for OC_{petro} chemical alteration during weathering. Up to 51% of OC contained in saprolites and deep A and E horizons lies between 150 and 185 kJ mol⁻¹ (termed mid- E ; table S4 and fig. S4, B and C)—higher than values corresponding to vascular plant OC (< 150 kJ mol⁻¹) yet lower than those for bedrock OC (≥ 185 kJ mol⁻¹). This observation could result from either (i) increasing vascular plant OC E by stabilization during aging in soils (24) or (ii) decreasing residual OC_{petro} E during oxidative weathering (20, 21). We assessed the relative importance of these mechanisms by using the ¹⁴C activity of each RPO fraction (table S5). As shown in Fig. 2B, low- E Fm values cluster near those of plant-wax fatty acids, whereas high- E material approaches an Fm value of zero. Meanwhile, mid- E OC spans an Fm range from 0.083 ± 0.002 to 0.912 ± 0.008 . We rule out the possibility that ¹⁴C-depleted mid- E OC exclusively reflects OC_{bio} aging because (i) this would require a biospheric component that has aged up to 20,000 ¹⁴C years, much longer than the centennial soil residence times in Taiwan (14), and (ii) plant-wax fatty acids were not detected in some saprolite samples (table S6). Thus, mid- E material must reflect a mixture of weathered OC_{petro} and moderately aged OC_{bio}.

We treat OC_{petro} that has been chemically altered during weathering as a distinct end-member described by Fm = 0.0 and a value of f_{mid} , the fraction of $p(0, E)$ contained within the mid- E range, which is greater than the highest observed saprolite value of 0.51 (14). Figure 3A shows that all hillslope samples, with the exception of one unweathered saprolite, are adequately explained by a mixture of OC_{bio} and chemically altered OC_{petro}. This end-member is also present in LiWu River POC collected during typhoon floods, as evidenced by the divergence from a vertical mixing line between OC_{petro} and OC_{bio} in Fig. 3A. Therefore, along with unweathered bedrock OC (19) sourced from deep incision and landsliding (11), our approach detected catchment-scale export of chemically altered OC_{petro} from Central Range hillslopes during typhoon flood events. Because calculated f_{mid} depends on our choice of mid- E range (here, 150 to 185 kJ mol⁻¹), it is possible that mixing

trends and end-member compositions are sensitive to changes in E boundary values. We tested this sensitivity by allowing these boundary values to vary by $\pm 10 \text{ kJ mol}^{-1}$ (14). Although quantitative differences exist (fig. S5), the resultant mixing trends are qualitatively robust, indicating that the importance of chemically altered OC_{petro} is insensitive to our choice of mid- E boundary values.

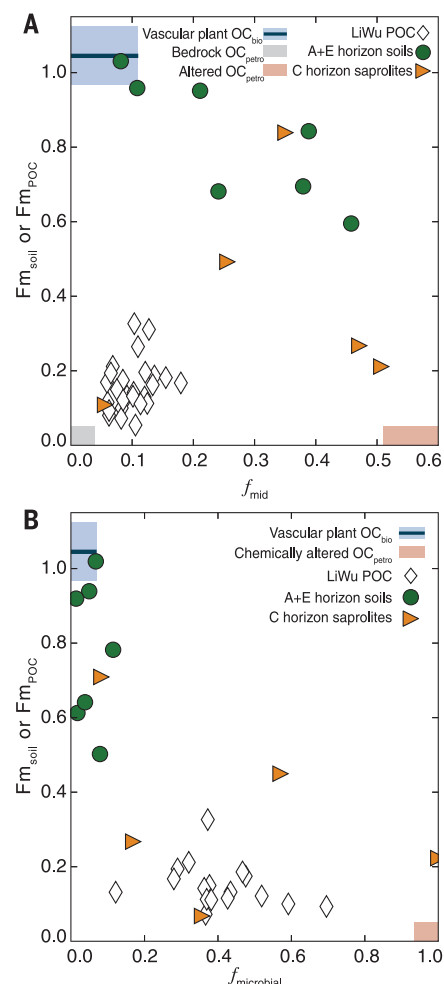


Fig. 3. Evidence for microbially mediated bedrock OC oxidation. (A) Bulk Fm versus f_{mid} relationships for soils (green circles and orange triangles) and LiWu POC (white diamonds). All soils, with the exception of the 0.5-m saprolite discussed in the text, are described by a mixing line between vascular plant OC_{bio} (blue) and chemically altered OC_{petro} (red) (14). LiWu River POC is dominated by bedrock OC_{petro} (gray) but does contain detectable chemically altered OC_{petro} , as evidenced by the deviation from a vertical mixing line between OC_{bio} and OC_{petro} . (B) Bulk Fm versus $f_{\text{microbial}}$ relationships for all samples in which fatty acid concentrations were analyzed (14). The relative abundance of microbial fatty acids increases with decreasing Fm across all samples, suggesting that microbial respiration is the source of chemically altered OC_{petro} . Measurement error bars ($\pm 1\sigma$) are smaller than marker sizes.

Fatty acid molecular distributions and $\delta^{13}\text{C}$ values imply that rapidly oxidized OC_{petro} in soils is incorporated into microbial biomass, supporting laboratory-based incubation studies (20, 22). We calculated $f_{\text{microbial}}$, the fraction of total fatty acids that are microbial in origin (25, 26), as a proxy for the relative abundance of heterotrophic versus vascular plant biomass (14). This approach excludes fungal contributions and is thus a minimum estimate of heterotrophic biomass. Figure 3B shows that bulk Fm is negatively correlated with $f_{\text{microbial}}$ across all soil and POC samples. We do not expect this trend to be linear owing to fatty acid production biases (25, 26). Still, this relationship suggests that heterotrophic biomass is more abundant in samples containing predominantly ^{14}C -free OC.

Sample limitation prevented measurement of microbial fatty acid ^{14}C activity (14), but the $\delta^{13}\text{C}$ values imply that bedrock OC is used as substrate (table S7) (26, 27). Bulk OC and plant-wax fatty acid $\delta^{13}\text{C}$ values correlate strongly in A and E horizons (coefficient of determination $r^2 = 0.959$; $P < 0.001$; $n = 7$), reflecting the predominance of OC_{bio} in these samples, but are uncorrelated in C horizons ($P > 0.05$; $n = 4$) owing to a lack of OC_{bio} contribution to saprolites (fig. S6). Still, if OC_{bio} were the sole substrate for heterotrophs, then microbial and plant-wax fatty acid $\delta^{13}\text{C}$ values should correlate strongly with a constant $\delta^{13}\text{C}$ offset (27) in all samples. This is not apparent in either A- and E-horizon ($P > 0.05$; $n = 7$) or saprolite ($P > 0.05$; $n = 4$) samples, indicating that vascular plant OC cannot be the only substrate. Rather, this lack of correlation requires a secondary microbial carbon source (20–22), namely, bedrock OC. We conclude that mid- E , ^{14}C -free material is a product of microbial bedrock oxidation, produced either directly by extracellular enzymes or indirectly after acid hydrolysis (20), and is manifest as ^{14}C -depleted living biomass (22) or as residual, chemically altered OC_{petro} (21).

Substantial bedrock OC replacement in saprolites implies that considerable weathering occurs $\leq 1 \text{ m}$ below the surface and that microbially mediated OC_{petro} oxidation can proceed at a pace matching the rapid exhumation in Taiwan. We propose that exhumation and hillslope erosion rates exert a first-order control on CO_2 emissions from OC_{petro} oxidation, because faster erosion will increase the rate of bedrock exposure to the weathering front (8). This is further supported by measurements of the dissolved rhenium flux from Taiwanese rivers, a proxy for OC_{petro} oxidation, which increases with erosion rate (5). However, the relationship between OC_{petro} oxidation and physical erosion rate cannot be linear. Large earthquakes and typhoons are known to cause widespread bedrock landsliding (28–30) and elevated export of OC_{petro} by rivers (19). Such events increase catchment-averaged erosion rates (28) but could decrease catchment-averaged OC_{petro} oxidation efficiency by bypassing the hillslope soil weathering window. OC_{petro} remineralization in Taiwan is incomplete, as evidenced by the abundance of

bedrock OC in sediments exported by rivers (19) and deposited in nearby coastal margins (31). We predict a dampened response of OC_{petro} -derived CO_2 emissions to further increases in erosion rates, because increasing landslide rates will result in less catchment area available for soil formation and weathering.

Microbially mediated oxidative weathering in Taiwanese hillslope soils offsets geologic CO_2 drawdown and O_2 production by silicate weathering and OC_{bio} burial (1, 5, 8, 22). The Φ_{ox} values calculated here are similar in magnitude to CO_2 source estimates from sulfide oxidation ($\geq 22.9 \pm 1.0 \text{ metric tons C km}^{-2} \text{ year}^{-1}$; LiWu basin only) (9) and to CO_2 sinks from silicate weathering ($3.1 \pm 0.1 \text{ metric tons C km}^{-2} \text{ year}^{-1}$; LiWu basin only; fig. S3C) (18) and OC_{bio} burial ($21 \pm 10 \text{ metric tons C km}^{-2} \text{ year}^{-1}$; Taiwan average; fig. S3D) (14, 32). This process is likely globally important, given that rapid soil formation is observed in other tropical and temperate orogenic settings such as the Southern Alps of New Zealand (33). We therefore hypothesize that CO_2 consumption is not favored in highly erosive mountain belts dominated by OC- and sulfide-rich low- and intermediate-grade metasedimentary lithologies. This comes from the observation that OC_{petro} and sulfide mineral oxidation is not limited by reaction kinetics even at high erosion rates (5, 8, 21), unlike silicate weathering and OC_{bio} export (4, 34). Conversely, the magnitude of the net CO_2 sink likely increases with physical erosion rate in orogens dominated by high-grade metamorphic and igneous rocks owing to their lower OC_{petro} and sulfide contents. Although the global fluxes and the time scales over which they impact atmospheric CO_2 and O_2 concentrations remain to be assessed, our results demonstrate the importance of microbially mediated OC_{petro} oxidation and its relationship to tectonic and erosive controls on the global carbon cycle and Earth's long-term climate.

REFERENCES AND NOTES

1. N. M. Bergman, T. M. Lenton, A. J. Watson, *Am. J. Sci.* **304**, 397–437 (2004).
2. P. Molnar, P. England, *Nature* **346**, 29–34 (1990).
3. J. Gaillardet, B. Dupré, P. Louvat, C. J. Allègre, *Chem. Geol.* **159**, 3–30 (1999).
4. V. Galy, B. Peucker-Ehrenbrink, T. Eglington, *Nature* **521**, 204–207 (2015).
5. R. G. Hilton, J. Gaillardet, D. Calmels, J.-L. Birck, *Earth Planet. Sci. Lett.* **403**, 27–36 (2014).
6. M. A. Torres et al., *Earth Planet. Sci. Lett.* **450**, 381–391 (2016).
7. R. A. Berner, K. Caldeira, *Geology* **25**, 955–956 (1997).
8. E. W. Bolton, R. A. Berner, S. T. Petsch, *Am. J. Sci.* **306**, 575–615 (2006).
9. M. A. Torres, A. J. West, G. Li, *Nature* **507**, 346–349 (2014).
10. S. J. Dadson et al., *Nature* **426**, 648–651 (2003).
11. N. Hovius, C. P. Stark, C. Hao-Tsu, L. Jiun-Chuan, *J. Geol.* **108**, 73–89 (2000).
12. R. Emberson, N. Hovius, A. Galy, O. Marc, *Earth Surf. Dyn.* **4**, 727–742 (2016).
13. K. Maher, C. P. Chamberlain, *Science* **343**, 1502–1504 (2014).
14. Materials and methods are available as supplementary materials.
15. C.-C. Tsai, Z.-S. Chen, C.-T. Duh, F.-W. Horng, *Proc. Natl. Sci. Counc. Repub. China B* **25**, 34–39 (2001).
16. O. Beyssac et al., *Tectonics* **26**, TC6001 (2007).
17. R. G. Hilton, A. Galy, N. Hovius, M.-J. Horng, H. Chen, *Geochim. Cosmochim. Acta* **74**, 3164–3181 (2010).
18. D. Calmels et al., *Earth Planet. Sci. Lett.* **303**, 48–58 (2011).
19. R. G. Hilton, A. Galy, N. Hovius, M.-J. Horng, H. Chen, *Geology* **39**, 71–74 (2011).

20. S. Schillawski, S. Petsch, *Global Biogeochem. Cycles* **22**, GB3002 (2008).
21. S. Chang, R. A. Berner, *Geochim. Cosmochim. Acta* **63**, 3301–3310 (1999).
22. S. T. Petsch, T. I. Eglinton, K. J. Edwards, *Science* **292**, 1127–1131 (2001).
23. J. D. Hemingway, D. H. Rothman, S. Z. Rosengard, V. V. Galy, *Biogeosciences* **14**, 5099–5114 (2017).
24. R. G. Keil, L. M. Mayer, in *Treatise on Geochemistry*, H. Holland, K. Turekian, Eds. (Elsevier, 2014), vol. 12, chap. 12.
25. A. Frostegård, E. Bååth, *Biol. Fertil. Soils* **22**, 59–65 (1996).
26. F. M. Hopkins *et al.*, *Soil Biol. Biochem.* **76**, 57–69 (2014).
27. N. Blair *et al.*, *Appl. Environ. Microbiol.* **50**, 996–1001 (1985).
28. N. Hovius *et al.*, *Earth Planet. Sci. Lett.* **304**, 347–355 (2011).
29. G. Li *et al.*, *Earth Planet. Sci. Lett.* **472**, 253–265 (2017).
30. O. Marc, N. Hovius, P. Meunier, T. Uchida, T. Gorum, *J. Geophys. Res. Earth Surf.* **121**, 640–663 (2016).
31. L.-W. Zheng *et al.*, *Earth Planet. Sci. Lett.* **465**, 103–111 (2017).
32. R. G. Hilton *et al.*, *Global Biogeochem. Cycles* **26**, GB3014 (2012).
33. I. J. Larsen *et al.*, *Science* **343**, 637–640 (2014).
34. A. J. West, *Geology* **40**, 811–814 (2012).

ACKNOWLEDGMENTS

We thank the NOSAMS (National Ocean Sciences Accelerator Mass Spectrometry) staff, especially M. Lardie-Gaylord and A. McNichol, and C. Johnson for laboratory assistance. J. Scheingross provided valuable comments on early versions of this manuscript. **Funding:** This research was supported by NSF Graduate Research Fellowship number 2012126152 and the WHOI Ocean Ventures Fund (J.D.H.); European Research Council Starting Grant 678779 ROC-CO2 (R.G.H.); and NSF grants OCE-0851015 and OCE-0928582 and WHOI Independent Study Award 27005306 (V.V.G.). **Author contributions:** J.D.H. and V.V.G. conceived the study; R.G.H., N.H., T.I.E., and M.-C.C. contributed

samples and analytical tools; J.D.H., N.H., and L.W. performed laboratory measurements; J.D.H., R.G.H., N.H., L.W., and V.V.G. analyzed data; and J.D.H., R.G.H., N.H., and V.V.G. wrote the manuscript with input from all authors. **Competing interests:** None declared. **Data and materials availability:** All data are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/209/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 to S8
References (35–65)

11 August 2017; accepted 21 February 2018
10.1126/science.aao6463

Photoperiodic control of seasonal growth is mediated by ABA acting on cell-cell communication

S. Tylewicz,^{1*} A. Petterle,¹ S. Marttila,² P. Miskolczi,¹ A. Azeez,^{1,3} R. K. Singh,¹ J. Immanen,⁴ N. Mähler,⁵ T. R. Hvidsten,^{5,6} D. M. Eklund,⁷ J. L. Bowman,⁸ Y. Helariutta,⁹ R. P. Bhalerao^{1†}

In temperate and boreal ecosystems, seasonal cycles of growth and dormancy allow perennial plants to adapt to winter conditions. We show, in hybrid aspen trees, that photoperiodic regulation of dormancy is mechanistically distinct from autumnal growth cessation. Dormancy sets in when symplastic intercellular communication through plasmodesmata is blocked by a process dependent on the phytohormone abscisic acid. The communication blockage prevents growth-promoting signals from accessing the meristem. Thus, precocious growth is disallowed during dormancy. The dormant period, which supports robust survival of the aspen tree in winter, is due to loss of access to growth-promoting signals.

Dormancy protects meristematic cells of perennial plants in temperate and boreal ecosystems by preventing growth during winter. Release from dormancy enables reinitiation of growth when favorable conditions return in spring (1). Shorter photoperiods as winter approaches (2) induce growth cessation, formation of a bud that encloses the arrested leaf primordia and shoot apical meristem (SAM) (Fig. 1A), and bud dormancy (3, 4). Longer photoperiods alone cannot promote growth in dormant buds; prolonged exposure to low temperatures is required to release dormancy (5, 6). We show that blockage of symplastic communication mediated by the action of abscisic acid (ABA) is part of the photoperiodically controlled dormancy mechanism in hybrid aspen.

Short photoperiods induce expression of ABA receptors and increase ABA levels in hybrid aspen buds (4, 7). ABA regulates dormancy (8). Therefore, we probed ABA's role in photoperiodic control of bud dormancy. First, we generated hybrid aspen plants with reduced ABA responses by expressing the dominant-negative *abi-1* allele

of *ABII*, a key ABA-signaling gene (9). Those hybrid aspens that expressed *abi-1* had reduced ABA responses, manifested by weak induction of the ABA-inducible gene *KIN2*, compared with that of wild-type (WT) controls (fig. S1). We then assessed bud dormancy by exposing WT and *abi-1* plants to 11 weeks of short photoperiod followed by transfer to long photoperiod without the low-temperature treatment required for dormancy release. Both WT and *abi-1* plants ceased growth and set buds after 4 weeks of short photoperiod (Fig. 1, A to C), but after 11 weeks of short photoperiod followed by long photoperiod, WT buds remained dormant, whereas *abi-1* buds reactivated growth within 11 to 15 days (Fig. 1, D to F). Thus, attenuation of ABA responses compromised photoperiodic control of bud dormancy and not growth cessation.

We investigated transcriptomic responses to short photoperiod in WT and *abi-1* apices in order to understand ABA-mediated control of dormancy. After 6 and 10 weeks of short photoperiod, respectively, we detected 9290 and 3053 differentially expressed genes in WT and 10,514 and 2149 differentially expressed genes in *abi-1* (line 1) apices (table S1). A large number of transcripts for plasmodesmata-associated proteins responded to short photoperiod. Plasmodesmata closure (by callosic dormancy sphincters) correlates with dormancy and their opening with dormancy release in diverse plants, including hybrid aspen and charophycean algae such as *Chara* (6, 10, 11). Of 187 poplar homologs of *Arabidopsis* genes encoding proteins enriched in plasmodesmata (12), 62 and 47 were induced after 6 and 10 weeks in WT apices, respectively, and of these, 53.2 and 76.6% were differentially expressed in *abi-1* relative to WT apices at these time points (table S2). Expression of *GERMIN-LIKE 10*; *REMORIN-LIKE 1* and 2, which are implicated in plasmodesmata function (13); and *CALLOSE SYNTHASE 1*, which is required for callose deposition (6), was progressively up-regulated, whereas

that of *GHI7-39*, a glucanase implicated in sphincter removal (6), was down-regulated in WT apices after 6 and 10 weeks of short photoperiod. These genes showed an altered response to short photoperiod in *abi-1* plants (fig. S2). Thus, ABA mediates short-photoperiod response of the plasmodesmata-related transcriptome.

Transcriptomic analysis prompted us to investigate ABA's role in plasmodesmata closure (Fig. 1, G to O). Under long photoperiod, WT and *abi-1* lines 1 and 3 had similar frequencies of "closed" plasmodesmata with dormancy sphincters (12.5 versus 17.4 and 13.5%, respectively). After 5 weeks of short photoperiod, corresponding frequencies were 78% in WT and 5.5 and 17.4% in *abi-1* apices, respectively, and after 10 weeks, frequencies increased to 83.6% in WT plants but fell to 2.2 and 0.5% in *abi-1* lines 1 and 3, respectively. Thus, ABA mediates plasmodesmata closure in response to short photoperiod. Plasmodesmata closure is not required for growth cessation (because growth cessation occurs in *abi-1* plants) and indicates association of plasmodesmata closure with bud dormancy, both being mediated by the same factor, ABA.

To investigate ABA-mediated plasmodesmata closure's role in short photoperiod-induced dormancy, we overexpressed PDLPI (PLASMODESMATA-LOCATED PROTEIN 1), which impairs trafficking via plasmodesmata (14) and phenocopying plasmodesmata blockage by dormancy sphincters, in *abi-1* plants (fig. S3). Both *abi-1*/*PDLPI* double transformants and parental *abi-1* plants ceased growth and formed buds under short photoperiod (Fig. 2, A to C), but subsequent exposure to long photoperiod only reactivated growth in the latter (Fig. 2, D to F). Thus, PDLPI expression suppressed *abi-1* plants' bud dormancy phenotype, although *KIN2* expression responses to ABA remained attenuated in *abi-1*/*PDLPI* (fig. S4). Thus, expression of PDLPI was sufficient to restore bud dormancy in *abi-1*/*PDLPI* plants without the restoration of general ABA responses.

PICKLE (*PKL*) is an antagonist of polycomb repression complex 2, which is implicated in seed dormancy (15, 16). *PKL* expression was down-regulated in WT plants but up-regulated in *abi-1* plants under short photoperiod (fig. S5). Hence, we investigated whether *PKL* could be involved in plasmodesmata closure and dormancy regulation mediated by ABA. Thus, we examined plasmodesmata in *abi-1* plants with suppressed *PKL* activity (*abi-1*/*PKLRNAi*) (RNAi, RNA interference) (fig. S6). Under long photoperiod, frequencies of plasmodesmata with dormancy sphincters were comparable in *abi-1* (13.1%) and *abi-1*/*PKLRNAi* lines 9 (19.4%) and 11 (18.4%) (Fig. 3, A to C). After 5 weeks of short photoperiod, the frequencies increased in the *abi-1*/*PKLRNAi* lines (to 34.4 and 28.5%, respectively), but not *abi-1* plants (16.4%) (Fig. 3, D to F). After 10 weeks of short photoperiod, the frequencies further increased in *abi-1*/*PKLRNAi* lines 9 and 11 to 84.6 and 74.5%, respectively, but fell in *abi-1* plants (5.2%) (Fig. 3, G to I). *PKL* down-regulation in *abi-1*/*PKLRNAi* also suppressed expression defects of plasmodesmata

¹Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, SE-901 87 Umeå, Sweden. ²Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Box 102, SE-230 53 Alnarp, Sweden. ³Plant Molecular Biology Laboratory, Jain R&D Laboratory, Agri Park, Jain Hills, Shirsol Road, Jalgaon, India. ⁴Department of Biosciences, Institute of Biotechnology, University of Helsinki, Viikinkaari 1, Post Office Box 65, Helsinki, Finland. ⁵Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, SE-901 87 Umeå, Sweden. ⁶Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway. ⁷Department of Plant Ecology and Evolution, Evolutionary Biology Centre, Uppsala University, SE-75236 Uppsala, Sweden. ⁸School of Biological Sciences, Monash University, Melbourne, VIC, Australia. ⁹Sainsbury Laboratory, Cambridge University, Bateman Street, Cambridge, UK.

*Present address: Department of Plant and Microbial Biology, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland. †Corresponding author. Email: risshi.bhalerao@slu.se

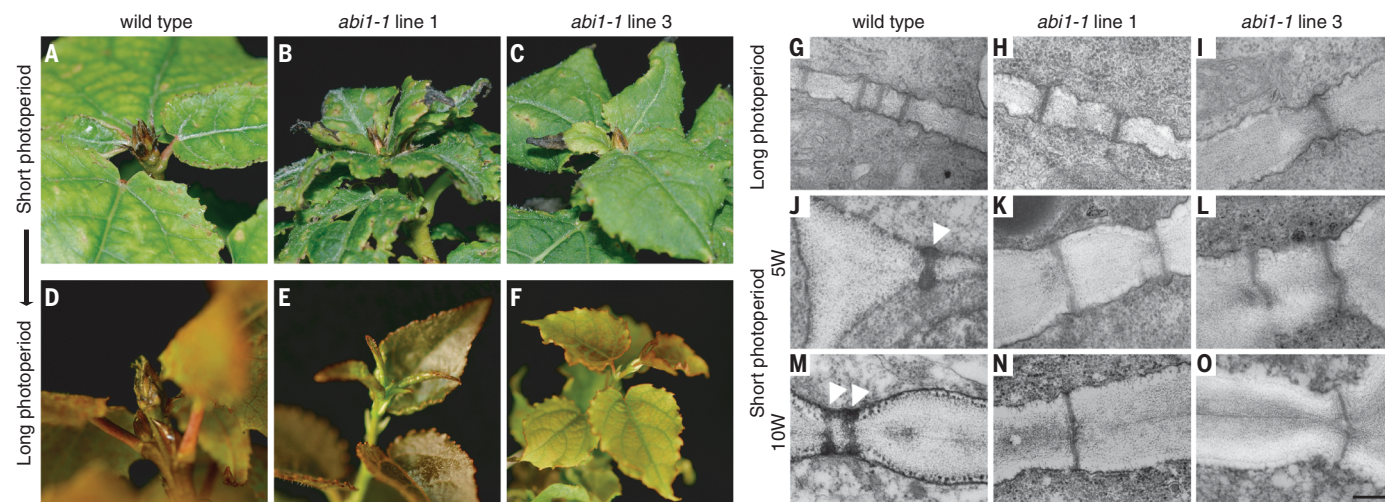


Fig. 1. Hybrid aspen plants with attenuated ABA responses fail to establish dormancy. (A to C) Buds of (A) wild type, (B) *abi-1* line 1, and (C) *abi-1* line 3 after 11 weeks of short photoperiod. (D to F) Unlike in (D) WT, buds burst in (E) *abi-1* line 1 and (F) *abi-1* line 3. (G to I) Transmission electron microscopy (TEM) micrographs of apices of actively growing

(G) WT plants, (H) *abi-1* line 1, and (I) *abi-1* line 3, showing plasmodesmata lacking electron-dense dormancy sphincters. (J to O) Sphincters are observed after 5 and 10 weeks of short photoperiod in apices of [(J) and (M)] wild-type plants (indicated with arrowheads), but not [(K) and (N)] *abi-1* line 1 or [(L) and (O)] *abi-1* line 3. Scale bar, 200 nm.

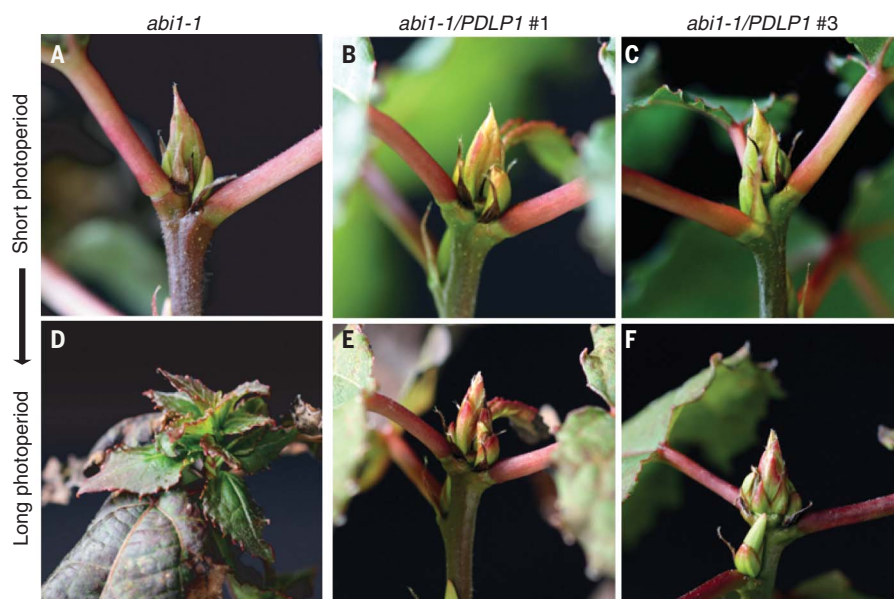


Fig. 2. *PDL1* expression restores bud dormancy in *abi-1* plants. (A to C) Buds of (A) *abi-1*, (B) *abi-1/PDL1* line 1, and (C) *abi-1/PDL1* line 3 after 11 weeks of short photoperiod. (D to F) Transfer to long photoperiod results in bud burst in (D) *abi-1* plants but not in (E) *abi-1/PDL1* line 1 or (F) *abi-1/PDL1* line 3.

markers evident in *abi-1* plants (fig. S7). Although both *abi-1* and *abi-1/PKLRNAi* plants ceased growth and set buds (Fig. 3, J to L), *abi-1/PKLRNAi* buds remained dormant and did not reactivate growth (unlike nondormant *abi-1* buds) after long photoperiod exposure after 11 weeks of short photoperiod (Fig. 3, M to O). Thus, *PKL* down-regulation restores plasmodesmata closure and bud dormancy defects in *abi-1* plants, suggesting that ABA mediates

plasmodesmata closure and bud dormancy by suppressing *PKL*.

Plasmodesmata closure could mediate dormancy by limiting access of SAM to growth-promotive signals. We investigated this hypothesis by analyzing responses of WT and *abi-1* buds to *FLOWERING LOCUS T1* (*FT1*), a seasonal growth regulator induced during dormancy release and before bud growth resumes (6, 17). We grafted scions of WT and *abi-1* plants exposed

to 10 weeks of short photoperiod (in order to induce plasmodesmata closure and dormancy) onto rootstocks of *FT1*-expressing plants (18). Although buds of WT scions did not reactivate growth, new leaves emerged from buds of *abi-1* scions under a continued short photoperiod (Fig. 4). Thus, plasmodesmata closure, as in WT plants, was associated with buds' failure to respond to *FT1* or *FT1*-derived growth-promotive signals, corroborating the involvement of plasmodesmata in photoperiodic control of ABA-mediated bud dormancy.

Thus, short photoperiods suppress *FT2*, which causes growth cessation and amplifies the ABA response by enhancing levels of ABA and ABA receptors (4, 7). ABA suppresses *PKL* and induces callose synthase to block plasmodesmata and maintains these blockages by repressing antagonistic glucanases (fig. S8). Hence, attenuating ABA responses not only results in a failure to induce plasmodesmata closure at dormancy onset but also in fewer subsequently closed plasmodesmata. Plasmodesmata closure through *PKL* down-regulation or *PDL1* expression, which both target cell-cell communication, suppresses dormancy defects in *abi-1* plants. *PDL1* expression restores dormancy without suppressing ABA response defects in *abi-1* plants. Thus, plasmodesmata closure is essential to dormancy and occurs downstream of ABA-mediated control of dormancy in response to shorter photoperiods.

With plasmodesmata closed, growth arrest is maintained even in the presence of growth-promoting signals. Reopening of closed plasmodesmata in dormant buds occurs slowly and only after prolonged exposure to low temperature. Hence, dormancy prevents precocious activation of growth. On the other hand, in the absence of dormancy and plasmodesmata closure, growth

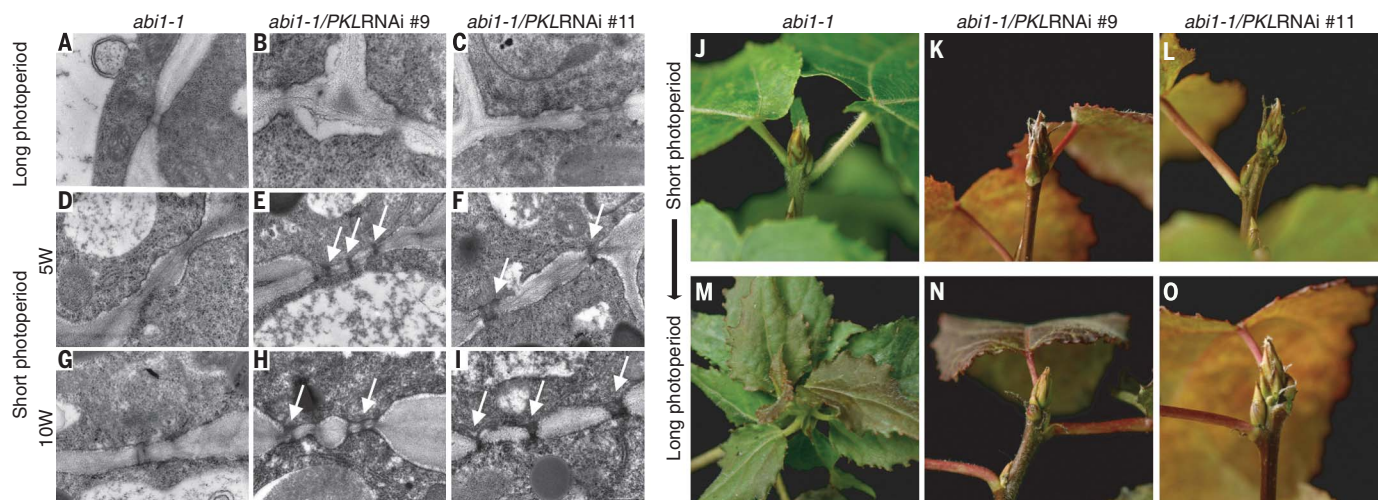


Fig. 3. PKL down-regulation restores dormancy sphincters and bud dormancy in *abi1-1* plants. (A to C) TEM micrographs of apices of actively growing (A) *abi1-1* plants, (B) *abi1-1/PKLRNAi* line 9, and (C) *abi1-1/PKLRNAi* line 11, showing plasmodesmata lacking electron-dense dormancy sphincters. (D to I) After 5 and 10 weeks of short photoperiod, sphincters were not observed in [(D) and (G)] *abi1-1* apices but were

present in *abi1-1/PKLRNAi* apices of [(E) and (H)] lines 9 and [(F) and (I)] 11 (arrows). Scale bar, 500 nm. (J to L) Buds of (J) *abi1-1* plants, (K) *abi1-1/PKLRNAi* line 9, and (L) *abi1-1/PKLRNAi* line 11 after 11 weeks of short photoperiod. (M to O) After a shift to long photoperiod, buds burst in (M) *abi1-1* plants but not in (N) *abi1-1/PKLRNAi* line 9 or (O) *abi1-1/PKLRNAi* line 11.



Fig. 4. FT1-expressing stocks can reactivate growth in *abi1-1* scions under short photoperiod. WT and *abi1-1* buds after 10 weeks of short photoperiod before grafting, and a further 2 and 7 weeks of short photoperiod after grafting of WT and *abi1-1* scions on FT1-expressing stocks. Buds remained dormant in WT scions but burst in *abi1-1* scions.

cessation induced by short photoperiod can be quickly reversed. Thus, dormancy, unlike growth cessation, adds robustness to the mechanism that is crucial for perennial survival and longevity in the face of changing seasons.

REFERENCES AND NOTES

1. R. K. Singh, T. Svystun, B. Aldahmash, A. M. Jönsson, R. P. Bhalerao, *New Phytol.* **213**, 511–524 (2017).
2. C. J. Weiser, *Science* **169**, 1269–1278 (1970).
3. J. E. Olsen et al., *Plant J.* **12**, 1339–1350 (1997).
4. T. Ruttink et al., *Plant Cell* **19**, 2370–2390 (2007).

5. A. Espinosa-Ruiz et al., *Plant J.* **38**, 603–615 (2004).
6. P. L. Rinne et al., *Plant Cell* **23**, 130–146 (2011).
7. A. Karlberg et al., *Plant Biotechnol.* **27**, 1–16 (2010).
8. S. Penfield, J. King, *Proc. Biol. Sci.* **276**, 3561–3569 (2009).
9. J. Leung et al., *Science* **264**, 1448–1452 (1994).
10. V. A. Shepherd, P. B. Goodwin, *Plant Cell Environ.* **15**, 137–150 (1992).
11. L. C. Jian, L. H. Sun, *Bot. Res.* **6**, 157–162 (1992).
12. L. Fernandez-Calvino et al., *PLOS ONE* **6**, e18880 (2011).
13. S. Raffaele, E. Bayer, S. Mongrand, *Plant Signal. Behav.* **4**, 915–919 (2009).
14. C. L. Thomas, E. M. Bayer, C. Ritzenhaller, L. Fernandez-Calvino, A. J. Maule, *PLOS Biol.* **6**, e7 (2008).
15. E. Aichinger et al., *PLOS Genet.* **5**, e1000605 (2009).
16. D. Bouyer et al., *PLOS Genet.* **7**, e1002014 (2011).
17. C. Y. Hsu et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10756–10761 (2011).
18. A. Azeez, P. Miskolczi, S. Tylewicz, R. P. Bhalerao, *Curr. Biol.* **24**, 717–724 (2014).

ACKNOWLEDGMENTS

Funding: Grants from Vetenskapsrådet (VR-2016-04430) and the Knut and Alice Wallenberg Foundation (2014-0032) to R.P.B. are gratefully acknowledged. **Author contributions:** S.T., A.P., S.M., P.M., A.A., R.K.S., and J.I. performed experiments. N.M. and T.R.H. performed the transcriptomics analysis. R.P.B., D.M.E., J.L.B., and Y.H. designed experiments and provided intellectual input. All authors contributed to writing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw RNA sequencing reads are available at the European Nucleotide Archive (www.ebi.ac.uk/ena) under accession no. PRJEB23073. All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/212/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S3
References (19–35)

30 May 2017; resubmitted 18 December 2017
Accepted 26 February 2018
Published online 8 March 2018
10.1126/science.aan8576

STRUCTURAL BIOLOGY

Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum

Katharina Braunger,^{1*} Stefan Pfeffer,^{2,*†} Shiteshu Shrima,³ Reid Gilmore,³ Otto Berninghausen,¹ Elisabet C. Mandon,³ Thomas Becker,¹ Friedrich Förster,^{4†} Roland Beckmann^{1†}

Protein synthesis, transport, and N-glycosylation are coupled at the mammalian endoplasmic reticulum by complex formation of a ribosome, the Sec61 protein-conducting channel, and oligosaccharyltransferase (OST). Here we used different cryo-electron microscopy approaches to determine structures of native and solubilized ribosome-Sec61-OST complexes. A molecular model for the catalytic OST subunit STT3A (staurosporine and temperature sensitive 3A) revealed how it is integrated into the OST and how STT3-paralog specificity for translocon-associated OST is achieved. The OST subunit DC2 was placed at the interface between Sec61 and STT3A, where it acts as a versatile module for recruitment of STT3A-containing OST to the ribosome-Sec61 complex. This detailed structural view on the molecular architecture of the cotranslational machinery for N-glycosylation provides the basis for a mechanistic understanding of glycoprotein biogenesis at the endoplasmic reticulum.

The mammalian translocon is responsible for cotranslational insertion of proteins into the endoplasmic reticulum (ER). The translocon is formed from the Sec61 complex, the oligosaccharyltransferase (OST) complex, and the translocon-associated protein (TRAP) complex (1). The Sec61 channel enables signal sequence-dependent protein translocation of soluble proteins through its central pore as well as integration of transmembrane proteins into the lipid bilayer through a lateral gate (2–5). OST catalyzes N-linked glycosylation of asparagine residues, an essential covalent protein modification (6–8). In higher eukaryotes, the catalytic OST subunit STT3 (staurosporine and temperature sensitive 3) is present in two paralogous forms (STT3A and STT3B) that assemble with a partially overlapping set of accessory subunits (Fig. 1A): RPN1 (ribophorin 1), RPN2 (ribophorin 2), OST48 (OST 48-kDa subunit), DAD1 (defender against cell death 1), TMEM258 (transmembrane protein 258), and OST4 (OST 4-kDa subunit) (9). STT3B complexes contain several specific subunits that include the paralogous oxidoreductases TUSC3 (tumor suppressor candidate 3) and MAGT1 (mag-

nesium transporter protein 1), whereas DC2 and KCP2 (keratinocyte-associated protein 2) are found only in STT3A complexes (10). The STT3A complex is thought to act cotranslationally and to be stably integrated into the translocon (10). The STT3B complex acts as a proofreader for sites missed by STT3A (11). Structures of monomeric bacterial and archaeal STT3 homologs have provided detailed insights into the catalytic mechanism (12–14). Genetic and biochemical data, as well as very recent high-resolution yeast OST structures (15, 16), indicate three subcomplexes of intimately interacting OST subunits. In the mammalian STT3A complex, these are RPN1 and TMEM258 (subcomplex I); STT3A, OST4, DC2, and KCP2 (subcomplex II); and RPN2, DAD1, and OST48 (subcomplex III) (7). The overall structure of mammalian OST in a native membrane environment has been established by cryo-electron tomography (cryo-ET) at medium resolution (1, 17–19); however, neither revealed structural details or the basis of STT3-paralog specificity.

To confirm STT3-paralog specificity in the ribosome translocon complex (RTC), we analyzed microsomes isolated from established Δ STT3A and Δ STT3B HEK cell lines, which do not express STT3A and STT3B, respectively (10), using cryo-ET. Immunoblots confirmed the absence of either STT3A or STT3B in the microsomal preparations of knockout cell lines, whereas both paralogs were present in microsomes prepared from control cells (Fig. 1B). Cryo-ET and in silico analysis of subtomograms showed that control microsomes harbored translocon populations that either included only TRAP (58%) or included both TRAP and OST (42%; Fig. 1C), as expected (17–19). The same

populations were found in a similar ratio in microsomes isolated from Δ STT3B cells (Fig. 1D), suggesting that translocon-associated OST was not affected by STT3B knockout. By contrast, no translocon-associated OST was observed after STT3A knockout (Fig. 1E), further indicating that RTCs harbor exclusively STT3A complexes (11). Instead of the TRAP-OST translocon complexes, a different, possibly partially assembled translocon population was observed after STT3A knockout.

We used single-particle cryo-electron microscopy (cryo-EM) to visualize solubilized mammalian RTCs translating the well-studied membrane glycoprotein bovine opsin (20) (figs. S1 and S2). Reconstructions yielded nonprogrammed and programmed RTCs showing an overall translocon architecture as observed in the native membrane (18, 19) except for TRAP, which appeared disordered or bound in substoichiometric amounts. Local resolution ranged from 3.5 to 4.5 Å for Sec61 and adjacent OST transmembrane helices (TMs) to 5 to 5.5 Å for more peripheral OST TMs (fig. S3). In the programmed, peptidyl-tRNA-containing complex, the nascent polypeptide density could be traced from the peptidyl (P)-site tRNA through the vestibule of the ribosomal tunnel projecting toward the cytoplasmic tip of Sec61 α TM10 (fig. S4). Sec61 was in a conformation very similar to the previously described “primed” state (Fig. 2, A and B, and fig. S5) (21), with a closed lateral gate (22, 23) and the plug helix (24) occluding the central pore.

Importantly, 28 additional TMs packed against Sec61 (Fig. 2, A and B), where OST is positioned in the native translocon (19). We generated a molecular model for mammalian STT3A, revealing high structural similarity to its fungal, archaeal, and bacterial homologs (fig. S6) as well as its orientation in context of the RTC (Fig. 2, C and D, and fig. S6). Clear density for the pyrophosphate group of the dolichol carrier was visible in the catalytic site (Fig. 2, B to D, and fig. S3D), suggesting that STT3A was in an active state. Glycosylation of the two consensus motifs in our substrate had already been completed (fig. S1C), and no peptide-substrate density was visible in the catalytic site. The TMs assigned to STT3A were surrounded by 15 additional TMs (Fig. 2, B and C, and fig. S7A). Of these TMs, 10 were located at the distal side of STT3A, facing away from Sec61. Three of them formed a bundle directly adjacent to STT3A TMs 1 and 2, whereas another bundle of seven TMs was in proximity to STT3A TMs 5 to 8. On the basis of the three established OST subcomplexes (7) and the number of TMs included in the bundles, we assigned the three-TM bundle to subcomplex I and the seven-TM bundle to subcomplex III. One TM of subcomplex I extended into the metazoan-specific cytoplasmic domain of RPN1, which formed a four-helix bundle according to secondary-structure predictions (fig. S7, B and C). It was intercalated between the OST TMs and the ribosome and contacted the linker between ribosomal RNA (rRNA) helix H19 and H20, rRNA expansion segment ES7a (H25), and the tail of ribosomal

¹Department of Biochemistry, Gene Center and Center for Integrated Protein Science Munich, University of Munich, 81377 Munich, Germany. ²Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. ³Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA. ⁴Cryo-Electron Microscopy, Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CH Utrecht, Netherlands.

*These authors contributed equally to this work.

†Corresponding author. Email: beckmann@genzentrum.lmu.de (R.B.); pfeffer@biochem.mpg.de (S.P.); f.g.forster@uu.nl (F.F.)

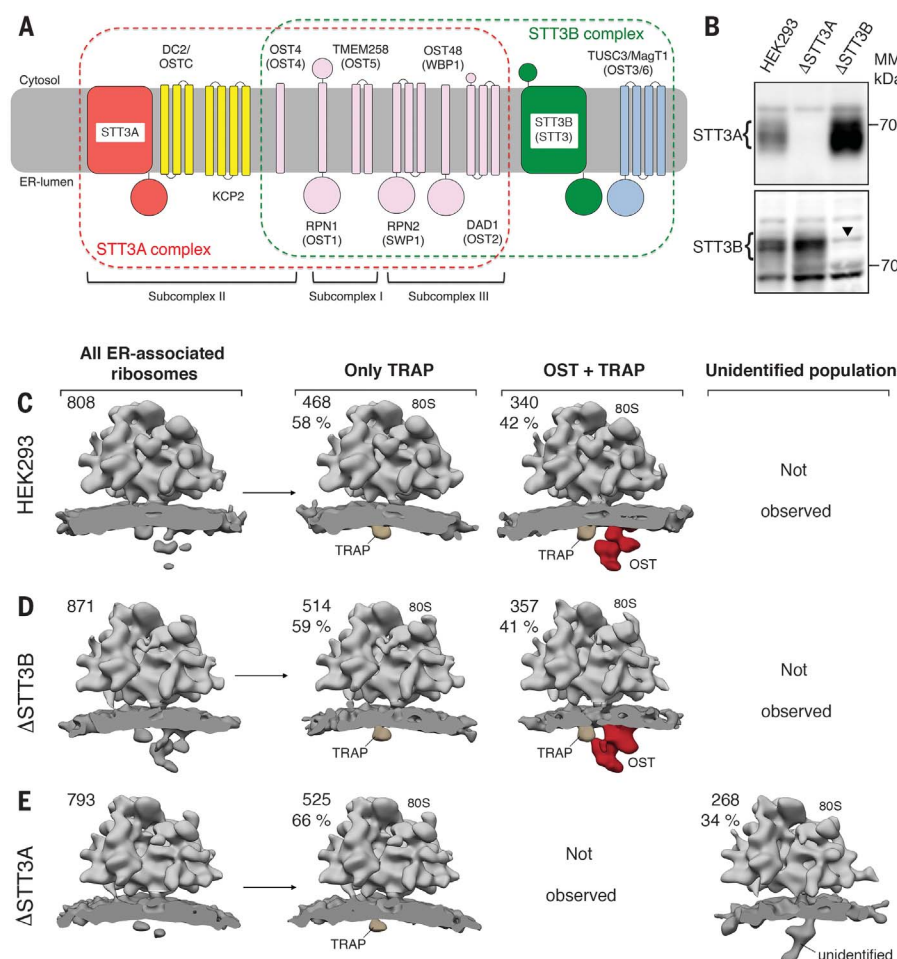


Fig. 1. RTCs harbor exclusively STT3A complexes. (A) Schematic representation and membrane topology of OST subunits for the STT3A (red frame) and STT3B complexes (green frame; yeast names in parentheses). Shared subunits are depicted in pink. OST subcomplexes are indicated for the STT3A complex. (B) Microsomes from wild-type or mutant human embryonic kidney (HEK) 293 cells were analyzed by immunoblotting using rabbit polyclonal antibodies. The arrowhead in the STT3B blot designates a nonspecific background band. MM, molecular mass. (C to E) Ribosome-bound translocon populations observed for microsomes from wild-type HEK293 (C), Δ STT3B (D), and Δ STT3A (E) cell lines after in silico sorting. The absolute number and percentage of subtomograms contributing to each class are given. All densities were filtered to 30-Å resolution. 80S, eukaryotic ribosome.

protein eL28 (Fig. 2, A and E). Antibodies against the cytosolic RPN1 segment inhibit protein translocation by interfering with ribosome binding to the translocon, confirming direct ribosome-RPN1 interaction (25). We further observed four extra TMs tightly associated with STT3A belonging to subcomplex II. One single TM, which we attributed to the single-spanning membrane protein OST4, was tightly intercalated between STT3A TMs 1, 3, 12, and 13 (fig. S7). The three remaining TMs located at the interface between STT3A TMs 10 to 13 and Sec61 were assigned to DC2. We built an atomic model for the three TMs of DC2 de novo based on excellent agreement between features resolved in our map and the predicted length and connectivity of DC2 TMs (Fig. 2F and fig. S3D). Recent biochemical data (10) show that

DC2 assumes a central role in recruiting OST into the translocon complex, and interactions of DC2 with both Sec61 and STT3A have been predicted. Indeed, DC2 contacted STT3A via its luminal C terminus (to STT3A TM13) (Fig. 2F), the cytosolic TM2-TM3 loop (to STT3A TM12-TM13 loop), and TM2 (close proximity to STT3A TM9-TM10 loop, also referred to as EL5). The amphipathic DC2 N terminus projected toward Sec61 on the micelle surface (Fig. 2F), whereas the luminal loop of DC2 interacted with the C termini of Sec61 β and Sec61 γ (Fig. 2F). We did not observe density for KCP2, likely because it tends to dissociate upon solubilization (10). We observed an additional weaker density for a TM segment intercalated between DC2 and Sec61 in the peptidyl-tRNA-containing map,

which was absent in the nonprogrammed map (Fig. 2, B and C, and fig. S2) and might correspond to the nascent opsin substrate or an as yet unknown translocon component.

We identified two interfaces integrating the STT3A complex into the RTC, one between the ribosome and the cytosolic RPN1 domain (Fig. 2, A and E) and one between DC2 and Sec61 (Fig. 2, B and F), both of which could explain STT3-paralog specificity. First, STT3B possesses a specific 47-amino acid soluble domain extending from STT3 TM1 into the cytosol directly beneath the cytosolic RPN1 helix bundle (Fig. 2, B and C). The STT3B-specific extension would thus be located in immediate proximity to the ribosome-OST interface, where it could interfere with ribosome binding. Second, STT3 TMs 10 to 13 and the cytosolic STT3 TM12-TM13 loop, identified as the major contact sites between DC2 and STT3A (Fig. 2, C and F), differed considerably between the STT3 paralogs (table S1 and fig. S8). This suggests that DC2 binds specifically to the STT3A paralog, which would exclude STT3B complexes from the RTC.

In our second (nonprogrammed) reconstruction (fig. S2), the general translocon architecture was very similar to the P-site tRNA-containing complex, and models for laterally closed Sec61 and OST fitted well as separate rigid bodies (Fig. 3A). Comparison with the model of the programmed RTC revealed a tilting movement between Sec61 and OST, with the cytosolic loops of Sec61 and the cytosolic RPN1 domain serving as hinge points on the ribosomal large subunit (Fig. 3B and movie S1). Furthermore, we improved image processing for an already published cryo-ET data set (18) of the native RTC with laterally opened Sec61 (fig. S9) to a resolution allowing rigid body fits of Sec61 and OST (Fig. 3A). Upon opening of the Sec61 lateral gate, the Sec61 α N-terminal domain and Sec61 β approached DC2. This induced a repositioning of the entire OST complex to accommodate the Sec61 conformational change (Fig. 3B and movie S1). Although the relative arrangement of DC2 and Sec61 differed substantially between the three observed conformational states, DC2 always stably interacted with Sec61. Thus, DC2 acts as a versatile module that provides robust integration of OST into the translocon complex, even under vastly differing conformational states of the translocon complex.

Our cryo-EM reconstructions define the exact position and orientation of the OST catalytic site in the context of the mammalian RTC and enable a detailed dissection of the interface between OST and the ribosome-Sec61 complex. This allowed us to interpolate the path for a nascent glycosylation substrate for cotranslational scanning on translocon-associated OST (Fig. 3C) and provided a molecular basis for STT3-paralog specificity in the RTC (Fig. 3D). The minimum distance between a TM segment at the Sec61 lateral gate and the catalytic site of STT3A was about 6.5 nm, explaining why glycosylation sites that are very close to TMs are often skipped by translocon-associated OST (26).

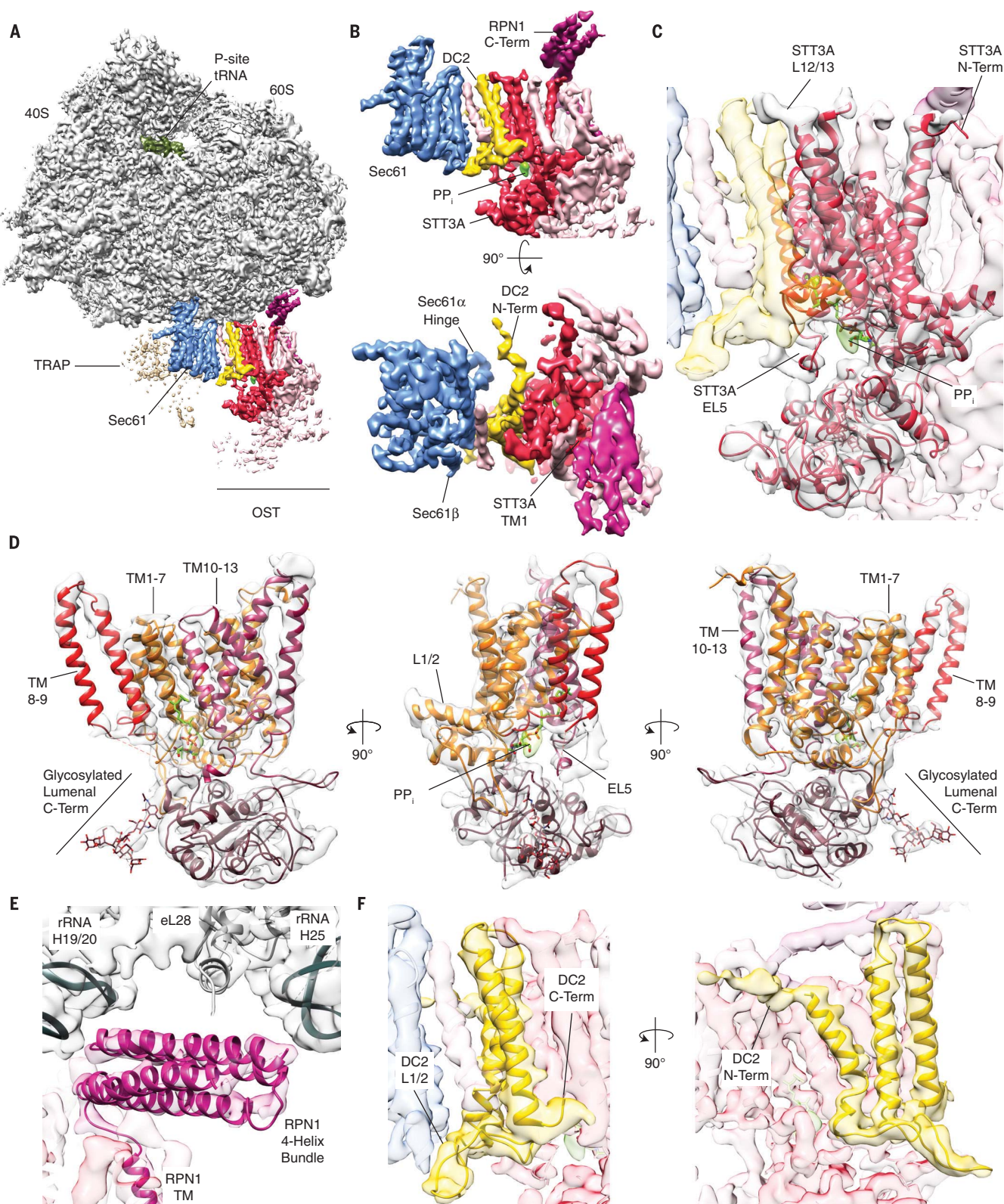


Fig. 2. Localization of STT3A, RPN1, and DC2 in ribosome-bound OST. (A) Cryo-EM structure of the active solubilized RTC. Ribosome and P-site tRNA are shown before focused refinement, low-pass filtered to 4 Å; the membrane region including Sec61, TRAP, and OST is shown after (fig. S2). 40S and 60S are eukaryotic ribosomal subunits. (B) Magnified view of the translocon region omitting TRAP,

as depicted in (A) (top) or rotated by 90° (bottom). PP_i, inorganic pyrophosphate. (C and D) Fitted homology model for mammalian STT3A. Density for phosphate groups in the catalytic center is green. L, loop. (E) Magnified view of the cytosolic RPN1 four-helix bundle binding to the ribosome. (F) Magnified view of the Sec61-OST interface with a fitted model for the DC2 TMs.

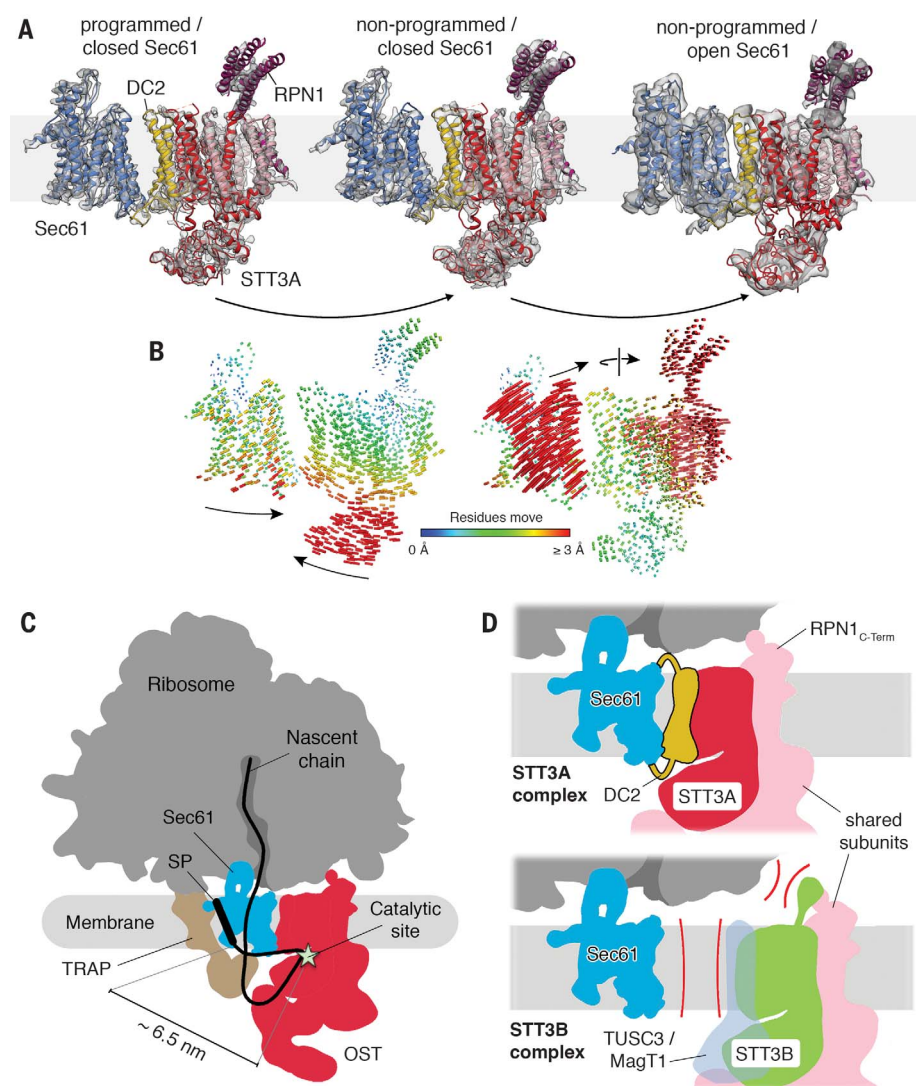


Fig. 3. Translocon dynamics and scheme for cotranslational N-glycosylation. (A) Models for Sec61 and OST were fitted into the RTC densities with laterally closed (left, programmed; center, nonprogrammed) and opened Sec61 (right). (B) Trajectories of α carbon atoms connecting the observed conformational states with color-coded length. (C) Schematic representation of the RTC with an interpolated example path for a nascent secretory protein. The STT3A catalytic site and a signal peptide (SP) or TM in the Sec61 lateral gate are separated by ~ 6.5 nm. (D) Molecular basis for STT3-paralog specificity in the RTC. The DC2 and RPN1 subunits tie the STT3A complex into the RTC (top). The lack of DC2 and potential interference of the STT3B-specific cytosolic domain with ribosome binding exclude STT3B complexes from the RTC (bottom).

REFERENCES AND NOTES

- S. Pfeffer, J. Dudek, R. Zimmermann, F. Förster, *Biochim. Biophys. Acta* **1860**, 2122–2129 (2016).
- R. M. Voorhees, R. S. Hegde, *Curr. Opin. Cell Biol.* **41**, 91–99 (2016).
- T. A. Rapoport, L. Li, E. Park, *Annu. Rev. Cell Dev. Biol.* **33**, 369–390 (2017).
- F. Cymer, G. von Heijne, S. H. White, *J. Mol. Biol.* **427**, 999–1022 (2015).
- R. Beckmann et al., *Science* **278**, 2123–2126 (1997).
- E. Mohorko, R. Glockshuber, M. Aebi, *J. Inher. Metab. Dis.* **34**, 869–878 (2011).
- D. J. Kelleher, R. Gilmore, *Glycobiology* **16**, 47R–62R (2006).
- M. Chavan, W. Lennarz, *Trends Biochem. Sci.* **31**, 17–20 (2006).
- D. J. Kelleher, D. Karaoglu, E. C. Mandon, R. Gilmore, *Mol. Cell* **12**, 101–111 (2003).
- S. Shrimai, N. A. Cherepanova, R. Gilmore, *J. Cell Biol.* **216**, 3625–3638 (2017).
- C. Ruiz-Canada, D. J. Kelleher, R. Gilmore, *Cell* **136**, 272–283 (2009).
- C. Lizak, S. Gerber, S. Numao, M. Aebi, K. P. Locher, *Nature* **474**, 350–355 (2011).
- S. Matsumoto et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17868–17873 (2013).
- M. Napiórkowska et al., *Nat. Struct. Mol. Biol.* **24**, 1100–1106 (2017).
- R. Wild et al., *Science* **359**, 545–550 (2018).
- L. Bai, T. Wang, G. Zhao, A. Kovach, H. Li, *Nature* (2018).
- J. Mahamid et al., *Science* **351**, 969–972 (2016).
- S. Pfeffer et al., *Nat. Commun.* **6**, 8403 (2015).
- S. Pfeffer et al., *Nat. Commun.* **5**, 3072 (2014).
- S. L. Meacock, F. J. Lecomte, S. G. Crawshaw, S. High, *Mol. Biol. Cell* **13**, 4114–4129 (2002).
- R. M. Voorhees, I. S. Fernández, S. H. Scheres, R. S. Hegde, *Cell* **157**, 1632–1643 (2014).
- R. M. Voorhees, R. S. Hegde, *Science* **351**, 88–91 (2016).
- M. Gogala et al., *Nature* **506**, 107–110 (2014).
- E. Park, T. A. Rapoport, *Nature* **473**, 239–242 (2011).
- Y. H. Yu, D. D. Sabatini, G. Kreibich, *J. Cell Biol.* **111**, 1335–1342 (1990).
- I. M. Nilsson, G. von Heijne, *J. Biol. Chem.* **268**, 5798–5801 (1993).

ACKNOWLEDGMENTS

We thank S. Rieder and M. Gogala for technical assistance, as well as C. Schmidt and J. Cheng for critical discussions.

Funding: This work was supported by funding from the German Research Council (SFB646 to R.B. and T.B.; GRK1721 to R.B. and F.F.) and the NIH under award number GM35678 (R.G.). R.B. acknowledges support from the Center for Integrated Protein Science Munich (CiPS-M) and the European Research Council (Advanced Grant CRYOTRANSLATION). F.F. received funding from the European Research Council under the European Union's Horizon2020 Programme [ERC Consolidator Grant Agreement 724425, BioGenesis and Degradation of Endoplasmic Reticulum proteins (BENDER)] and the Deutsche Forschungsgemeinschaft (FO 716/4-1). We also acknowledge the support of a Ph.D. fellowship from Boehringer Ingelheim Fonds (to K.B.). **Author contributions:** K.B., S.P., T.B., E.C.M., R.G., F.F., and R.B. designed the study and wrote the manuscript. S.S. purified and characterized human microsomal membranes. K.B. purified and biochemically analyzed stalled, solubilized RTCs. S.P. collected cryo-ET data, and O.B. collected cryo-EM data. S.P. processed and interpreted the cryo-ET data. K.B. processed the cryo-EM data and, together with T.B. and S.P., built the models and analyzed the structures. **Competing interests:** None declared. **Data and materials availability:** The cryo-EM density maps and corresponding atomic models reported in this paper have been deposited in the EM Data Bank and Protein Data Bank with the accession codes EMD-4306, EMD-4307, EMD-4308, EMD-4309, EMD-4310, EMD-4311, EMD-4312, EMD-4313, EMD-4314, EMD-4315, and PDB 6FTG for cryo-ET and EMD-4316, EMD 4317, and PDB 6FTI and 6FTJ for single-particle cryo-EM.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/215/suppl/DC1

Materials and Methods

Figs. S1 to S9

Table S1

References (27–39)

Movie S1

18 December 2017; accepted 26 February 2018

Published online 8 March 2018

10.1126/science.aar7899

STRUCTURAL BIOLOGY

Structure of the nuclear exosome captured on a maturing preribosome

Jan Michael Schuller,^{1*} Sebastian Falk,^{1*} Lisa Fromm,² Ed Hurt,^{2†} Elena Conti^{1†}

The RNA exosome complex processes and degrades a wide range of transcripts, including ribosomal RNAs (rRNAs). We used cryo-electron microscopy to visualize the yeast nuclear exosome holocomplex captured on a precursor large ribosomal subunit (pre-60S) during 7S-to-5.8S rRNA processing. The cofactors of the nuclear exosome are sandwiched between the ribonuclease core complex (Exo-10) and the remodeled “foot” structure of the pre-60S particle, which harbors the 5.8S rRNA precursor. The exosome-associated helicase Mtr4 recognizes the preribosomal substrate by docking to specific sites on the 25S rRNA, captures the 3' extension of the 5.8S rRNA, and channels it toward Exo-10. The structure elucidates how the exosome forms a structural and functional unit together with its massive pre-60S substrate to process rRNA during ribosome maturation.

The eukaryotic RNA exosome is a conserved 3'-5' degradation machinery that functions in the turnover, surveillance, and processing of coding and noncoding RNAs, in both the nucleus and the cytoplasm (1, 2). The processing of ribosomal RNA (rRNA) precursors is a prominent function of the nuclear exosome (3). In yeast, ribosome biogenesis starts with the synthesis of a polycistronic transcript, from which the 18S, 5.8S, and 25S rRNAs are generated by a series of processing reactions (4, 5). One of the most complex steps in rRNA biogenesis is the degradation of the internal transcribed spacer 2 (ITS2), an intervening sequence located between the 5.8S and 25S rRNAs that is almost completely removed before the pre-60S ribosomal particle is exported to the cytoplasm (4) (fig. S1). ITS2 removal requires the action of the exosome and is indeed the pathway that led to the discovery of this complex in *Saccharomyces cerevisiae* (6).

The yeast exosome contains a core complex of 10 proteins (Exo-10), which include a single processive 3'-5' exoribonuclease (Rrp44) and nine catalytically inactive subunits (Exo-9) (1, 2, 7). RNA substrates reach the ribonuclease via an internal channel that traverses the entire core complex and can accommodate up to 30 nucleotides (8, 9). In the nucleus, Exo-10 functions with four conserved cofactors: the distributive 3'-5' exoribonuclease Rrp6, its binding partner Rrp47, the small protein Mpp6, and the 3'-5' RNA helicase Mtr4 (1, 3). Rrp6-Rrp47 and Mpp6 stably associate with the exosome core and together contribute to transiently recruit Mtr4 (10–13). In turn, Mtr4 is transiently recruited by ribosome biogenesis factors to catalyze the removal of rRNA spacer sequences (14).

The removal of ITS2 from the pre-60S particle starts with cleavage reactions that generate

the 5' end of the mature 25S rRNA but leave behind a 5.8S rRNA precursor with a long 3' end extension (7S) (fig. S1) (4, 5). Subsequent trimming of the 7S pre-rRNA by the exosome occurs through the sequential action of the two nuclear exosome ribonucleases (15, 16). Rrp44 first shortens the 3' end of the 7S pre-rRNA to a 5.8S rRNA form extended by 30 nucleotides (5.8S+30); Rrp6 then takes over this intermediate and shortens the extension further (fig. S1) (15, 16). Similar pre-rRNA intermediates have been observed in mammalian cells, suggesting that the mechanism of exosome-mediated 7S-to-5.8S rRNA processing is conserved from yeast to human (17).

The individual steps in ribosomal biogenesis not only entail the progressive shortening of rRNA precursors but also correlate with discrete preribosomal particles that differ in the composition of ribosomal proteins and transiently associated biogenesis factors (4). Recent cryo-electron

microscopy (cryo-EM) reconstructions have revealed the architecture of pre-60S particles containing the 7S pre-rRNA, showing how ribosomal biogenesis factors assemble around part of ITS2 and form the so-called “foot” structure of the particle (18). The finding that one of these biogenesis factors, Nop53, recruits the Mtr4 helicase (14) has paved the way for visualizing the structure of a nuclear exosome as it processes the 5.8S rRNA in a pre-60S ribosome particle.

We recently reconstituted the yeast 7S pre-rRNA processing reaction in vitro using endogenous 7S-containing pre-60S particles (purified by tagging Nop53) together with an active recombinant nuclear exosome holo-complex (Exo-10–Rrp6–Rrp47–Mpp6–Mtr4, referred to as Exo-14n) (19). For the structural analysis, we stalled the exosome on the pre-60S using an Exo-14n complex with a catalytically inactive Rrp6 (12), which accumulates unprocessed 5.8S+30 pre-rRNA (19) (fig. S1). Single-particle cryo-EM analysis of the purified pre-60S–Exo-14n complex yielded EM density maps ranging between 3.9- and 4.6-Å resolution (figs. S2 to S5 and table S1), of sufficient quality to unambiguously fit all the known atomic models (see materials and methods) (fig. S6). The resulting pseudo-atomic model reveals the architecture of the entire pre-60S–Exo-14n assembly intermediate, stalled on a 5.8S+30 pre-rRNA (5.8S+30 particle) (Fig. 1).

The inner core of the in vitro-processed pre-60S particle has a very similar overall structure as compared with the 7S pre-rRNA containing pre-60S particles (7S particles) previously isolated from yeast via either Nog2 (18) or Arx1 (20). However, there are pronounced differences. First, the L1 stalk, a flexible structural element formed within domain V of the 25S rRNA, has swiveled about 30° into a half-inward conformation, with its tip contacting the immature unrotated 5S ribonucleoprotein (RNP) (fig. S7). Second, the foot

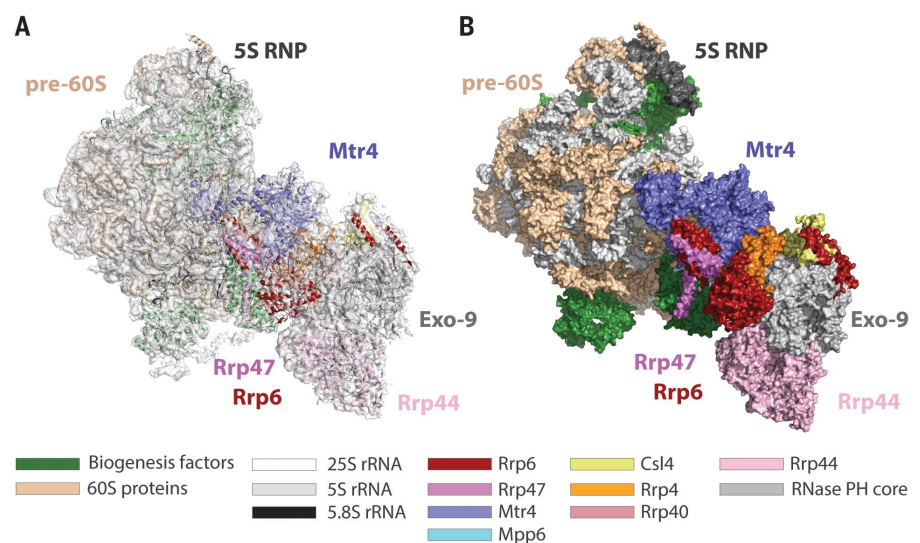


Fig. 1. Overall structure of the yeast pre-60S–Exo-14n complex. (A) Cryo-EM density and (B) surface representation of the pre-60S–Exo-14n structure fitted with known atomic structures. The color-coding scheme for the different proteins and RNAs is indicated at the bottom. The 5.8S rRNA is embedded within the complex and not visible in the surface representation.

¹Department of Structural Cell Biology, Max Planck Institute (MPI) for Biochemistry, Munich, Germany. ²Biochemistry Centre, University of Heidelberg, Heidelberg, Germany.

*These authors contributed equally to this work.

†Corresponding author. Email: conti@biochem.mpg.de (E.C.); ed.hurt@bzh.uni-heidelberg.de (E.H.)

structure at the bottom of the pre-60S particle has been almost completely remodeled. In the earlier 7S particles, the foot is formed by five ribosome biogenesis factors, which coat the structured part at the 3' extension of the 5.8S rRNA (18) (fig. S8, A and B). In the 5.8S+30 particle, only one of these assembly factors (Nop7) has remained bound in the same conformation (Fig. 1 and fig. S8, A and B). No ordered density is visible for Nop53, which had been used as the bait for pre-60S purification, suggesting that it may be flexibly attached after remodeling or dissociated during the EM sample preparation (fig. S8C). Furthermore, the convoluted structure at the 3' extension of the 5.8S rRNA has been unfolded and trimmed and is now embedded in a single-stranded conformation within the exosome channel (see below). The physical space previously occupied by the ITS2 RNP in the 7S particle is now occupied by the bulky Mtr4 and the other exosome cofactors (Fig. 2 and fig. S8B).

The Mtr4 helicase provides the main connection between the pre-60S and Exo-10. Mtr4 contains a catalytic core (a DExH-type helicase

region formed by two RecA and a helical bundle domain) and a regulatory “arch” (21, 22) [formed by a helical “stalk” and a KOW (Kyrpides, Ouzounis, and Woese) domain]. In our cryo-EM structure, Mtr4 binds the 25S rRNA via a bidentate interaction mediated both by the arch and by the DExH core (Fig. 2A and fig. S6B). Within the arch, the KOW domain contacts domain I of the 25S rRNA (at helices 15 and 16) (Fig. 2A) using structural elements that had been previously shown to bind double-stranded RNA in nuclear magnetic resonance mapping experiments (23). In this orientation, the Nop53-binding site on the KOW domain is solvent accessible (14, 23) (fig. S8D), suggesting that the arch can in principle bind both Nop53 and the 25S rRNA during the early stages of recruitment to the 7S particle (23). In general, these KOW-rRNA interactions, which we observe in our map, rationalize previous functional data that the arch of Mtr4 is required for rRNA processing *in vivo* (22).

The DExH core of Mtr4 contacts domain V of the 25S rRNA (Fig. 2A). The helical bundle domain approaches a eukaryotic-specific element

of the 25S rRNA (helix 79 in expansion segment ES31), whereas the RecA2 domain contacts an adjacent surface at the base of the L1 stalk (helix 76, near ribosomal protein L8). Altogether, these interactions push the L1 stalk upward, possibly causing long-range effects to the tip of the L1 stalk and stabilizing it in its half-inward conformation. Importantly, some of the contacts between Mtr4 and domain V of the 25S rRNA would only be feasible after the foot structure has been remodeled and the biogenesis factor Rlp7 been removed. It is thus possible to envisage how Mtr4 could signal the state of ITS2 processing to the L1 stalk, which in turn could trigger the next ribosome biogenesis steps (e.g., the recruitment of Rix1-Rea1 and rotation of the 5S RNP) (24).

The DExH-binding and KOW-binding regions in the 25S rRNA are separated by about 90 Å (Fig. 2A). To span this distance, the arch domain of Mtr4 moves from the bent conformation captured in previous crystal structures (21, 22) to a more extended state. Interestingly, a similar conformational change has been observed with

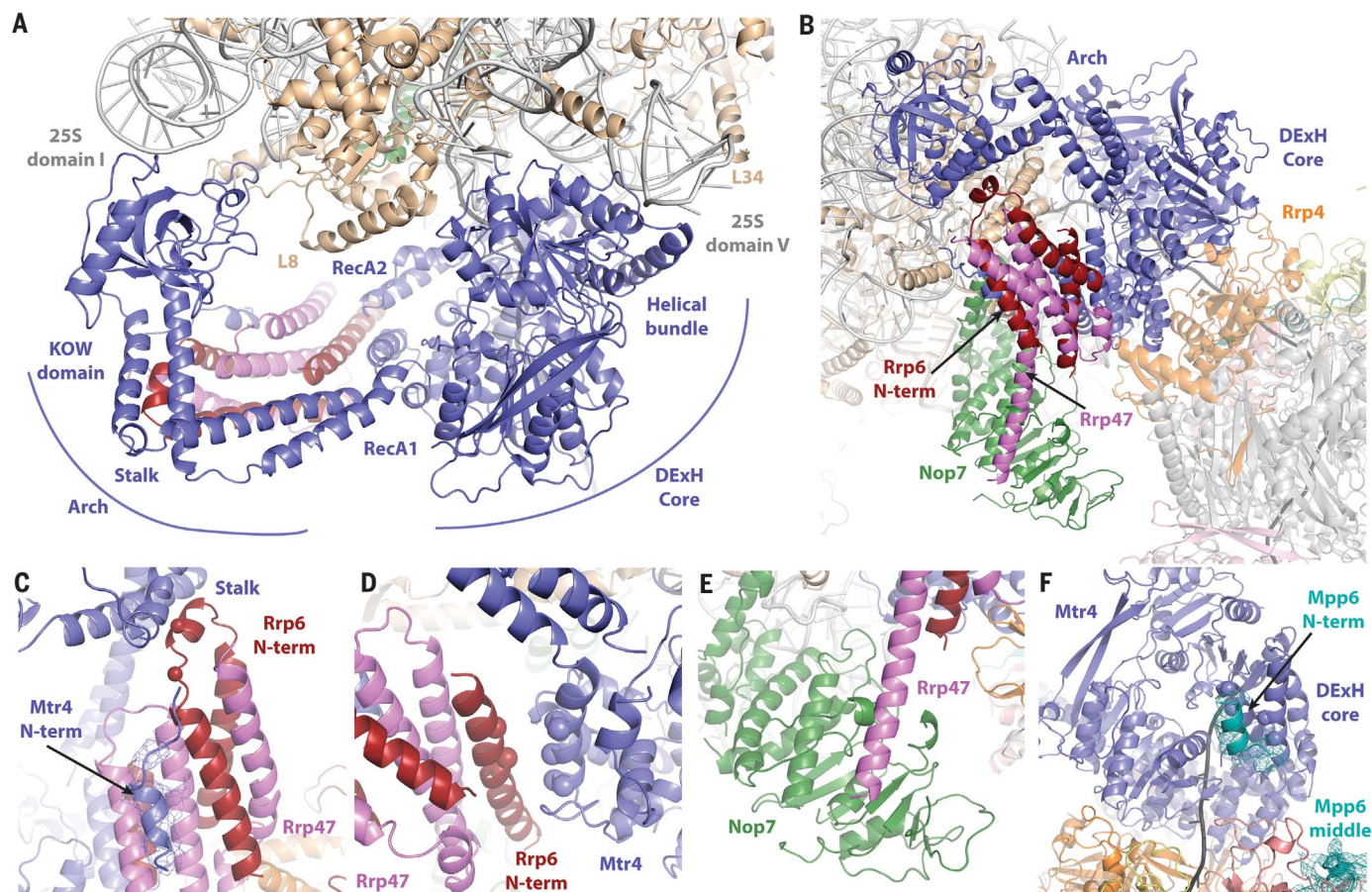
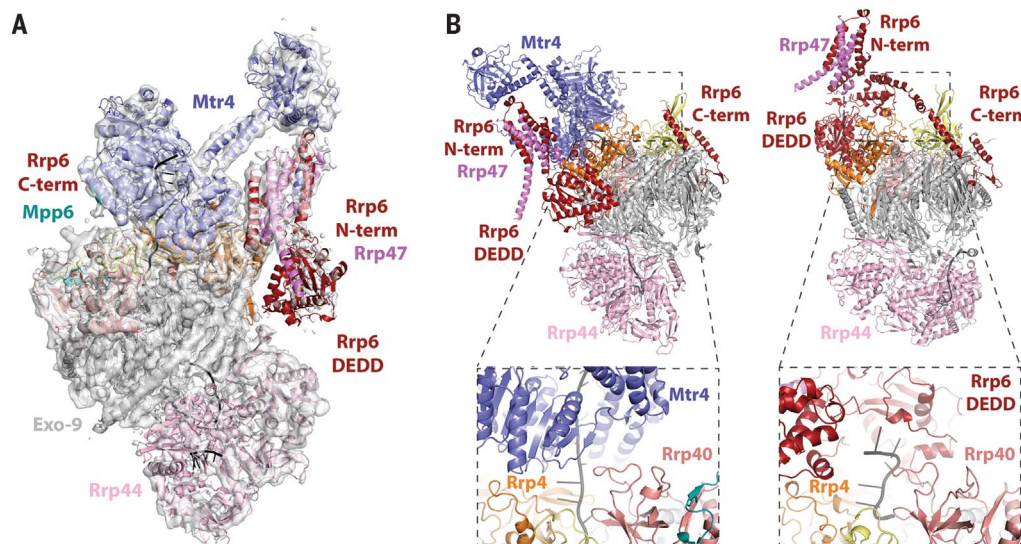


Fig. 2. The nuclear cofactors of the RNA exosome. (A) Mtr4 (blue) with the pre-60S particle (25S rRNA gray, ribosomal proteins wheat). (B) Rrp6_N-Rrp47_N (red and pink) with Mtr4 and the biogenesis factor Nop7 (green). (C) Rrp6_N-Rrp47_N concave surface with the N-terminal region of Mtr4 (30), shown with the corresponding cryo-EM density. Red spheres represent the position of residues mutated in a previous study [Rrp6 Asp²⁷ and Phe³⁰ (30)].

(D) Rrp6_N-Rrp47_N convex surface with the DExH core of Mtr4. Spheres identify positions of conserved negatively charged residues of Rrp6 and conserved positively charged residues of Mtr4. (E) C-terminal helix of Rrp47_N with Nop7. (F) Bottom surface of the Mtr4 DExH core with additional cryo-EM density (attributed to N terminus of Mpp6, cyan). Blue spheres represent the position of residues mutated in Mtr4 that abolish binding to Mpp6.

Fig. 3. Open and closed conformations of nuclear exosome complexes.

(A) View of the Exo-14n structure from the complex with the pre-60S particle (with cryo-EM density) showing the edge-on position of Mtr4 on top of Exo-9. **(B)** (Left) Exo-14n complex rotated $\sim 180^\circ$ around a vertical axis with respect to (A) showing Rrp6 in an open conformation. (Right) Exo-12n crystal structure (16) in the same orientation, showing Rrp6 in a closed conformation. The zoom-in views at the bottom show how Mtr4 and Rrp6 dock on the same surface of the exosome subunit Rrp4 (orange).



the homologous cytoplasmic helicase Ski2 upon binding to the 80S ribosome (25). In our cryo-EM structure, the extended conformation of the Mtr4 arch appears to be stabilized by the Rrp6_N-Rrp47_N module, a tightly intertwined heterodimer formed by the N-terminal domains of the two proteins (26) (Fig. 2B). Fitting the characteristic crescent-shaped structure of Rrp6_N-Rrp47_N was unambiguous in the EM density (fig. S6C). Confirming the interpretation, the EM reconstruction showed additional density on the concave surface of the Rrp6_N-Rrp47_N crescent (Fig. 2C), consistent with the binding of the Mtr4 N-terminal region (26). On the convex surface of Rrp6_N-Rrp47_N, a helix of Rrp6 lined by negatively charged residues approaches the helical bundle domain of Mtr4 at a conserved positively charged surface (Fig. 2D and fig. S6D). At the tip of Rrp6_N-Rrp47_N, a conserved loop of Rrp6 reaches the stalk helices of the Mtr4 arch (Fig. 2C and fig. S2C). This observation rationalizes previous *in vivo* data that mutations of conserved residues in this loop result in a 5.8S rRNA processing defect in yeast (26). Finally, a characteristic feature of Rrp6_N-Rrp47_N is the presence of a long α helix in Rrp47_N (26). This helix protrudes by more than 20 Å from the crescent and attaches to the pre-60S particle by binding to the only remaining biogenesis factor at the remnant foot structure, Nop7 (Fig. 2E and fig. S6E).

Besides visualizing how Mtr4 docks to the pre-60S particle, the EM reconstruction reveals how it binds to the exosome complex. The Exo-9 core is formed by an upper ring of three “cap” proteins (Rrp4, Rrp40, and Csl4) and a lower ring of six ribonuclease (RNase) PH-like proteins (1, 7). The helicase core of Mtr4 is positioned edge-on at the top of Exo-9 (Fig. 3A). The helical bundle and RecA1 domains of Mtr4 bind the cap protein Rrp4 (Fig. 3B, left panel) and in particular approach a conserved loop of Rrp4 that is known to be essential *in vivo* (27). Previous structural studies have shown that the same surface of Rrp4 binds the ribonuclease domain of Rrp6 (12, 13, 28)

(Fig. 3B, right panel). This “closed” conformation of Rrp6 appears to be the resting position in exosome complexes lacking Mtr4 (12). In the pre-60S-Exo-14n complex, Mtr4 binding onto Rrp4 appears to have displaced the Rrp6 ribonuclease module into an “open” conformation, on the side of Exo-9. In this open conformation, the Rrp6 ribonuclease module is rather flexible, with only a portion [the helicase and RNaseD C-terminal (HRDC) domain] accounted for in the EM density (Fig. 3A), reminiscent of (although not identical to) another flexible open conformation observed in previous crystallographic studies (12). Essentially, the only part of Rrp6 that remains unchanged in all structures determined to date is the C-terminal exosome-binding domain (Fig. 3B and fig. S6F) (8, 12, 13, 28).

The incorporation of yeast Mtr4 into Exo-14n also requires Mpp6 (10, 26). Structural studies have shown how the middle domain of Mpp6 binds the cap protein Rrp40 at the top of Exo-9 (10, 11). From biochemical studies, the N-terminal domain of Mpp6 is expected to contribute to binding Mtr4 (10) and channeling RNA through it (11), but the mechanisms have remained unclear. In the reconstruction, we noticed a density feature on the helicase core of Mtr4 that would be unexplained by the fitting of available crystal structures and that would correspond dimensionally to a helix (Fig. 2F and fig. S6G). This structural feature docks onto conserved hydrophobic residues at the bottom of the DExH core (Ile⁴⁴³ and Ile⁴⁸⁹) and points toward the middle domain of Mpp6 (Fig. 2F and fig. S9). We reasoned that this density might correspond to the conserved N-terminal domain of Mpp6. Indeed, isothermal titration calorimetry experiments showed that Mpp6 residues 1 to 67 bind the DExH core of Mtr4 with a dissociation constant K_d of $\sim 25 \mu\text{M}$ (fig. S9A). The interaction was impaired when using the I443R/N446R or I489R/E493R mutants of Mtr4 or when deleting the conserved N-terminal segment of Mpp6 (residues 1 to 26) in pull-down experiments (fig. S9C and S9D).

After fitting the Exo-14n proteins, we identified and traced the 3' extension of the 5.8S rRNA in a prominent density that emerges from the pre-60S particle and extends into Exo-14n (Fig. 4). After the final nucleotide of the 5.8S rRNA (nucleotide 158), the ribonucleotide chain continues and enters into the DExH core of Mtr4. Here, the density follows the same path that had been observed in the crystal structure of RNA-bound Mtr4 (21) (Fig. 4). Upon exiting the helicase, the density weakens as it crosses the solvent region between the edge-on base of Mtr4 and the top of Exo-9. Well-defined density starts again as the RNA reaches the cap proteins and enters the internal channel of the exosome core. RNA traverses Exo-9, as previously observed in the crystal structure of Exo-10-Rrp6_C (8). The major difference is that the RNA chain ends in the PIN domain of Rrp44 rather than in the exoribonuclease domain. Such a path from Exo-9 to the PIN domain had already been suggested in previous studies (29, 30). In the context of our reconstruction, the most plausible interpretation is that we captured a situation/state whereby the 3' extension of 5.8S has been trimmed to ~ 30 nucleotides but cannot be handed over to Rrp6 for further trimming (because Rrp6 exoribonuclease is inactivated) and hence is re-captured in the exosome channel. Considering that Exo-14n has a footprint of 40 nucleotides in RNase protection assays (11), the path toward the PIN domain might simply reflect the best fit for a 30-nucleotide extension in a “resting” state of Exo-14n, when Mtr4 is in an edge-on position on top of Exo-9. The Mtr4-channeling conformation of the nuclear exosome that we observed in our reconstruction is likely to be relevant not only for the pre-60S substrate. Indeed, RNase protection assays of Exo-14n bound to a generic single-stranded RNA recapitulate the predictions from the cryo-EM reconstruction, namely that the arch domain of Mtr4 is required for RNA channeling (fig. S10).

This study shows how the nuclear RNA exosome remodels the pre-60S particle, both in

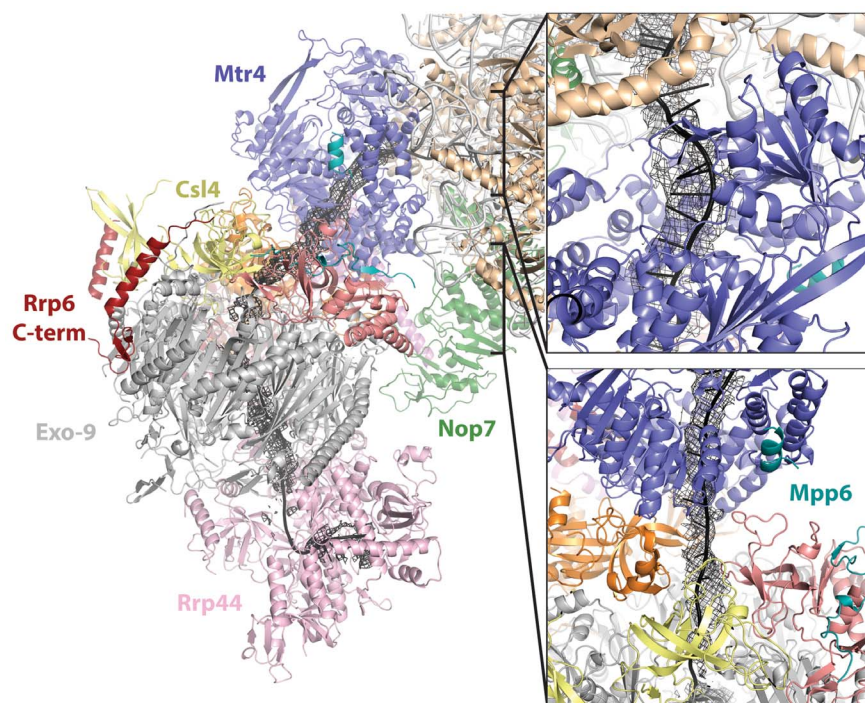


Fig. 4. The path of the 3' extension of the 5.8S rRNA from the pre-60S particle to the exosome. (Left) Close-up view of the cryo-EM structure, with the density corresponding to the 3' extension of the 5.8S rRNA as a black mesh. (Right) Zoom-in panels showing the RNA as it exits the pre-60S particle and enters Mtr4 (top) and as it exits Mtr4 and enters Exo-9 (bottom).

composition and structure, thereby signaling the status of ITS2 processing to the ribosome core. The exosome complex itself is remodeled upon binding to the preribosome: Its cofactors undergo dramatic conformational changes as they rearrange to channel the 3' extension of the 5.8S rRNA through the Mtr4 helicase and into the degradative chamber. Trapping the exosome in action on a pre-60S particle has given an unprecedented snapshot of how this RNA shredding machine works on such a complex substrate. Although the macromolecular complexes that degrade RNAs and synthesize proteins have so far been studied individually, this work sets the stage to elucidating how different machineries in eukaryotic gene expression are physically coupled and coordinated.

REFERENCES AND NOTES

- J. C. Zinder, C. D. Lima, *Genes Dev.* **31**, 88–100 (2017).
- A. Chlebowski, M. Lubas, T. H. Jensen, A. Dziembowski, *Biochim. Biophys. Acta* **1829**, 552–560 (2013).
- K. Kilchert, S. Wittmann, L. Vasiljeva, *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).
- J. L. Woolford Jr., S. J. Baserga, *Genetics* **195**, 643–681 (2013).
- E. Thomson, S. Ferreira-Cerca, E. Hurt, *J. Cell Sci.* **126**, 4815–4821 (2013).
- P. Mitchell, E. Petfalski, A. Shevchenko, M. Mann, D. Tollervey, *Cell* **91**, 457–466 (1997).
- D. L. Makino, F. Halbach, E. Conti, *Nat. Rev. Mol. Cell Biol.* **14**, 654–660 (2013).
- D. L. Makino, M. Baumgärtner, E. Conti, *Nature* **495**, 70–75 (2013).
- F. Bonneau, J. Basquin, J. Ebert, E. Lorentzen, E. Conti, *Cell* **139**, 547–559 (2009).
- E. V. Wasmuth, J. C. Zinder, D. Zattas, M. Das, C. D. Lima, *eLife* **6**, e29062 (2017).
- S. Falk, F. Bonneau, J. Ebert, A. Kögel, E. Conti, *Cell Reports* **20**, 2279–2286 (2017).
- D. L. Makino et al., *Nature* **524**, 54–58 (2015).
- E. V. Wasmuth, K. Janusz, C. D. Lima, *Nature* **511**, 435–439 (2014).
- M. Thoms et al., *Cell* **162**, 1029–1038 (2015).
- C. Allmang et al., *EMBO J.* **18**, 5399–5410 (1999).
- M. W. Briggs, K. T. D. Burkard, J. S. Butler, *J. Biol. Chem.* **273**, 13255–13263 (1998).
- L. Tafforeau et al., *Mol. Cell* **51**, 539–551 (2013).
- S. Wu et al., *Nature* **534**, 133–137 (2016).
- L. Fromm et al., *Nat. Commun.* **8**, 1787 (2017).
- B. Bradatsch et al., *Nat. Struct. Mol. Biol.* **19**, 1234–1241 (2012).
- J. R. Weir, F. Bonneau, J. Hentschel, E. Conti, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12139–12144 (2010).
- R. N. Jackson et al., *EMBO J.* **29**, 2205–2216 (2010).
- S. Falk et al., *RNA* **23**, 1780–1787 (2017).
- B. Tutuncuoglu, J. Jakovljevic, S. Wu, N. Gao, J. L. Woolford Jr., *RNA* **22**, 1386–1399 (2016).
- C. Schmidt et al., *Science* **354**, 1431–1433 (2016).
- B. Schuch et al., *EMBO J.* **33**, 2829–2846 (2014).
- H. Malet et al., *EMBO Rep.* **11**, 936–942 (2010).
- J. C. Zinder, E. V. Wasmuth, C. D. Lima, *Mol. Cell* **64**, 734–745 (2016).
- K. Drażkowska et al., *Nucleic Acids Res.* **41**, 3845–3858 (2013).
- E. V. Wasmuth, C. D. Lima, *Mol. Cell* **48**, 133–144 (2012).

ACKNOWLEDGMENTS

We would like to thank I. Schäfer, R. Prabu, and F. Beck for discussions on EM; P. Reichelt for help with yeast growths; D. Fleming for initial negative-stain images; F. Bonneau for the assay in fig. S10; E. Stegmann for technical support in protein purification; P. Cramer and T. Schulz for giving us access to yeast fermentation; M. Strauss for giving superb assistance and training to access the microscopes of the MPIB cryo-EM Facility; and members of our groups for comments and discussions. **Funding:** This study was supported by the Max Planck Gesellschaft E.C., the European Commission (ERC Advanced Investigator Grants EXORICO to E.C. and Glowsome to E.H.), the Deutsche Forschungsgemeinschaft (DFG SFB646, SFB1035, GRK1721, and CIPSM to E.C. and HU363/10-5 and HU363/12-1 to E.H.), and the Louis Jeantet Foundation (to E.C.). **Author contributions:** E.C. and E.H. initiated the project; L.F., S.F., and J.M.S. identified initial biochemical conditions; J.M.S. and S.F. carried out cryo-EM sample preparation; J.M.S. collected cryo-EM data and performed image processing; J.M.S. built the structure with help from S.F.; S.F. carried out recombinant in vitro assays; and J.M.S., S.F., L.F., E.C., and E.H. analyzed the structure and wrote the paper. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The cryo-EM density maps are deposited in the Electron Microscopy Data Bank under accession numbers EMD-4301 and EMD-4302. The atomic model is deposited in the Protein Data Bank (PDB) under accession numbers 6FSZ and 6FT6. All other data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/219/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
References (31–36)

21 November 2017; accepted 22 February 2018
Published online 8 March 2018
10.1126/science.aar5428

IMMUNOLOGY

Germinal center antibody mutation trajectories are determined by rapid self/foreign discrimination

Deborah L. Burnett,^{1,2} David B. Langley,¹ Peter Schofield,^{1,2} Jana R. Hermes,¹ Tyani D. Chan,^{1,2} Jennifer Jackson,¹ Katherine Bourne,¹ Joanne H. Reed,¹ Kate Patterson,¹ Benjamin T. Porebski,³ Robert Brink,^{1,2} Daniel Christ,^{1,2*} Christopher C. Goodnow^{1,2*}

Antibodies have the specificity to differentiate foreign antigens that mimic self antigens, but it remains unclear how such specificity is acquired. In a mouse model, we generated B cells displaying an antibody that cross-reacts with two related protein antigens expressed on self versus foreign cells. B cell anergy was imposed by self antigen but reversed upon challenge with high-density foreign antigen, leading to germinal center recruitment and antibody gene hypermutation. Single-cell analysis detected rapid selection for mutations that decrease self affinity and slower selection for epistatic mutations that specifically increase foreign affinity. Crystal structures revealed that these mutations exploited subtle topological differences to achieve 5000-fold preferential binding to foreign over self epitopes. Resolution of antigenic mimicry drove the optimal affinity maturation trajectory, highlighting the value of retaining self-reactive clones as substrates for protective antibody responses.

Antibodies often distinguish nearly identical foreign and self antigens, such as the glycolipids on *Campylobacter jejuni* cell walls and those on human nerve cells, with fewer than 0.1% of infected people producing cross-reactive antibodies that result in paralysis and Guillain-Barré syndrome (1). Apparent limits to antibody discrimination of self versus foreign antigens are exploited by HIV, lymphocytic choriomeningitis virus, and Lassa fever virus. These viruses establish persistent infections and evade antibodies by mimicking self glycoproteins and cloaking their foreign envelope proteins with self glycans (2–5). Although self-reactivity can be removed from antibodies by V(D)J recombination (6) or by V-region hypermutation (7–9), the cellular basis and mutational pathways for resolving foreign-self mimicry after infection or immunization remain undefined.

We engineered bone marrow chimeric mice (Fig. 1, A and B, and figs. S1 to S4) in which the majority of developing B cells reaching the spleen from the bone marrow were polyclonal and expressed CD45.2 (CD45.2⁺). However, 1% of transitional B cells and 0.1% of mature follicular B cells were CD45.1⁺ SW_{HEL} cells, which carry HyHEL10 antibodies on their surfaces. HyHEL10 antibodies have a defined structure and low affinity for a self protein [hen egg lysozyme with three substitutions (HEL^{3X}) (10–13); $1/K_D$ (equilibrium

dissociation constant) = $1.2 \times 10^7 \text{ M}^{-1}$] and for a structurally similar foreign protein [duck egg lysozyme (DEL); $1/K_D = 2.5 \times 10^7 \text{ M}^{-1}$]. In one group of chimeric mice, the self protein was displayed on all cells as an integral membrane protein, mHEL^{3X}, encoded by a transgene with a ubiquitin promoter (14). When SW_{HEL} B cells were self-reactive, they reached the spleen as short-lived anergic cells with decreased surface immunoglobulin M (IgM) but normal surface IgD (Fig. 1B and figs. S1 to S4), located primarily in the T cell zone (Fig. 1C) as in other anergic models (15–17). The frequency of anergic SW_{HEL} cells was lower than the frequency of circulating anergic IgD⁺ IgM^{lo} VH4-34⁺ B cells, which recognize ubiquitous cell surface antigens and mutate away from self-reactivity in humans (8).

We first tested whether self-reactive SW_{HEL} B cells could respond to a foreign antigen that perfectly mimicked self antigen. Sheep red blood cells (SRBCs) were covalently coupled with self antigen at surface densities equivalent to those on endogenous mouse red blood cells (MRBCs) or 30-fold higher (Fig. 1D). Despite equal levels of T cell help for germinal center (GC) responses by the diverse repertoire of other B cells (Fig. 1F), self-reactive SW_{HEL} B cells entered GCs only when SRBCs carried high antigen density (Fig. 1G). SRBCs with low antigen density could nevertheless induce GC responses from SW_{HEL} B cells that were not self-reactive. These results are consistent with previous evidence that helper T cells cooperate with anergic B cells only when B cell–receptor cross-linking by foreign antigen is greater than that induced by self antigen (18).

Next, we tested the response of self-reactive SW_{HEL} B cells to DEL, which differs from self antigen at four residues that make contact with

the HyHEL10 heavy chain (H chain) (figs. S5 and S6A). GC reactions were initiated with unconjugated SRBCs, and 11 days later, SW_{HEL} B cells were recruited into the reactions synchronously by a booster immunization with DEL coupled at high density to SRBCs (Fig. 2A). Four days after immunization with DEL-conjugated SRBCs, SW_{HEL} B cells constituted ~20% of all GC B cells and were present in comparable total numbers regardless of self-reactivity (Fig. 2B and figs. S5 and S6, B and C). When the SW_{HEL} GC B cells were self-reactive, they had lower densities of surface IgG1 per cell (Fig. 2C and fig. S6D), likely caused by engagement with self antigen on neighboring cells. At this early time point, the frequencies and numbers of IgG1⁺ and IgG1⁺ SW_{HEL} B cells with low binding to self antigen were increased when the cells were self-reactive (Fig. 2, C and D, and fig. S6C). These low-binding cells had increased frequencies of missense mutations (fig. S6, E and F), with 55% having acquired a Ser⁵²→Arg⁵² (S52R) or Ser⁵²→Asn⁵² (S52N) mutation in complementarity-determining region 2 (CDR2) (Fig. 2E). Both mutations greatly decreased affinities for both self and foreign proteins (fig. S7 and table S1).

To determine whether rapid selection for mutant GC B cells with decreased affinity for self protein was followed by maturation of affinity for foreign protein, we analyzed antibody mutations 4, 7, and 11 days after SW_{HEL} B cells were challenged with DEL-conjugated SRBCs (Fig. 3A and fig. S8). On day 4, the frequencies of S52R and S52N mutations were again significantly increased (11.55 versus 3.55%; $P = 0.0093$) when SW_{HEL} B cells were self-reactive. However, the frequencies decreased on days 7 and 11. An Ile²⁹→Phe²⁹ (I29F) mutation in CDR1 became prevalent instead on day 7, occurring as a single substitution in 31% of SW_{HEL} B cells when they were self-reactive compared with only 1.7% when they were not. I29F conferred the property of distinguishing foreign from self protein, causing a 10-fold decrease in self affinity and a 2.6-fold increase in foreign affinity (Fig. 3A, fig. S7, and table S1).

The I29F mutation became paired with Ser⁵²→Thr⁵² (S52T) and Tyr⁵³→Phe⁵³ (Y53F) mutations in CDR2. This pattern emerged in a small subset of self-reactive cells on day 7, but these mutations became most prevalent as pairs or a trio by day 11. S52T and Y53F were rarely found individually, but combined with the I29F foundation mutation, they increased foreign-self discrimination. Cells with the combined mutations retained $1 \times 10^6 \text{ M}^{-1}$ affinity for self but showed progressively increasing foreign affinity, up to to $6 \times 10^9 \text{ M}^{-1}$. Strong epistatic (nonadditive) effects were observed. For example, the I29F-S52T-YF3F trio increased the apparent differential binding energy ($\Delta\Delta G$) for binding foreign antigen by -3.3 kcal/mol , compared with -1.6 kcal/mol expected for additive effects of the individual mutations (table S1). This trio of mutations became even more prevalent when self-reactive SW_{HEL} B cells were recruited at the outset of the GC reaction and analyzed 15 days later (Fig. 3B

¹Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, New South Wales 2010, Australia. ²St. Vincent's Clinical School, University of New South Wales Sydney, Sydney, Australia. ³Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.

*Corresponding author. Email: d.christ@garvan.org.au (D.C.); c.goodnow@garvan.org.au (C.C.G.)

and fig. S9). Thus, an antibody that was initially unable to distinguish foreign from self antigen had evolved a 5000-fold differential binding to foreign antigen over self antigen by first mutating away from binding self antigen and subsequently mutating toward binding foreign antigen. SW_{HEL} -derived cells that had lost self-binding but retained foreign binding were also frequent among the IgG1⁺ memory B cell compartment (fig. S10). Foreign antigen-specific IgG1 serum titers were increased in mice with initially self-reactive SW_{HEL} B cells (fig. S11).

A different, less optimal evolutionary trajectory prevailed when SW_{HEL} B cells were not self-reactive. This trajectory was dominated by acquisition of a CDR2 mutation (Y58F) alone, paired, or in trio with S52T and Y53F (Fig. 3). Y58F alone or with S52T and Y53F increased self-affinity by a factor of four, explaining why this trajectory was not taken by self-reactive SW_{HEL}

cells. The Y58F-S52T-Y53F trio increased foreign affinity to $2 \times 10^9 M^{-1}$, which was one-third of the affinity obtained with the I29F-S52T-Y53F trio selected through the self-reactive trajectory.

To understand how these three mutations conferred a 5000-fold differential binding to foreign protein over self, we used x-ray crystallography to analyze the structure of HyHEL10^{I29F,S52T,Y53F} in complex with DEL (Fig. 4, table S2, and movie S1) compared to that of wild-type HyHEL10 (HyHEL10^{WT}) in complex with HEL (19). I29F resulted in a structural rearrangement of the CDR1 loop to accommodate the larger phenylalanine side chain. Displacement of this loop (arrow 1 in Fig. 4C) opened up additional structural adjustments of CDR2 (arrow 2 in Fig. 4C) and, in particular, repositioned Y53F to interact with a hydrophobic pocket formed on the surface of DEL by the short Ala75 (A75) side chain, which is in contrast to the much longer leucine

in HEL. The CDR2 backbone adjustments also allowed replacement of the smaller S52 side chain with threonine. Thus, our structural analyses were in agreement with the observed mutational trajectory, whereby the I29F foundation mutation introduces structural rearrangements into CDR1 and CDR2. These rearrangements enable secondary mutations at positions 52 and 53, which selectively increase foreign affinity in an epistatic manner. Binding studies confirmed that I29F confers 50-fold-lower binding to self versus foreign antigen by exploiting the Leu⁷⁵→Ala⁷⁵ (L75A) foreign pocket coupled with the adjacent Glu⁷³→Lys⁷³ charge reversal (table S1). This

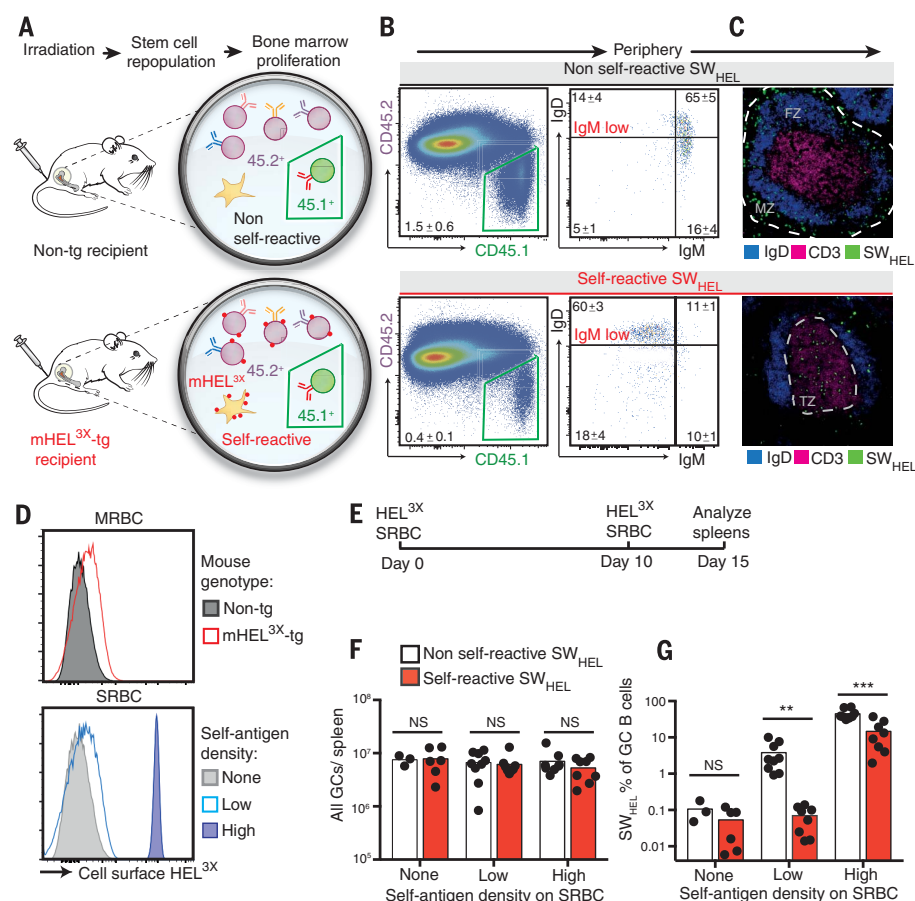


Fig. 1. Recruitment of anergic cells into GCs requires higher foreign antigen density.

Construction of parallel groups of hematopoietic chimeras (A) and analysis of their spleens by (B) flow cytometry of all B cells (left) or CD45.1⁺ SW_{HEL} B cells (right) ($n = 14$ spleens per group; values are mean percentages \pm SEM) or (C) immunohistology showing localization of SW_{HEL} B cells (green), other B cells (blue), and T cells (pink). Dashed lines indicate borders between splenic compartments. FZ, follicular zone; MZ, marginal zone; TZ, T cell zone. (D) Relative abundances of self HEL^{3X} on MRBCs from mHEL^{3X}-tg or nontransgenic mice and on foreign SRBCs conjugated with 0 (none), 0.1 (low), or 10 (high) $\mu g/ml$ HEL^{3X}. (E) Timing of chimera immunizations. (F) Total GC cells per spleen. (G) Percentages of SW_{HEL} cells among GC B cells. NS, not significant ($P > 0.05$); $^{**}P < 0.01$; $^{***}P < 0.001$; Student's t test. Data points represent one chimera (two experiments, 16 to 26 chimeras in each).

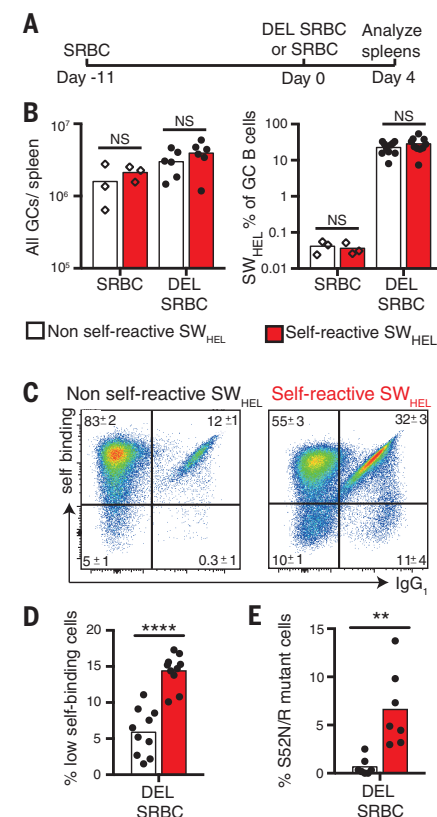


Fig. 2. Binding to similar foreign and self antigens triggers rapid mutation away from self affinity.

(A) Timing of chimera immunizations. (B) Total GC cells per spleen and percentages of SW_{HEL} cells among GC cells in chimeras receiving DEL-coupled SRBCs or unconjugated SRBCs. (C and D) Analysis of SW_{HEL} GC B cells showing (C) the percentages (means \pm SEM) that bind 0.14 μM HEL^{3X} or express cell surface IgG1 and (D) the percentages of nonbinding cells. (E) Percentages of sorted and individually sequenced SW_{HEL} GC cells with an S52N or S52R mutation. NS, $P > 0.05$; $^{**}P < 0.01$, $^{***}P < 0.0001$; Student's t test. Each data point represents one mouse. Data are from at least two independent experiments, each involving three to four mice in DEL-coupled SRBC groups.

effect was further confirmed by solving the structure of HyHEL10^{I29F} in complex with DEL (fig. S12).

We next identified anergic B cells in the mHEL^{3X} transgenic (mHEL^{3X}tg) mice within a polyclonal repertoire that displayed micromolar affinity for the same self antigen and tested whether these B cells too could resolve antigenic mimicry. HEL^{3X}-binding B cells constituted 2.7% of IgD⁺ IgM^{lo} anergic B cells and 0.5% of all splenic B cells (fig. S13A). These

were sorted and added at 0.5% frequency to unselected CD45.1⁺ B cells, and the polyclonal mixture was injected together with T cells into mHEL^{3X}tg RagT^{-/-} mice immunized with DEL-conjugated SRBCs. In the recipients, 96% of the GC response was derived from the unselected CD45.1⁺ B cells, presumably recognizing mostly SRBC antigens. In contrast, 61% of the DEL-binding GC response was derived from the polyclonal HEL^{3X}-binding anergic CD45.2⁺ B cells

(fig. S13B). Only 9.7% of these cells still bound self antigen, whereas 53% bound foreign DEL selectively (fig. S13C). Thus, in a normal repertoire, cells with micromolar affinity for self HEL^{3X} are dominant contributors to the GC response against the self mimic DEL and rapidly lose binding to self.

The findings here extend evidence for auto-antibody redemption in human antibodies (7–9) by showing that mutation away from self-reactivity

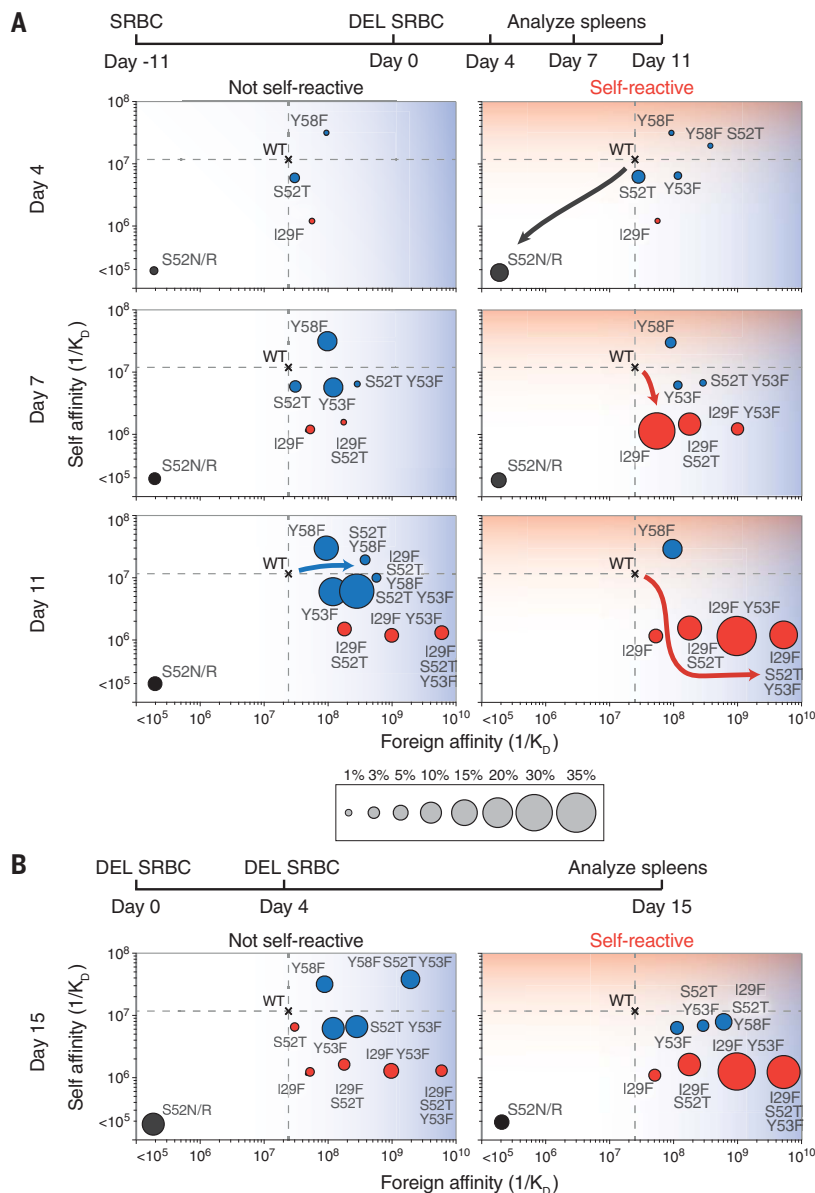


Fig. 3. Self-reactive cells follow distinct mutational trajectories to lose self-binding capacity, leading to optimal affinity for foreign antigen. Chimera immunization time points were selected to (A) synchronize recruitment of SW_{HEL} B cells into established GCs or (B) recruit SW_{HEL} B cells into GCs from the outset. SW_{HEL} GC B cells were subjected to single-cell sequencing. Dashed lines show the affinities of unmutated (WT) antibody for self and foreign proteins. Circles show the affinities of recurring mutant antibodies for self and foreign proteins. Areas of the circles denote the percentages of SW_{HEL} B cells with the indicated mutations. Red circles indicate mutations more frequent in self-reactive SW_{HEL} B cells, and blue circles represent mutations more frequent in the non-self-reactive SW_{HEL} B cells. Data are from one experiment and are representative of two experiments, each involving two to three mice per group at each time point. N/R, Asn or Arg.

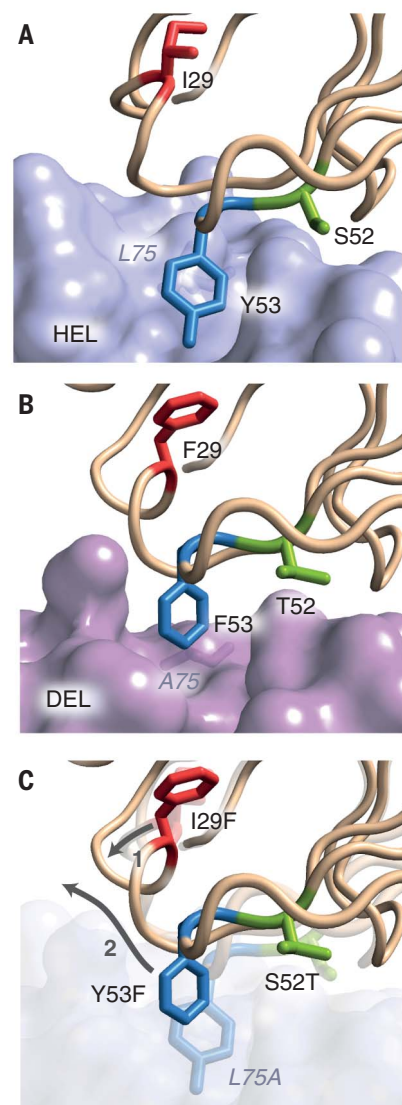


Fig. 4. Structural basis of mutation away from self. X-ray crystallographic structures of (A) unmutated HyHEL10 in complex with HEL and (B) HyHEL10^{I29F,S52T,Y53F} triple mutant antibody in complex with DEL. (C) Overlay of both structures showing the structural rearrangement of the CDR1 loop caused by the I29F mutation (arrow 1) and the complementary structural adjustments of positions 52 and 53 in the CDR2 loop to exploit the L75A pocket in the foreign antigen (arrow 2).

precedes mutation toward foreign affinity to create distinctive epistatic trajectories. Self-reactivity, rather than being a barrier to immunization, directed cells down an alternative trajectory, which produced a higher final affinity for the foreign immunogen. The higher threshold to activate anergic cells and recruit them to GC reactions is nevertheless an important constraint: for instance, a low density of Env molecules on HIV virions may fail to activate anergic B cells with moderate cross-reactive affinity for self glycans attached to foreign and self polypeptides, precluding mutation trajectories away from self-reactivity.

Antibody mutation away from self-reactivity in GC reactions defers the need to acquire stringent self-tolerance until after an infection. This process is complementary to the concept of purging self-reactive antibodies from the preimmune repertoire before they can be tested for binding foreign antigen (1, 6, 20–22) as well as to Jerne's hypothesis of mutation away from self in the bone marrow and bursa (23). Both concepts create a "holes in the repertoire" problem if applied too stringently (24, 25). Crucially, autoantibody redemption minimizes the potential for microbes to evolve antigens that are "almost self," which could otherwise be recognized only by preimmune antibodies that had been deleted or edited in the bone marrow. Mutation away from self in response to one foreign antigen may allow progeny B cells to respond to an unrelated foreign antigen later. For example, intestinal microbes may induce polyspecific B cells to mutate away from self, providing a self-tolerant repertoire that would not be available in individuals treated

with antibiotics or raised in a more hygienic environment. The evolution of an antibody along a limited set of mutation trajectories, driven by two selection pressures for higher affinity for one ligand and lower affinity for another, provides an example of deterministic molecular evolution. Our findings provide insights into the GC reaction and the evolution of specificity in antibody-antigen interactions.

REFERENCES AND NOTES

1. B. R. Wakerley, N. Yuki, *Expert Rev. Clin. Immunol.* **9**, 627–639 (2013).
2. B. F. Haynes *et al.*, *Science* **308**, 1906–1908 (2005).
3. D. D. Pinschewer *et al.*, *J. Clin. Invest.* **114**, 988–993 (2004).
4. R. Wyatt *et al.*, *Nature* **393**, 705–711 (1998).
5. H. B. Gristick *et al.*, *Nat. Struct. Mol. Biol.* **23**, 906–915 (2016).
6. D. Nemazee, M. Weigert, *J. Exp. Med.* **191**, 1813–1817 (2000).
7. J. Tan *et al.*, *Nature* **529**, 105–109 (2016).
8. J. H. Reed, J. Jackson, D. Christ, C. C. Goodnow, *J. Exp. Med.* **213**, 1255–1265 (2016).
9. Z. Sabouri, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2567–E2575 (2014).
10. T. G. Phan *et al.*, *J. Exp. Med.* **197**, 845–860 (2003).
11. T. G. Phan *et al.*, *J. Exp. Med.* **203**, 2419–2424 (2006).
12. E. A. Padlan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5938–5942 (1989).
13. D. Paus *et al.*, *J. Exp. Med.* **203**, 1081–1091 (2006).
14. T. D. Chan *et al.*, *Immunity* **37**, 893–904 (2012).
15. J. G. Cyster, S. B. Hartley, C. C. Goodnow, *Nature* **371**, 389–395 (1994).
16. D. A. Fulcher, A. Basten, *J. Exp. Med.* **179**, 125–134 (1994).
17. C. C. Goodnow *et al.*, *Nature* **334**, 676–682 (1988).
18. M. P. Cooke *et al.*, *J. Exp. Med.* **179**, 425–438 (1994).
19. M. Acchione *et al.*, *Mol. Immunol.* **47**, 457–464 (2009).
20. F. M. Burnet, *The Clonal Selection Theory of Acquired Immunity* (Vanderbilt Univ. Press, 1959).
21. J. Lederberg, *Science* **129**, 1649–1653 (1959).
22. H. Wardemann *et al.*, *Science* **301**, 1374–1377 (2003).
23. N. K. Jerne, *Eur. J. Immunol.* **1**, 1–9 (1971).

24. D. Nemazee, *Immunol. Today* **17**, 25–29 (1996).
25. A. S. Perelson, G. F. Oster, *J. Theor. Biol.* **81**, 645–670 (1979).

ACKNOWLEDGMENTS

We thank the Garvan Institute Australian BioResources, Garvan Molecular Genetics, and the Garvan Flow Cytometry Facility for animal husbandry, genotyping, and cell sorting. **Funding:** This work was supported by program grants 1016953 and 1113904, project grant 1108800, and fellowships 585490 and 1081858 from the National Health and Medical Research Council; by Discovery grants 160104915 and 140103465 from the Australian Research Council; and by The Bill and Patricia Ritchie Foundation.

Author contributions: D.L.B. performed and analyzed mouse experiments; P.S. performed and analyzed binding affinity experiments; D.B.L. performed and analyzed crystallography experiments; J.J. and K.B. generated antibodies and antigens; T.D.C., J.H.R., and J.R.H. developed mHEL^{3X}tg mice; R.B. devised and developed the mHEL^{3X} × SW_{HEL} system; C.C.G. supervised B cell biology aims; D.C. supervised structural and biophysical aims; D.L.B., P.S., D.B.L., R.B., D.C., and C.C.G. designed and interpreted experiments; D.L.B., D.B.L., P.S., K.P., R.B., D.C., and C.C.G. prepared figures; B.T.P. generated movie S1; and D.L.B. drafted and R.B., D.C., and C.C.G. revised the manuscript.

Competing interests: The authors have no competing interests.

Data and materials availability: Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 5VJO and 5VJQ. All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6385/223/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 and S2
References (26–35)
Movie S1

14 July 2017; resubmitted 2 December 2017

Accepted 20 February 2018

10.1126/science.aao3859

Light your way.

Get the full picture with trusted tumor immune response results from our in-house validated antibodies.

When it comes to your tumor immune response research, blind spots are unacceptable. Independent testing has demonstrated that 75% of antibodies in today's market are non-specific or simply do not work at all.* But at Bethyl, we've manufactured and validated every antibody we make on site to ensure target specificity and sensitivity. All to guarantee our antibodies will function as designed in your assay 100% of the time. More than 10,000 independent citations over the past 15 years have proven that, at Bethyl, we put a lot in every drop.

See our data at bethyl.com/immuno-oncology

*Weller, MG, *Analytical Chemistry Insights*:11, 21-27 (2016).

Antibodies shown: Rabbit anti-PD-L1 (red, A700-020) & Lamin-A/C (green, A303-430A) in FFPE lung.
©2018 Bethyl Laboratories, Inc. All rights reserved.



Really Good Antibodies



CALL FOR NOMINATIONS FOR A \$1 MILLION PRIZE THE BIGGEST PRIZE IN THE WORLD FOR INNOVATION IN THE FIELD OF ALTERNATIVE FUELS FOR TRANSPORTATION

**THE 2018 ERIC AND SHEILA SAMSON
PRIME MINISTER'S PRIZE IS
A \$ 1 MILLION INTERNATIONAL PRIZE
FOR INNOVATION IN ALTERNATIVE
FUELS FOR TRANSPORTATION.**

If you are actively engaged in innovative, paradigm-shifting research or groundbreaking technological development worthy of significant and widespread attention, you can qualify for this prestigious prize.

For more information visit
prize.fuelchoicesinitiative.com

DEADLINE FOR NOMINATIONS
MAY 31, 2018
OPEN TO CITIZENS OF ALL COUNTRIES

For further information and forms:

Go to: prize.fuelchoicesinitiative.com

Or contact: OmerM@most.gov.il



**Ministry of Science
and Technology**



PNAS Congratulates 2017 Cozzarelli Prize Recipients

PNAS has selected six outstanding PNAS articles for the 2017 Cozzarelli Prize. The award was established in 2005 as the PNAS Paper of the Year Prize and renamed the Cozzarelli Prize in 2007 to honor late PNAS Editor-in-Chief Nicholas R. Cozzarelli. Articles receiving the Cozzarelli Prize were chosen from more than 3,200 research reports published by PNAS in 2017 and represent the six broadly defined classes under which the National Academy of Sciences is organized.

CLASS I: PHYSICAL AND MATHEMATICAL SCIENCES

Origin of the RNA world: The fate of nucleobases in warm little ponds

Ben K. D. Pearce, Ralph E. Pudritz, Dmitry A. Semenov, and Thomas K. Henning (2017) *Proc Natl Acad Sci USA* 114:11327–11332

CLASS II: BIOLOGICAL SCIENCES

Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence

Linlin Zhang, Anyi Mazo-Vargas, and Robert D. Reed (2017) *Proc Natl Acad Sci USA* 114:10707–10712

CLASS III: ENGINEERING AND APPLIED SCIENCES

Nondestructive nanostraw intracellular sampling for longitudinal cell monitoring

Yuhong Cao, Martin Hjort, Haodong Chen, Fikri Birey, Sergio A. Leal-Ortiz, Crystal M. Han, Juan G. Santiago, Sergiu P. Pașca, Joseph C. Wu, and Nicholas A. Melosh (2017) *Proc Natl Acad Sci USA* 114:E1866–E1874

CLASS IV: BIOMEDICAL SCIENCES

Three-dimensional visualization and a deep-learning model reveal complex fungal parasite networks in behaviorally manipulated ants

Maridel A. Fredericksen, Yizhe Zhang, Missy L. Hazen, Raquel G. Loreto, Colleen A. Mangold, Danny Z. Chen, and David P. Hughes (2017) *Proc Natl Acad Sci USA* 114:12590–12595

CLASS V: BEHAVIORAL AND SOCIAL SCIENCES

Language from police body camera footage shows racial disparities in officer respect

Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt (2017) *Proc Natl Acad Sci USA* 114:6521–6526

CLASS VI: APPLIED BIOLOGICAL, AGRICULTURAL, AND ENVIRONMENTAL SCIENCES

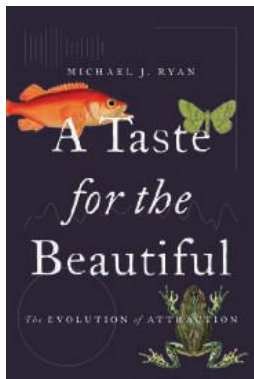
Methyl-compound use and slow growth characterize microbial life in 2-km-deep subseafloor coal and shale beds

Elizabeth Trembath-Reichert, Yuki Morono, Akira Ijiri, Tatsuhiko Hoshino, Katherine S. Dawson, Fumio Inagaki, and Victoria J. Orphan (2017) *Proc Natl Acad Sci USA* 114:E9206–E9215

PNAS

www.pnas.org

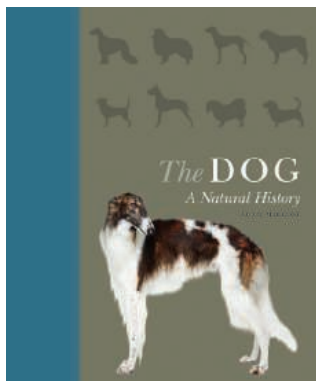
Podcast interviews with the authors will be available at:
www.pnas.org/site/misc/cozzarelliprize.xhtml



"Michael Ryan guides readers through the complex and fascinating path of the evolution of beauty with the authority of somebody who is responsible for some of the subject's most important milestones."

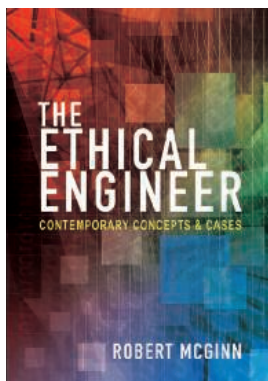
—Leonida Fusani, University of Vienna and University of Veterinary Medicine, Vienna, Austria

Cloth \$27.95



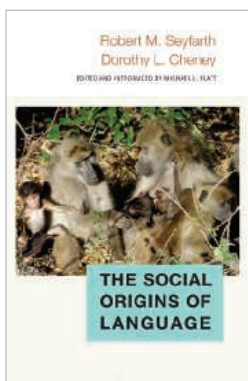
An accessible and richly illustrated introduction to the natural history of dogs—from evolution, anatomy, cognition, and behavior to the relationship between dogs and humans

Cloth \$27.95



Presenting key ethics concepts and eighteen rich case studies, Robert McGinn illuminates the ethical dimension of engineering practice and helps students and professionals determine engineers' context-specific ethical responsibilities.

Paper \$55.00



"An inspiring discourse on the social function of communication. In response to a magnificent essay by Seyfarth and Cheney, the grandmasters of primate communication studies, leading scholars engage in a vivid debate on the origins of language. A must-read!"

—Julia Fischer, author of *Monkeytalk*

Cloth \$32.95

 PRINCETON UNIVERSITY PRESS



NEW! Low-Noise Ultra-Fast Digital Patch Clamp Amplifier System

- High bandwidth for fastest signal characterization
- Single-channel and whole-cell patch clamp recordings
- Digital compensation circuitry for precision and signal fidelity
- Quick and easy setup
- Bundled SutterPatch® software built on Igor Pro platform



dPatch®

The next generation Digital Patch Clamp Amplifier System. Combining high-speed, high-resolution digital processing, precision A/D circuitry, integrated data acquisition and bundled SutterPatch® software, the dPatch system provides capabilities previously out of reach for the electrophysiologist. Available in either a single- or double-headstage configuration, the dPatch meets the requirements of today's experiments and anticipates the demands of tomorrow's.


SUTTER INSTRUMENT®

PHONE: +1.415.883.0128 | FAX: +1.415.883.0572
EMAIL: INFO@SUTTER.COM | WWW.SUTTER.COM

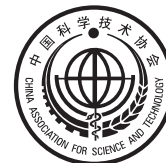
Step up your job search with *Science Careers*



ScienceCareers

FROM THE JOURNAL SCIENCE  AAAS

CALL FOR PAPERS



SciencePartnerJournals.org/research

Research

OFFICIAL JOURNAL OF CAST

An Open Access Journal

Research is a comprehensive, interdisciplinary, and selective **Open Access journal from the China Association of Science and Technology (CAST)**, published quarterly and distributed by the **American Association for the Advancement for Science (AAAS)**. *Research* provides an international platform for academic exchange, research collaboration, and technological advancements. The journal will publish fundamental research in the life and physical sciences as well as important findings or issues in engineering and applied science.

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



Thermometers

Verifying sample temperatures has never been easier with the new, cable-free H-B Frio-Temp Calibrated Electronic Verification Lollipop Stem Thermometers from Bel-Art-SP Scienceware. They are compact, all-in-one units with the display attached directly to the probe instead of with a cable, as

commonly found in electronic verification thermometers. The entire digital unit can be placed inside a refrigerator, incubator, or room to verify the temperature of the samples within. Cabled electronic verification thermometers are useful in ovens, freezers, and other applications requiring temperature verification without disruption of the sample environment. However, for applications where the temperature is verified less frequently or can be checked during routine entry of the environment, these compact thermometers are the superior choice. Available with general calibration, or with calibration specific to refrigerators, incubators, or ambient temperatures, all of the Lollipop Stem Thermometers are supplied with calibration documents meeting the requirements of ISO 17025.

Bel-Art-SP Scienceware

For info: 800-423-5278
www.belart.com

Protein A Kit

The Gyrolab Protein A Kit offers increased efficiency and throughput for quantification of residual Protein A ligands. Two ready-to-use versions of the kit—Native Protein A and MabSelect Sure—are designed for use with Gyrolab systems, and include an enhanced protocol for automated acid pretreatment of harvested samples from downstream purification of biotherapeutics. Determination of residual Protein A is an essential regulatory requirement for the safety evaluation of biotherapeutics. During purification of biotherapeutics, Protein A can leach from the chromatography support and coelute with the therapeutic antibody product. If bound to immunoglobulins, Protein A can increase the risk of adverse reactions.

Gyros Protein Technologies

For info: +46-(0)-18-56-63-00
www.gyrosproteintechnologies.com

Blood-Cell Washer

Thermo Scientific's CW3 Cell Washer enables precise, thorough, reproducible, and rapid blood-cell washing of up to 24 tubes in a 3-min run. Users can select one of the preset programs that have been designed to run in automatic mode, accelerate blood-cell washing cycles, and achieve significant time savings while allowing for maximum efficiency and sample safety. Furthermore, the flexibility to use 12- or 24-place rotors and standardized, repeatable procedures enables the instrument to adapt to the processes and protocols already in place within the lab. Effective blood-cell washing is extremely important for removing plasma and unwanted antibodies to ensure that patients will not be affected by adverse reactions and that blood-transfused products will retain all their beneficial attributes.

Thermo Fisher Scientific

For info: 800-955-6288
www.thermofisher.com/cellwasher

Life Science Imaging Software

Olympus Life Science's imaging software cellSens features an automatic object-tracking function—a dedicated solution to analyze and document dynamic processes within living samples. The software can adapt to different scenarios, such as samples containing fast-moving objects or growing cells, while precise manual tracking is also possible. Measured objects are reported as track diagrams, tables, and images, offering a range of possibilities to analyze, present, and export results. cellSens also features a user-friendly onscreen operation. Together with 4K-ready microscopy accessories, such as the UC90 camera, the software presents a highly detailed image together with a clear, easy-to-use interface. For effective collaboration and discussion, a dedicated conference mode is also available—the image is displayed full-screen, and annotation tools are just a single click away. cellSens gives users access to more data and saves time in both routine and advanced microscopy applications, making it an ideal choice for research and clinical laboratories.

Olympus

For info: 704-877-8801
www.olympus-lifescience.com

Compact Gas Analyzers

The Hidden ExQ quantitative gas analyzer is a multifeatured, compact mass spectrometer system that provides continuous online analysis of dynamic gas streams at pressures from subatmospheric up to 30 bar. This fully integrated system is now configurable for benchtop and rack-mounted operations. Connection to the process is via a flexible, heated inert capillary line, with a sample-consumption rate less than 10 mL/min and a response time of less than 300 ms at near-atmospheric pressures. The comprehensive range of process-interface options enables analysis of up to 16 individual process streams. Adaptors are available for direct connection to most standard thermogravimetric/thermal analysis instruments. External process data, sample temperature, or mass, for example, can be imported for combined integration and presentation with the mass spectral data. Minor species composition is measurable down to just 100 ppb, with a choice of mass range of 200 amu and 300 amu. An optional CO analyzer is available where monitoring and data integration of low levels of carbon monoxide is a requirement.

Hidden Analytical

For info: 888-964-4336
www.hiddenanalytical.com

Imaging System

Syngene's G:BOX is a range of automated multipurpose gel- and blot-imaging systems. Using HI-LED (high-intensity) lighting and updated image-capture software, these flexible systems guarantee cost-effective imaging and faster workflow with a huge range of fluorescence gel and blot applications. Featuring the choice to add a full spectrum of high-intensity blue, green, red, and infrared HI-LEDs that are up to 200 times brighter than standard LEDs, the new G:BOX options provide faster exposure times and great images in just one click. All systems in the G:BOX range are controlled via GeneSys software, which now includes a simple icon selection of preset, stain-free protein gel imaging conditions. The icon is based on optimum filter and lighting conditions that can accurately detect nanogram levels of protein on a stain-free gel; the software autocalibrates to each gel or blot size to generate publication-quality images every time.

Syngene

For info: +44-(0)-1223-727123
www.syngene.com/g-box-chemi-xx6

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

STAND TOGETHER

Be a Force for Science



GET THE FACTS

Understand the science behind the issues that matter.

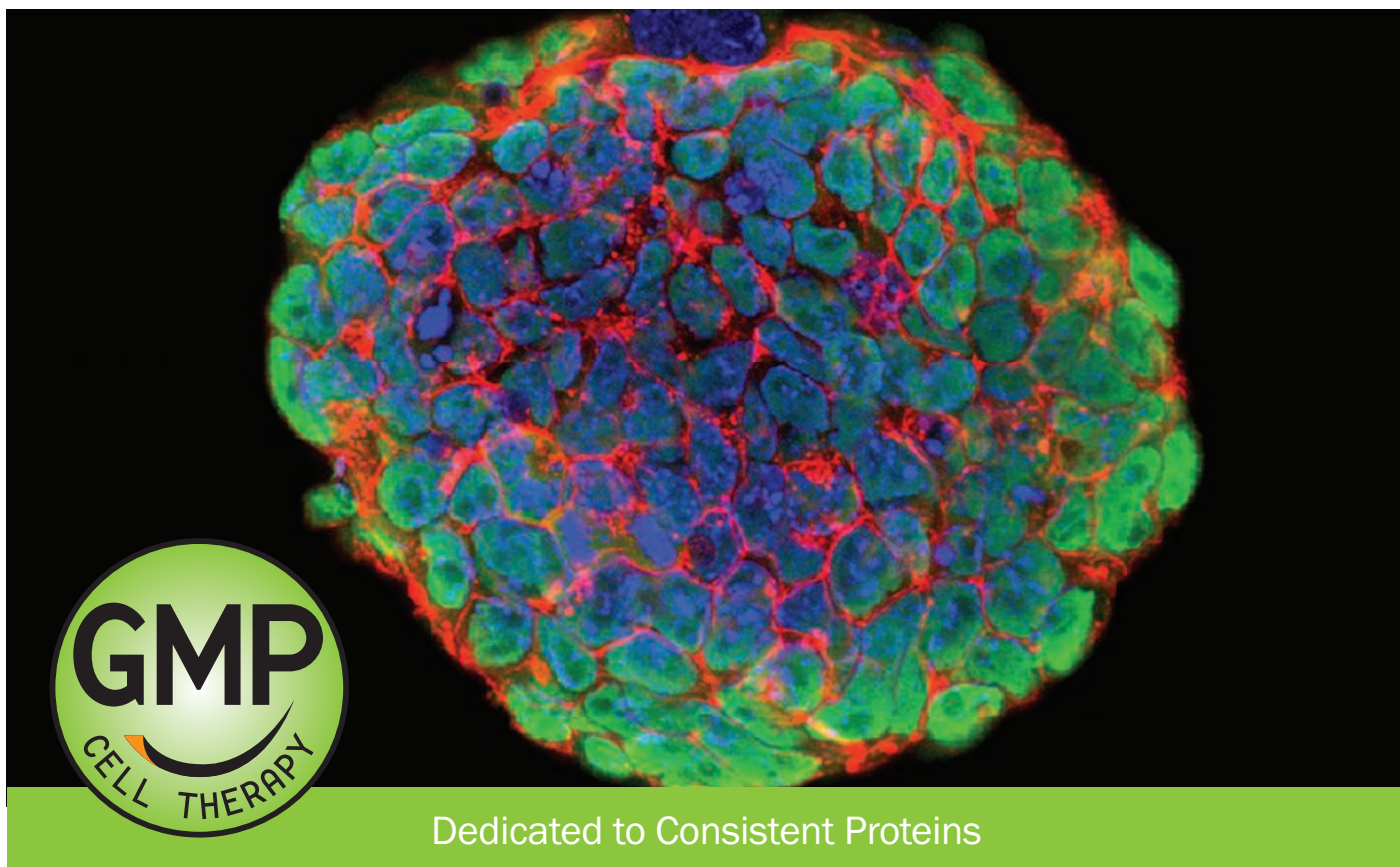
FOLLOW AAAS ADVOCACY

Champion public discussion and evidence-based policy.

TAKE ACTION

Learn ways you can become an advocate and stand up for science.

GMP-Grade Growth Factors



The Widest Selection for

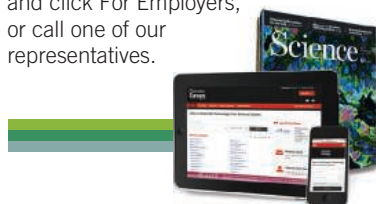
- Cell Manufacturing
- Immunotherapy
- Regenerative Medicine
- Stem Cell Therapy

Learn more | rndsystems.com/gmp

Science Careers

SCIENCE CAREERS ADVERTISING

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.



AMERICAS

+1 202 326-6577
+1 202 326-6578
advertise@sciencecareers.org

EUROPE, INDIA, AUSTRALIA, NEW ZEALAND, REST OF WORLD

+44 (0) 1223 326527
advertise@sciencecareers.org

CHINA, KOREA, SINGAPORE, TAIWAN, THAILAND

+86 131 4114 0012
advertise@sciencecareers.org

JAPAN

+81 3-6459-4174
advertise@sciencecareers.org

CUSTOMER SERVICE

AMERICAS

+1 202 326-6577
REST OF WORLD
+44 (0) 1223 326528

advertise@sciencecareers.org

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

ScienceCareers
FROM THE JOURNAL SCIENCE 

ScienceCareers.org

myIDP: A career plan customized for you, by you.



For your career in science, there's only one **Science**



Recommended by
leading professional
societies and the NIH

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with:

 AAAS



FASEB
Federation of American Societies
for Experimental Biology





Implementation of International Conventions

Global environmental issues have aroused concern over the sustainability of human development, while also providing a cooperation opportunity for solving the common problems facing mankind. The College of Environmental Sciences and Engineering (CESE), Peking University established the research team performing integrated and multiply disciplinary studies covering a wide range of research areas, from fundamental academic issues, technologies and alternatives, and policies of global environmental problems, which provides important support for the negotiations and implementation of various global environmental agreements.

The Montreal Protocol on the Protection of the Ozone Layer, praised by Kofi Annan as “perhaps the single most successful international environmental agreement to date”, is a prime example of global cooperation on environmental issues. Our research team was the most important institution providing decision-making support to the Chinese government for implementing the Montreal Protocol. As early as the late 1980s, our team worked as a counselor for the government before acceding to the Vienna Convention for the

Protection of the Ozone Layer. In the 1990s, we discovered that, from the perspective of atmospheric chemistry and physics, there were climate and environmental side effects as a result of using HFC-134a as a substitute. In 1993, China's Country Program for the phase-out of Ozone-Depleting Substances (ODS), spearheaded by our team, was adopted by the United Nations Environment Program and the Chinese government. It was the first national program of the Montreal Protocol and became a model for other countries as they developed their own national programs. Moreover, we proposed and constructed an innovative sector-based ODS phaseout plan which was proved to be cost-effective. According to the assessment by World Bank, it helped reduce 27% of the cost and was recommended by the Multilateral Fund as a mechanism suitable for other countries. The ODS quota trading system, designed by the team, was adopted by the Chinese government as one of the earliest examples of market based instruments for pollution control in China.

The protection of the ozone layer is closely related to climate change. In the past decades, our research team has studied the

climate impact of man-made chemicals, using discharging and monitoring methods from bottom-up and top-down to analyze the sources of CFCs, HCFCs and HFCs, together with their impacts. The study showed that in the effort to protect the ozone layer, China has reduced large amounts of halogenated hydrocarbon greenhouse gases and alleviated climate warming together with the international community. Our suggestions on the phase-out plan of CFCs and HCFCs were adopted and implemented by the Chinese government, and in just the year 2010, the net annual avoided emissions of halogenated hydrocarbons greenhouse gases was equivalent to 1 billion tons of carbon dioxide. An article, published by the team in early 2016, pointed out that China, whose HFCs were rapidly increasing, would play an unprecedentedly important role in the global reduction of HFCs. The article proposed a HFCs control schedule in China, this helps the Chinese government to sign the Kigali amendment to the Montreal Protocol in October 2016.

Solving the global environmental issues is contingent on social, economic, scientific knowledge and technological alternatives. The PKU team will continue its research on the fundamental issues and implementation strategies of international conventions on climate change, ozone layer depletion and chemicals (persistent organic pollutants and mercury). The research will involve the risk identification and assessment of these issues (relevant substances or contaminants), paying more attention to risk management as related to both development and protection at the national, regional and global levels.

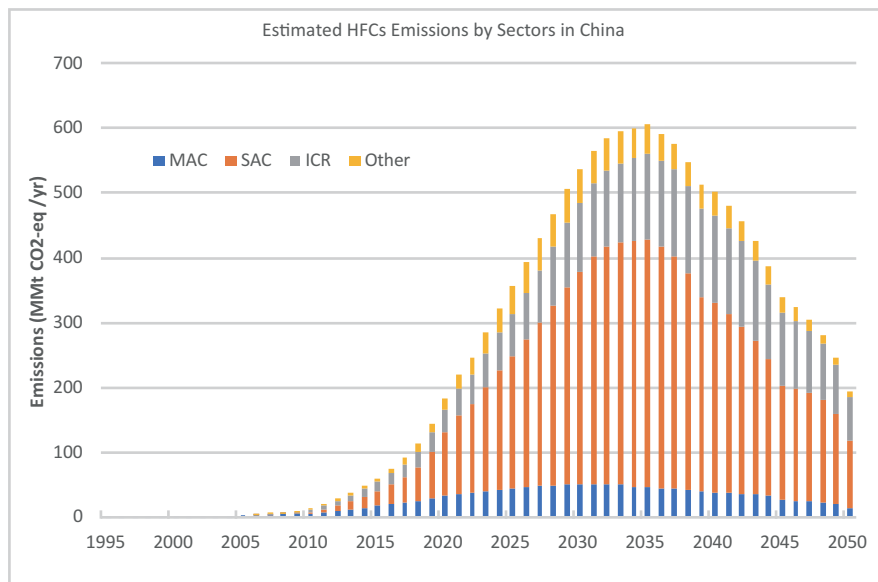
CESE cordially welcomes job applicants and visiting scholars with expertise in related areas such as environmental management and implementation of international conventions. Feel free to contact us:

Website: <http://cese.pku.edu.cn>

Email: huiliu@pku.edu.cn

Tel: 010-62754126; Fax: 010-62751480

Address: Environmental Building, Peking University, No.5 Yiheyuan Road, Beijing, China. 100871





Nanchang University Welcomes Young Talents at Home and Abroad to Attend The First International Young Scholar Forum

I. Introduction to the University and Forum

Nanchang University is located in Nanchang City, a famous historical and cultural city – the "Hero City" in China, which has a history of nearly one hundred years. It is a national "211 Project" key construction university, China's "Double First-Class" construction of first-rate disciplines and the only university in Jiangxi Province with high-level overall construction, and is one of the 14 universities in the central and western regions that are supported by the Ministry of Education in a construction mode of ministerial-provincial cooperation.

The university has five campuses, namely Qianhu Lake, Qingshan Lake, Donghu Lake, Poyang Lake, and Fuzhou. The main campus - Qianhu Lake covers an area of 4,500 mu and the building area of the school building is 1,300,000 square meters.

The university now has three departments, including humanities and social sciences, science and engineering, and medical science. There are two national key disciplines (food science, materials physics and chemistry) and one national key (cultivation) discipline (material processing engineering). There are currently a national key laboratory, a national engineering technology research center, a key research base of humanities and social sciences of the Ministry of Education, two key laboratories of the Ministry of Education, three engineering research centers of the Ministry of Education, and nine collaborative innovation centers of Jiangxi Province. There are authorized first-level disciplines for 15 academic doctoral degrees, 11 post-doctoral research mobile stations, and the university has 5 affiliated hospitals. Five disciplines including chemistry, clinical medicine, agricultural science (based mainly on food science and engineering), engineering, and materials science entered the top 1% of ESI in the United States. The discipline of food science and engineering was evaluated as A in the fourth round of China Discipline Ranking, ranking third in universities nationwide.

The First International Young Scholar Forum of Nanchang University will be held on May 15-17. This forum aims to build a platform for exchanges and cooperation among outstanding young scholars at home and abroad, promote interdisciplinary research and academic innovation, increase mutual cooperation and trust, and deepen the comprehensive understanding of Nanchang University, and realize promotion of high-level talents to build up their careers in Jiangxi Province and work together to jointly promote the "Double First-Class" construction of Nanchang University. We sincerely welcome young talents at home and abroad to participate in this forum.

II. Subject Areas

Materials Science and Engineering, Physics, Chemistry, Biomedical Engineering, Information and Communication Engineering, Mechanical Engineering, Food Science and Engineering, Public Health and Preventive Medicine, Bioengineering, Environmental Science and Engineering, Chemical Engineering and Technology, Biology, Clinical Medicine, Basic Medicine, Pharmacy, Journalism, Marxist Theory, Applied Economics, Theoretical Economics, Business Administration, Statistics, Chinese Language and Literature, Philosophy, Chinese History, Computer Science and Technology.

III. Invitees

The invitees shall be the outstanding young scholars who generally are under 40 years of age (born after January 1, 1978), have a PhD degree overseas with continuous overseas research work experience for 2 years or more, or have obtained a doctorate in China with more than 3 consecutive years of research work experience abroad but has not yet returned to China and intends to join Nanchang University.

Invitees should have one of the following conditions: Appointee of National "1000 Youth Talents Plan", Young Top-notch Talent of Ten Thousand Talent Program, Young Changjiang Scholars, win-

ners of the National Natural Science Foundation for Outstanding Young Scientists, or qualified personnel of corresponding levels.

IV. Sponsorship

The university uniformly arranges accommodation (free of charge) and offers a round-trip travel allowance (economy class air-ticket or second class seat high-speed rail ticket), with an upper limit of 15,000 yuan per person in Europe and the United States, a maximum of 7,000 yuan per person in the Asia-Pacific region, and a domestic (including Hong Kong, Macao and Taiwan) limit of 5,000 yuan per person.

V. Schedule

Registration Deadline: April 20, 2018 (determined according to the circumstances)

Invitation to the forum: March 20 to April 30, 2018

Forum Registration: May 15, 2018

Forum Duration: May 16-17 (Main Forum, Sub Forum, etc.)

Contact

Contact: Teacher Tian

Tel: 0791-83969076

E-mail: 373959571@qq.com
gcr@ncu.edu.cn

Application Materials

1. Information of the applicant: including but not limited to resume, academic achievements, award certificate, etc.;
2. Personal information (within 200 words);
3. Personal photos (within 100k).

Treatment to National "Four Categories of Youth Talents"

1. National support (for Appointee of National "1000 Youth Talents Plan"): The research fund is 1-3 million yuan with the one-off subsidy of 500,000 yuan.

2. Local support: Jiangxi Province provides 5 million yuan of innovation and entrepreneurship development funds (for Appointee of National "1000 Youth Talents Plan"). Those who are selected as the "Double 1000

Legal Scholars" in Jiangxi Province will be given a project grant of 1 million to 3 million yuan per person. Humanities and social sciences is funded by 200,000-500,000 yuan as the project grant, and no more than 30% of the funds can be used to improve personal living conditions.

3. The university provides:

1) According to the "Interim Measures for the Introduction of High-level Talents" issued by the university and in line with the standards for introduction of subject-leading talents, the university will provide an annual salary of 500,000-800,000 yuan;

2) With the appointment to professorship (staffing of government affiliated institutions, and fixed position), the university will provide a good scientific research platform, office space and research space;

3) Based on the funding for the start-up of scientific research provided by the State and Jiangxi Province, the university will provide sufficient scientific research start-up funds to meet the needs of the discipline construction according to the circumstances, with details to be discussed;

4) The university will provide a set of circulating house within the campus together with a setting-in allowance of 600,000 yuan and a talent reward of 500,000 yuan;

5) The university will provide one-stop high-quality educational resources for the talents' children from primary school to high school.

** Young scholars who achieve a better evaluation in the national "Four Categories of Youth Talents" project review, pass the expert communication review and enter the meeting's assessment section may be introduced according to the university's "Interim Measures for the Introduction of High-level Talents". The university will provide an annual salary of 250,000-500,000 yuan and a setting-in allowance with necessary research funding according to relevant standards.*



The International Center of Future Science (ICFS) was established in 2016 on the basis of the International Talent Recruitment Center, which was co-founded by the State Administration of Foreign Experts Affairs of China and Jilin University. Focused on the newest development of international science and technology and oriented to the major demands of global development and expansion, ICFS is dedicated to the great mission of exploring the unknown and promoting human cognition. With the aim to bring out epoch-marking academic thoughts and new outstanding interdisciplinary results, ICFS has been established to carry out scientific research in the highest standards by gathering teams of top-ranking scientists from all over the world in the fields of chemistry, physics, materials, electronics, biomedicine, etc. As a key talent zone where scientists have unique opportunities for their research, ICFS enjoys preferential policy support and excellent development platform from both the Chinese government and Jilin University. Currently, ICFS has six research teams working on different research areas, including Advanced Energy and Environmental Materials, High-Pressure Science and Technology, Translational Immunology, Millimeter

Microwave Technology and Applications, Advanced Metallic Materials, PaleoChemistry and PaleoEvolution of Ancient Life.

Open Positions and Requirements

Position 1: Distinguished Scientists

Distinguished Scientists should be (1) the members of Chinese, European and American academy of sciences and engineering, or (2) the recipients of Innovative Talents Long-Term Project of National Thousand Talents Program, or (3) the recipients of Changjiang Scholars Program of Ministry of Education of China, or (4) the recipients of the National Science Fund for Distinguished Young Scholars, or (5) the recipients of the tenured professors of world leading universities and research institutions.

Position 2: Future Scientists

The Future Scientist should be the recipients of (1) National Thousand Youth Talents Program, or (2) National Program for Support of Top-notch Young Professionals, or (3) Excellent Young Scientists Fund, or (4) Changjiang Youth Scholars Program of Ministry of Education of China.

Position 3: Post-Doctoral Research Fellow

The applicants should hold a doctoral degree granted by world-renowned universities or research institutions. In principle, the

applicants should be under 35 years old.

Salary and Benefit

ICFS provides competitive annual salary ranging from \$125,000 to \$360,000 for distinguished scientists and \$55,000~\$180,000 for future scientists. Housing benefits, insurance, medical treatment and welfare are in accordance with the relevant policies of the government and Jilin University. Initiation funds, laboratories, and research assistants will be provided according to the demand of the corresponding research field. The annual salary for post doctorate researchers would be \$45,000~\$60,000.

Distinguished Scientists and Future Scientists will enjoy beneficial policies for graduate student admissions. ICFS will help the spouses of Distinguished Scientists and Future Scientists find their jobs in Jilin University. Their children will get the entrance into the kindergarten, primary school, middle school, and high school affiliated to Jilin University.

Contact Us

Please send your Curriculum Vitae, 5 representative publications, and your research plan to the chief scientist of the research team you are interested in.

Research Team	Background Required	Chief Scientist
Advanced Energy and Environmental Materials	chemistry, physics, materials, electronics, computer science, etc.	Prof. Jihong Yu (jihong@jlu.edu.cn)
High-Pressure Science and Technology	physics, chemistry, materials, electronics, computer science, etc.	Prof. Yanming Ma (mym@jlu.edu.cn)
Translational Immunology	medicine, molecular biology, immunology, nanomedicine, stem cells, pharmacy, animal model development, etc.	Prof. Yongguang Yang (yonggg@jlu.edu.cn)
Advanced Metallic Materials	materials, physics, chemistry, etc.	Prof. Huiyuan Wang (wanghuiyuan@jlu.edu.cn)
PaleoChemistry and PaleoEvolution of Ancient Life	chemistry, physics, materials, electronics, computer science and graphics, earth sciences, life sciences, paleobiology, etc.	Prof. Robert R. Reisz (robert.reisz@utoronto.ca); Prof. Timothy D. Huang (timd_huang@yahoo.com)

UPMC HILLMAN CANCER CENTER

Affiliated with the University of Pittsburgh School of Medicine

Associate Director for Clinical Investigations, UPMC Hillman Cancer Center, Pittsburgh, PA

The UPMC Hillman Cancer Center, celebrating its 27th year as a leading center for cancer research, is recruiting outstanding applicants for the position of **Associate Director for Clinical Investigations**. The UPMC Hillman Cancer Center clinical research efforts located in Shadyside as well as the expanding network, including nearly 150 employees in various regulatory clinical research and leadership roles. The Associate Director for Clinical Investigations will provide oversight and support for re-designing and re-organizing our infrastructure for clinical research, in order to increase efficiency, reduce activation timelines and expand clinical trial availability to the many point-of-care sites in the UPMC Hillman Cancer Center network. As Dr. Ferris's representative, the AD will also provide physician leadership to Clinical Research Services in collaboration with UPMC Hillman Cancer Center Chief Operating Officer and Vice President Stephanie K. Dutton, MPA and Bhanu Pappu, PhD, MHA, Vice President for Clinical Research Operations and Strategy. Based on the individual's research field, the candidate will be hired into the appropriate Department in the School of Medicine at the University of Pittsburgh. The rank of Associate or Full Professor will be based on the candidate's experience.

Successful candidates will have an exceptional clinical, scientific research and administrative record and will join in tenure-track or tenured faculty positions that are commensurate with prior training and experience. A competitive salary and research start-up package will be provided, as well as potential for laboratory and office space within the state-of-the-art Hillman Cancer Center or Magee-Womens Research Institute.

Located in the city of Pittsburgh (routinely ranked as one of the top most livable and affordable U.S. cities), Hillman (previously known as the University of Pittsburgh Cancer Institute) is an NCI-designated Comprehensive Cancer Center with 344 members; 10 research programs in basic, translational, clinical, and population sciences; 13 shared resources that receive funding from our NCI Cancer Center Support Grant; and an FY17 institutional funding base of nearly \$157 million. In FY17, the University of Pittsburgh ranked #5 in overall NIH funding. Hillman Cancer Center serves a catchment area of **29 Western Pennsylvania counties** and provides unique opportunities to collaborate with clinical and translational research programs involved in cancer patient care.

To apply for a position, please send your *curriculum vitae*, a one-page summary of your research plans (together with recommendations) to Hillman Director Robert L. Ferris, MD, PhD, care of thompsonla3@upmc.edu. Applications will be reviewed and evaluated on an ongoing basis, following the receipt of all required materials. *The University of Pittsburgh is an Affirmative Action, Equal Opportunity Employer. EEO/AA/M/F/Vets/Disabled.*

Robert L. Ferris, MD, PhD, Director, UPMC Hillman Cancer Center
c/o Lola Thompson, 5150 Centre Avenue, Suite 500
Pittsburgh, PA 15232



Science Careers

FROM THE JOURNAL SCIENCE AAAS

Follow us for jobs,
career advice
and more!



@ScienceCareers



/ScienceCareers



Science Careers

ScienceCareers.org



Professor of Genetics and Director of the Center for Genomic Health

Department of Genetics,
Yale University School of Medicine
Yale New Haven Hospital

The Department of Genetics at Yale University School of Medicine is searching for a Professor of Genetics with an outstanding record of transformative scientific achievements in Human Genetics and Genomics. We expect that the candidate will lead a vigorous cross-disciplinary research program focused on identifying and characterizing genetic drivers of human disease. As a leader of human genetics both within the department and across the Yale School of Medicine, the successful candidate will have the opportunity to recruit other human geneticists to the Genetics Department and lead a new program in precision medicine as the Scientific Director of the Yale Center for Personalized Medicine and Genomic Health.

We are looking for a dynamic, internationally recognized scientist (Ph.D., M.D., or M.D./Ph.D.) with an outstanding research record of scientific discoveries, as well as a strong track record of training the best innovators in the field of human genetics and genomics.

To apply, please submit your CV to <http://apply.interfolio.com/45539> to the attention of Antonio Giraldez, Chair of Genetics. Applications will be reviewed starting April 2018 and will continue until the position is filled. Inquiries should be addressed to neltja.brewster@yale.edu.

Yale University is an Affirmative Action/Equal Opportunity Employer. Yale values diversity among its students, staff, and faculty and strongly welcomes applications from women, persons with disabilities, protected veterans, and underrepresented minorities.

Search more jobs online

Access hundreds of job postings
on ScienceCareers.org.

Expand your search today.



Science
Careers

AAAS

JOB FOCUS: IMMUNOLOGY

moderna

Multiple Immunology Research Openings

Moderna Therapeutics
Cambridge, MA

We have multiple Immunology Research opportunities available to join Moderna's team! We look forward to hearing from you!

Our Mission: Deliver on the promise of mRNA science to create a new generation of transformative medicines for patients.

www.modernatx.com/careers



Apply for EIT Health Summer Schools across Europe. Train to become a healthcare leader of tomorrow

The EIT Health Summer Schools support innovation by empowering the healthcare leaders and innovators of tomorrow. The two-to-three week programmes provide students with the necessary skills to identify needs related to healthcare and to transform innovative ideas into well-elaborated business cases that can attract investment.

From the role of artificial intelligence to the application of serious game design in healthcare – The EIT Health Summer Schools address the most promising game changers of today.

For more information on the dates and programmes, please visit:
<https://www.eithealth.eu/summer-schools>



JOB FOCUS: IMMUNOLOGY



DEPARTMENT OF IMMUNOLOGY
UNIVERSITY of WASHINGTON

ASSISTANT, ASSOCIATE or FULL PROFESSOR FACULTY POSITION

The Department of Immunology at the University of Washington seeks a highly qualified applicant for a fulltime tenure-track or tenured faculty position. Candidates at all levels will be considered although candidates at the Associate Professor or Full Professor level are particularly encouraged to apply. Candidates for this position must hold a PhD and/or MD (or foreign equivalent) degree in immunology or related discipline and have a strong record of published research in immunology. The University of Washington faculty engage in teaching, research and service. The successful candidate will be expected to teach at both the undergraduate and graduate level and lead a strong research program. The Department of Immunology offers excellent laboratory space, access to cutting-edge technologies and a highly collaborative environment. Additional information regarding the department can be found at <http://immunology.washington.edu/>.

This position will remain open until filled. Please submit an application (including a cover letter addressed to Dr. Joan Goverman, Professor and Chair, Department of Immunology and your curriculum vitae, a brief description of proposed research, as well as names and addresses of three references) at: <http://apply.interfolio.com/49647>

University of Washington is an Affirmative Action and Equal Opportunity Employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, gender expression, national origin, age, protected veteran or disabled status, or genetic information.

PRIZES

3rd Hideyo Noguchi Africa Prize

Calling for Nominations

The Prize, to be awarded by the Government of Japan on the occasion of 2019 Tokyo International Conference on African Development (TICAD) in memory of a Japanese microbiologist Dr. Hideyo NOGUCHI, aims to honor individuals with outstanding achievements in the fields of medical research and medical services to combat infectious and other diseases in Africa, thus contributing to the health and welfare of the African people and of all humankind.

Medical Research

- Basic medical research
- Clinical medical research
- Research in all fields of life science closely related to medicine

Medical Services

- Field-level medical/public health activities to combat diseases and advance public health

Please access at:

<https://www.jsps.go.jp/english/e-noguchiafrica/index.html>

DEADLINE: July 31st, 2018

Laureates for the First and Second Hideyo Noguchi Africa Prize



2008
Dr. Brian Greenwood
(Medical Research)



2008
Dr. Miriam Were
(Medical Services)



2013
Dr. Peter Piot
(Medical Research)



2013
Dr. Alex G. Coutinho
(Medical Services)

My path to contentment

I leapt into the air, screaming at the top of my lungs with tears rolling down my cheeks as the news sank in. I had lost both of my parents when I was 16 years old, and I had often been sent home from school for unpaid tuition as I worked my way to a bachelor's degree in my home country of Zimbabwe. But now, I was a Fulbright fellow. I was convinced that the award would propel my career to unconceivable heights. It was all the sweeter when I thought of my mother and how she used to cry over my report cards. At the time, I thought it was because I had not done well enough, but I later realized she was crying because she could not bear the idea that her poverty would keep me from reaching my full potential. I carried the burden of wanting to do her proud, and the fellowship was a huge step in that direction.

It would also help me prove to the world that I was more than my family's poverty. The numerous first-time opportunities the fellowship afforded—flying on a plane, staying in a hotel, moving to the United States—earned me respect in my small farming town. The chance to study and work abroad raised my own expectations sky-high as well.

But as I completed my Ph.D. about 5 years later, it became clear that, even with a Fulbright fellowship, I would not achieve all I had dreamed of. A degree from a solid but not world-renowned university and publications in journals with middling impact factors were not enough to secure the prestigious postdoc I thought I needed to achieve my long-term goals: opening my own lab and securing tenure. So, on 4 July 2016, while the rest of America celebrated its independence, I took a flight back to Zimbabwe—jobless, dejected, and hopeless.

Back home, people respected me. A bank teller insisted on putting “Dr.” on my ATM card. At community gatherings, elderly people offered me their seats when they learned I had a Ph.D. Yet, as I continued to unsuccessfully pursue a postdoc at a top-notch institution, I was haunted by the feeling that I was a failure.

I was eventually offered a postdoc position at a university in China—but it was not at a top school like Peking, Tsinghua, or Fudan University, so I ignored it. About 2 weeks later, I was invited to an onsite interview at ETH Zürich in Switzerland. Finally, I had an opportunity to work with famous researchers at a world-class university! I was elated. But I didn't get the job.

With that, my desire to be respected and valued by top



“I resolved that what mattered most was my commitment and diligence.”

researchers died. I was done trying to join the elite. After all, I couldn't change the grad school I attended or the ranking of the journals in which I had published. And I remembered the words of my mental health counselor in grad school, when the stress of writing a dissertation, job hunting, and trying to be there for my young family had driven me into depression: “Edmond, you do not need documented validation for you to know your worth.”

At that time, the advice didn't make sense. After all, I needed good publications to graduate. I needed better publications to get a postdoc and ultimately a tenured position. But now, it finally sank in. The rat race had to stop. I resolved that what mattered most

was my commitment and diligence rather than what others thought of my scientific contributions. I could do great science at a small, unknown university. So I decided to take the position in China.

I've been here a year now. Navigating the language and cultural barriers has been an enjoyable adventure. Focusing on what excites me rather than trying to fulfill the expectations of academia has been liberating. And whenever I find myself slipping back into old ways of thinking, I remember my wife's question: “How many orphaned kids from an unknown farming town graduated from high school and have an undergrad degree or a Ph.D.?” With that, I am proud of what I have accomplished, and that is enough. ■

Edmond Sanganyado is a postdoctoral fellow at Shantou University in China. Do you have an interesting career story? Send it to SciCareerEditor@aaas.org.