In Taiwan's mountains, a laboratory for landslides p. 938

Tomorrow's Earth pp. 950 & 969

Enhanced resolution of chromatin wins *Science* & SciLifeLab Prize p. 964

Science

$15
22 NOVEMBER 2019
sciencemag.org

AAAS

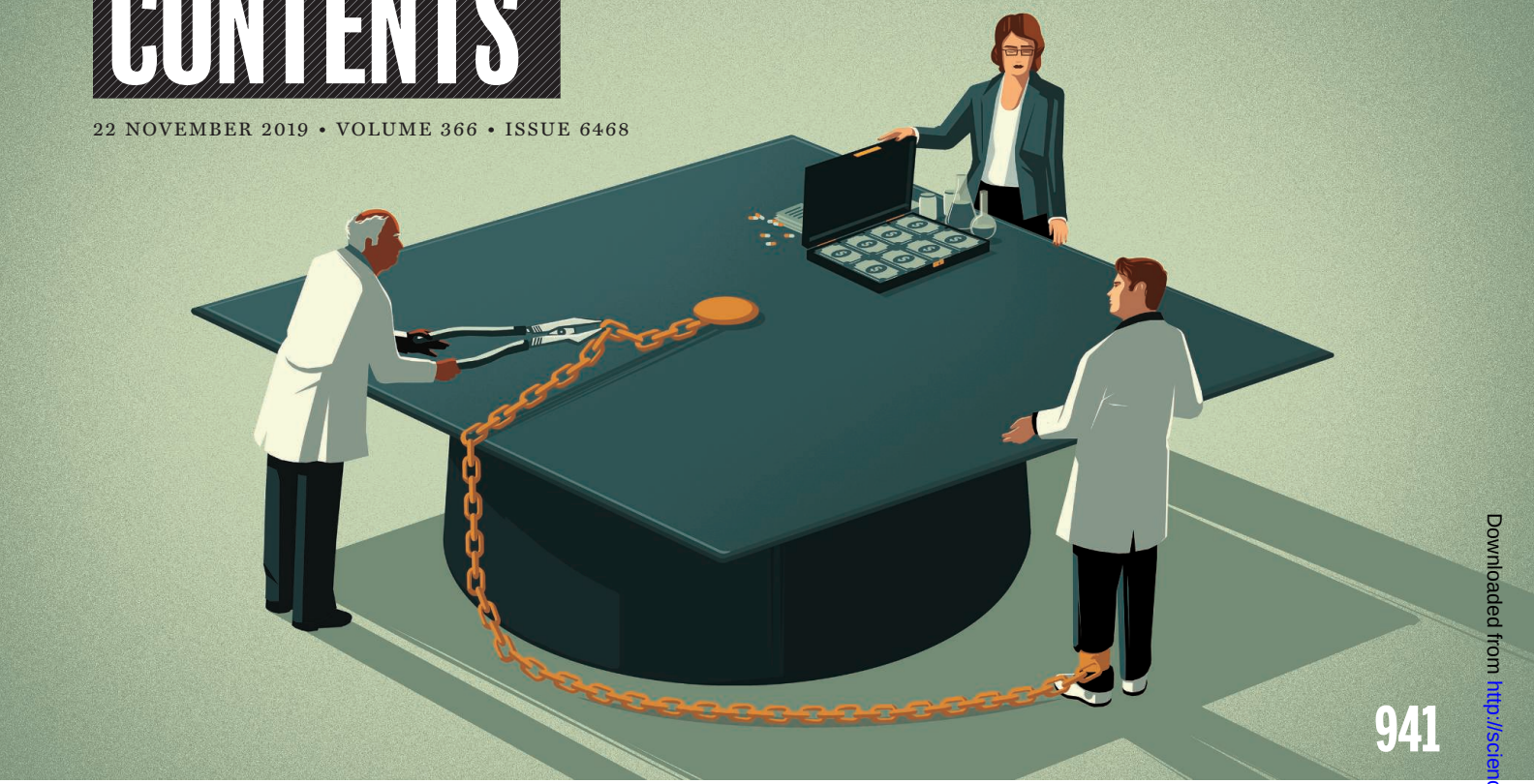
BEETLE HORNS

Beetle thoracic horns share a developmental origin with wings pp. 946 & 1004



CONTENTS

22 NOVEMBER 2019 • VOLUME 366 • ISSUE 6468



941

Downloaded from http://science.sciencemag.org/ on November 28, 2019

NEWS

IN BRIEF

930 News at a glance

IN DEPTH

932 Bold plan to retrieve Mars samples takes shape
NASA and Europe settle on suite of missions to bring back rocks from the Red Planet
By D. Clery and P. Voosen

934 After 20 years, golden rice nears approval
Bangladesh may become the first country to adopt transgenic rice enriched in vitamin A *By E. Stokstad*

935 Hong Kong protests hit universities
Clashes have had little impact on research, but long-term effects on schools are unclear
By D. Normile

936 Researchers sound alarm on European data law
Privacy regulations impede international studies on diabetes and Alzheimer's disease
By T. Rabesandratana

937 Australian blazes will 'reframe our understanding of bushfire'
Fire scientist David Bowman discusses the unfolding crisis
By J. Pickrell

FEATURES

938 The landslide laboratory
Landslides aren't just hazards. An experiment in Taiwan could reveal their outsize effect on the climate *By K. Kornei*
PODCAST

941 Double dip
An NIH program pays school debt to keep scientists in academia. Many break the rules by also taking industry money *By C. Piller*

INSIGHTS

PERSPECTIVES

944 The world in a song
Similar songs are used in similar contexts around the world *By W. T. Fitch and T. Popescu*
RESEARCH ARTICLE p. 970

946 The multistep morphing of beetle horns
Genes that specify insect wings initiate horn development in dung beetles *By H. F. Nijhout*
REPORT p. 1004

947 Compelled to drink: Why some cannot stop
Individual differences in aversive stimulus processing may explain compulsive alcohol drinking
By K. Nixon and R. A. Mangieri
REPORT p. 1008

948 A catalytic one-two punch
Cooperative catalysis enables the controlled synthesis of unnatural amino acids *By J. M. Ovian and E. N. Jacobsen*
REPORT p. 990

950 Unraveling the claims for (and against) green growth
Can the global economy grow indefinitely, decoupled from Earth's limitations?
By T. Jackson and P. A. Victor
REVIEW p. 969

952 Designer sense-response systems
Artificial chemically induced protein dimerization yields biological sensor-actuator devices
By R. A. Chica
REPORT p. 1024

953 Turning lateral roots into nodules
The evolutionary origin of legume root nodules that help them grow is revealed
By A. Bishopp and M. J. Bennett
REPORT p. 1021

ILLUSTRATION: STEPHAN SCHMITZ/FOLIO ART

Published by AAAS

955 Artificial intelligence for global health
Socially responsible technologies promise to help address health care inequalities
By A. Hosny and H. J. W. L. Aerts

POLICY FORUM

957 How low should we go when warning for earthquakes?
Social responses to alerts are critical but understudied *By E. S. Cochran and A. L. Husker*

BOOKS *ET AL.*

959 Excavating Uruk
A richly illustrated volume reveals what life was like in the world's first city *By A. Robinson*

960 Trash, treasure, and everything in between
A journalist traces the evolution of the large, but largely unknown, global secondhand market *By K. Dooley*

LETTERS

961 EU Court: Science must justify future hunting
By Y. Epstein et al.

961 Time to ban lead hunting ammunition
By C. Sonne et al.

962 Beyond meat: Ecological functions of livestock
By V. Proença and C. M. G. L. Teixeira

PRIZE ESSAY

964 Three-dimensional genome structure of a single cell
Chromosome organization in diploid cells reveals a structural basis for smell
By L. Tan

RESEARCH

IN BRIEF

966 From *Science* and other journals

REVIEW

969 Nanomaterials
Energy storage: The future enabled by nanomaterials *E. Pomerantseva* et al.
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAN8285
PERSPECTIVE p. 950

RESEARCH ARTICLES

970 Psychology of music
Universality and diversity in human song
S. A. Mehr et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAX0868
PERSPECTIVE p. 944; PODCAST

971 Signal transduction
Structural mechanism of a Rag GTPase activation checkpoint by the lysosomal folliculin complex
R. E. Lawrence et al.

977 Paleontology
Exceptional continental record of biotic recovery after the Cretaceous–Paleogene mass extinction *T. R. Lyson* et al.

REPORTS

983 2D magnetism
Direct observation of van der Waals stacking–dependent interlayer magnetism
W. Chen et al.

987 Correlated electrons
Strange metallicity in the doped Hubbard model *E. W. Huang* et al.

990 Organic chemistry
Highly enantioselective carbene insertion into N–H bonds of aliphatic amines *M.-L. Li* et al.
PERSPECTIVE p. 948

995 Infection
Host resistance factor SLC11A1 restricts *Salmonella* growth through magnesium deprivation *O. Cunrath and D. Bumann*

999 Computer science
Preventing undesirable behavior of intelligent machines *P. S. Thomas* et al.

1004 Trait development
Beetle horns evolved from wing serial homologs *Y. Hu* et al.
PERSPECTIVE p. 946

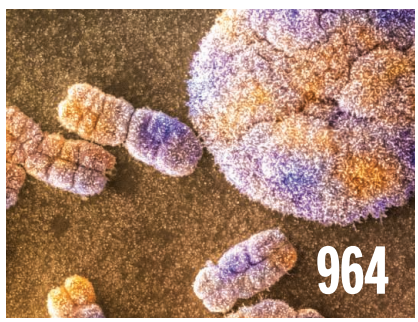
1008 Neuroscience
A cortical-brainstem circuit predicts and governs compulsive alcohol drinking
C. A. Siciliano et al.
PERSPECTIVE p. 947

1013 Immunometabolism
Glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion
R. D. Leone et al.

1021 Plant science
A shared gene drives lateral root development and root nodule symbiosis pathways in *Lotus T. Soyano* et al.
PERSPECTIVE p. 953

1024 Protein design
Computational design of a modular protein sense-response system *A. A. Glasgow* et al.
PERSPECTIVE p. 952

1029 Cancer
Peritumoral activation of the Hippo pathway effectors YAP and TAZ suppresses liver cancer in mice *I. M. Moya* et al.



DEPARTMENTS

929 Editorial
The context of diversity *By Cato T. Laurencin*

1042 Working Life
Sidestepping politics to teach climate
By Ragupathy Kannan

ON THE COVER



Onthophagus sagittarius dung beetle with prominent horns on its head and thorax. The thoracic horn, an evolutionary novelty whose origin has been a long-standing mystery, is derived from the same serially reiterated tissues that gave rise to wings in the more posterior thoracic segments. By textbook definition, evolutionary novelty requires the absence of homology, yet thoracic horns, considered a classic example of this phenomenon, appear to have originated by relying deeply on their homology to wings. See pages 946 and 1004. *Photo: Alexander Wild*

Science Staff 928
New Products 1035
Science Careers 1036

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 © 2019 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): $1971; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

IMAGE: POWER AND SYRED/SCIENCE SOURCE

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

Science AAAS

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Monica M. Bradford

Editors, Research Valda Vinson, Jake S. Yeston **Editor, Insights** Lisa D. Chong

DEPUTY EDITORS Julia Fahrenkamp-Uppenbrink (UK), Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, 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), Valerie B. Thompson, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon (UK), Keith T. Smith (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Julia Katris, Nida Masiulis, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Alana Warnke, Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronmel Navas, Hilary Stewart (UK), Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (UK)

News Editor Tim Appenzeller

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

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Xing Liu **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR PHOTO EDITOR** Emily Petersen

Interim Chief Executive Officer and Executive Publisher Alan Leshner

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman, Meron Kebede **FINANCIAL ANALYST** Alexander Lee **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SENIOR SALES COORDINATOR** Shirley Young **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Jacob Hedrick, Antoinette Hodal, Lori Murphy **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Tori Velasquez, Mike Romano, Ashley Hylton **DIGITAL MARKETING SPECIALIST** Asleigh Rojanavongse **SENIOR DESIGNER** Kim Huynh **TRADE SHOW AND MEETINGS ASSOCIATE** Andrew Clamp

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

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGMENT** Kris Bishop **ASSOCIATE DIRECTOR, PRODUCT DEVELOPMENT AND SPJ** Hannah Heckner **SR. PRODUCT ASSOCIATE** Robert Koepke **DIGITAL PRODUCT STRATEGIST** Michael Hardesty **SPJ ASSOCIATE** Samantha Bruno Fuller

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **MARKETING MANAGER** Kess Knight **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan **MANAGER, SYSTEMS AND OPERATIONS** Brian Holiahn **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor, Glen Cox **US WEST COAST SALES** Lynne Stickrod **US SALES MANAGER, SCIENCE CAREERS** Claudia Paulsen-Young **US SALES REP, SCIENCE CAREERS** Tracy Anderson **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Bryony Cousins **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **PROJECT MANAGER** Kilo Lan **ASCA CORPORATION, JAPAN** Kaoru Sasaki (Tokyo), Miyuki Tani (Osaka) **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **CONTRACTS AND LICENSING ASSOCIATE** Lili Catlett

---

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

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

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

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

EDITORIAL
science_editors@aaas.org

NEWS
science_news@aaas.org

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

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

MEDIA CONTACTS
scipak@aaas.org

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

INSTITUTIONAL SALES AND SITE LICENSES
sciencemag.org/librarian

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

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

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

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS
aaas.org/membercentral

AAAS BOARD OF DIRECTORS
CHAIR Margaret A. Hamburg
PRESIDENT Steven Chu
PRESIDENT-ELECT Claire M. Fraser
TREASURER Carolyn N. Ainslie
INTERIM CHIEF EXECUTIVE OFFICER Alan Leshner
BOARD Cynthia M. Beall, May R. Berenbaum, Rosina M. Bierbaum, Ann Bostrom, Stephen P. A. Fodor, S. James Gates, Jr., Laura H. Greene, Kaye Husbands Fealing, Maria Klawe, Robert B. Millard, William D. Provine

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

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

Adriano Aguzzi, *U. Hospital Zürich*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Foundation*
Judith Allen, *U. of Manchester*
Sebastian Amigorena, *Institut Curie*
James Analytis, *U. of California, Berkeley*
Paola Arlotta, *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*
Peter Bearman, *Columbia U.*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Gabriele Bergers, *VIB*
Bradley Bernstein, *Mass. General Hospital*
Alessandra Biffi, *Harvard Med. School*
Peer Bork, *EMBL*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *U. of California, Santa Cruz*
Ron Brookmeyer, *U. of California, Los Angeles* **(S)**
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Research*
Carter Tribley Butts, *U. of California, Irvine*
György Buzsáki, *New York U. School of Med.*
Blanche Capel, *Duke U.*
Annmarie Carlton, *U. of California, Irvine*
Lars-Erik Cederman, *ETH Zürich*
Nick Chater, *U. of Warwick*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Alan Cowman, *Walter & Eliza Hall Inst.*
Carolyn Coyne, *U. of Pittsburgh*
Roberta Croce, *VU Amsterdam*
Jeff L. Dangl, *U. of North Carolina*
Tom Daniel, *U. of Washington*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Frans de Waal, *Emory U.*
Claude Desplan, *New York U.*
Sandra Díaz, *Universidad Nacional de Córdoba*
Hong Ding, *Inst. of Physics, CAS*
Jennifer Dionne, *Stanford U.*
Dennis Discher, *U. of Penn.*
Jennifer A. Doudna, *U. of California, Berkeley*
Bruce Dunn, *U. of California, Los Angeles*
William Dunphy, *Caltech*
Christopher Dye, *U. of Oxford*
Todd Ehlers, *U. of Tübingen*
Jennifer Elisseeff, *Johns Hopkins U.*
Tim Elston, *U. of North Carolina*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *U. of Penn.*
Karen Ersche, *U. of Cambridge*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Michael Feuer, *The George Washington U.*
Toren Finkel, *U. of Pittsburgh Med. Ctr.*
Gwenn Flowers, *Simon Fraser U.*
Peter Fratzl, *Max Planck Inst. Potsdam*
Elaine Fuchs, *Rockefeller U.*
Eileen Furlong, *EMBL*
Jay Gallagher, *U. of Wisconsin*
Susan Gelman, *U. of Michigan*
Daniel Geschwind, *U. of California, Los Angeles*
Karl-Heinz Glassmeier, *TU Braunschweig*
Ramon Gonzalez, *U. of South Florida*
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, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. College London*
Jian He, *Clemson U.*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Hans Hilgenkamp, *U. of Twente*
Kai-Uwe Hinrichs, *U. of Bremen*
David Hodell, *U. of Cambridge*
Lora Hooper, *UT Southwestern Med. Ctr.*
Fred Hughson, *Princeton U.*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS and U. of Arizona*
Kai Johnsson, *EPFL*
Peter Jonas, *IST Austria*
Matt Kaeberlein, *U. of Washington*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *U. of California, Berkeley*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med. School*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alexander Kolodkin, *Johns Hopkins U.*
Thomas Langer, *U. of Cologne*
Mitchell A. Lazar, *U. of Penn.*
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, *Washington U. in St. Louis*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Fabienne Mackay, *U. of Melbourne*
Anne Magurran, *U. of St. Andrews*
Oscar Marín, *King's College London*
Charles Marshall, *U. of California, Berkeley*
Christopher Marx, *U. of Idaho*
Geraldine Masson, *CNRS*
C. Robertson McClung, *Dartmouth College*
Rodrigo Medellín, *U. of Mexico*
Graham Medley, *London School of Hygiene & Tropical Med.*
Jane Memmott, *U. of Bristol*
Edward Miguel, *U. of California, Berkeley*
Tom Misteli, *NCI, NIH*
Yasushi Miyashita, *U. of Tokyo*
Alison Motsinger-Reif, *NC State U.* **(S)**
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *U. of California, Berkeley*
Beatriz Noheda, *U. of Groningen*
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*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Samuel Pfaff, *Salk Inst. for Biological Studies*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Matthieu Piel, *Institut Curie*
Kathrin Plath, *U. of California, Los Angeles*
Martin Plenio, *Ulm U.*
Elvira Poloczanska, *Alfred-Wegener-Inst.*
Julia Pongratz, *Ludwig Maximilians U.*
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
David Randall, *Colorado State U.*
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, *Med. U. of Vienna*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *U. of California, Santa Barbara*
Rebecca Sear, *London School of Hygiene & Tropical Med.*
Vladimir Shalaev, *Purdue U.*
Jie Shan, *Cornell 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 Med.*
Lucia Sivilotti, *U. College London*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina* **(S)**
Mark Smyth, *QIMR Berghofer*
Pam Soltis, *U. of Florida*
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Science*
V. S. Subrahmanian, *U. of Maryland*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *U. of Cambridge*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Research*
Wim van der Putten, *Netherlands Inst. of Ecology*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Kathleen Vohs, *U. of Minnesota*
David Wallach, *Weizmann Inst. of Science*
Jane-Ling Wang, *U. of California, Davis* **(S)**
David Waxman, *Fudan U.*
Jonathan Weissman, *U. of California, San Francisco*
Chris Wikle, *U. of Missouri* **(S)**
Terrie Williams, *U. of California, Santa Cruz*
Ian A. Wilson, *Scripps Research* **(S)**
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *U. of Penn. School of Med.*
Jonathan Zehr, *U. of California, Santa Cruz*
Maria Zuber, *MIT*

Published by AAAS

EDITORIAL

The context of diversity

The term "diversity," which came about in connection with the passage of the U.S. Civil Rights Act of 1964, has been expanding to include an ever-growing list of identities—from race, gender, and sexual orientation to physical appearance, belief systems, thought styles, socioeconomic status, and rural/urban geographic location, among others. This is a welcome extension of representation, but this added texture has a downside—it threatens to muddle targets and obscure actions when achieving diversity is the goal. This consequence is particularly serious in the context of addressing equity for specific underrepresented racial and ethnic groups. Next week, the U.S. National Academies will convene the Roundtable on Black Men and Black Women in Science, Engineering and Medicine to focus on confronting issues that threaten the future of Blacks broadly in science. Forging systemic changes that bring Black diversity at all education and career levels will hopefully bring racial equity to practices in these fields and in doing so, expand the benefits of science, engineering, and medicine to society.

There are unintended negative consequences of the expanded definition of diversity. With so many groups, success in achieving diversity is increasingly measured in a pick-and-choose manner, where progress is defined through any lens that shows success. Also, with so many groups, diversity is often described through the lens of gender, leaving other groups as seemingly less important, or unimportant. And with so many groups, it has become easier for diversity efforts to disregard the historical and present drivers of discrimination that concepts of diversity began with. In other words, the greater context of inclusion and equity can get lost, making strides to diversify meaningless. The latter point is particularly relevant to Blacks in the United States who have experienced slavery, legally enforced segregation and discrimination, and now battle conscious and unconscious racism, and mass incarceration. Institutionalized racism, past and present, has resulted in the disregard, disrespect, and dismissal of Black people from all walks of life, and this is true in science, engineering, and medicine.

"Embracing the expanding definition of diversity is easy, but using the word with focus...for achieving diversity will take great attention."

These may be factors in the crisis-level changes seen across the academic landscape of Blacks in science, engineering, and medicine. For example, the number of Black males entering medical school between 2013 and 2014 in the United States was only 500, a historic low. Black men represented only 37.7% of Blacks entering medical school, which represented only 2.5% of all students entering medical school. This occurred during a historic increase in the number of medical schools in the nation. While this was happening at the trainee level, the U.S. National Academy of Medicine's most recent election in 2019 had no Black men in a class that recently increased by over 30% in size. Thus, there is a crisis taking place at all points in the medical educational and career spectrum for this particular group.

In response to this downward trend of Blacks in science and medicine, a number of individuals, including me, convened a U.S. National Academies workshop in 2017 that focused specifically on the growing absence of Black men in medicine in the United States. The ideas became a blueprint for actions that address not only Black men in medicine, but also the trajectory for Black women, and issues in engineering and science overall.

Embracing the expanding definition of diversity is easy, but using the word with focus so as not to weaken the paths for achieving diversity will take great attention. Next week, as leaders from academia, industry, government, foundations, and other nonprofits gather at the U.S. National Academy for this historic first meeting, the goal will be to begin to understand the barriers, explore opportunities, and develop actionable plans to increase the number of Blacks pursuing science, engineering, and medicine. The Roundtable will have a laser focus on racism and bias, early to graduate education, financing, public advocacy, mentorship, and mental health/behavioral factors. We're at the starting point of a roadmap that could potentially break cycles so rooted in the past for Blacks, and perhaps also help other groups navigate their pursuit of success in science too.

–Cato T. Laurencin



PHOTO: UCONN PHOTOS

Cato T. Laurencin is the Albert and Wilda Van Dusen Distinguished Endowed Professor of Orthopaedic Surgery, and the chief executive officer of the Connecticut Convergence Institute for Translation in Regenerative Engineering at UConn Health, Farmington, Connecticut, USA. He is the University Professor at the University of Connecticut, Storrs, Connecticut, USA. laurencin@uchc.edu

Downloaded from http://science.sciencemag.org/ on November 28, 2019

10.1126/science.aba2319

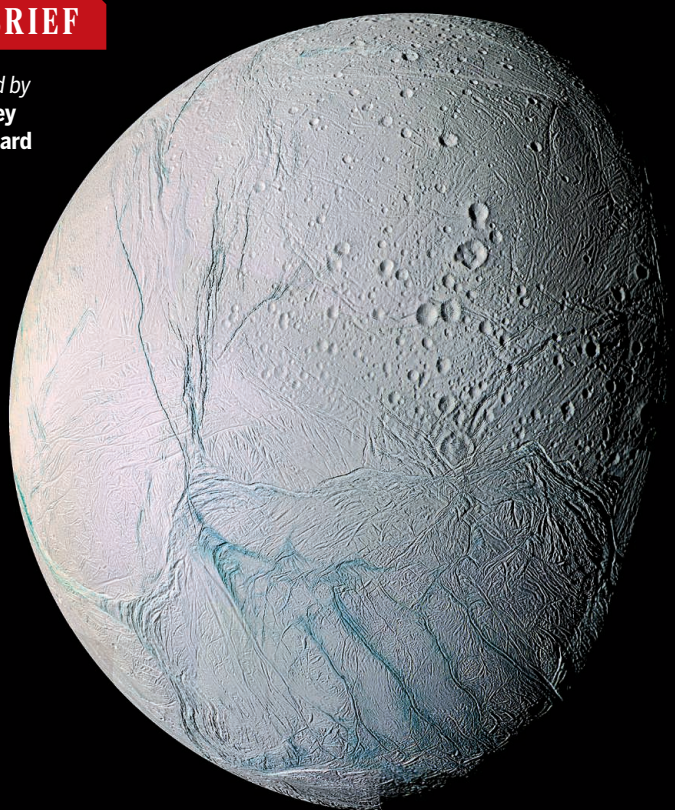
Published by AAAS

NEWS

> “In the interest of protecting personal privacy, we elected to not host the NIH dataset.”
>
> **A Google spokesperson**, in *The Washington Post*, revealing that in 2017, the company almost made public 100,000 x-rays with potentially identifying information as part of a research project.

IN BRIEF

Edited by **Jeffrey Brainard**



PLANETARY SCIENCE

How a moon of Saturn got its stripes

Researchers say they have solved a long-standing mystery about why Saturn’s frozen moon Enceladus features four parallel, water-spewing slashes known as tiger stripes, not found on any other satellite in the Solar System. The answer involves natural cooling that allowed an ice shell lying atop the moon’s liquid ocean to thicken. The shell became stressed and ruptured. A 130-kilometer-long fissure extended down into the ocean, creating a geyser that spewed snow on its flanks. The weight of that snow created stresses that cracked open parallel fissures, wrote Douglas Hemingway of the Carnegie Institution for Science in Washington, D.C., and co-authors, who posted a paper this month on the preprint server arXiv. The geysers, photographed in 2005 by NASA’s Cassini spacecraft, contain organic molecules, making Enceladus a top candidate to explore for life elsewhere in the Solar System.

CRISPR’s first clinical success?

BIOMEDICINE | CRISPR Therapeutics of Zug, Switzerland, told investors and biotech analysts this week that, in a first, it has used the CRISPR gene editor to help two people with blood diseases—sickle cell anemia in one and beta thalassemia in the other—caused by mutations to their hemoglobin gene. Collaborating with Boston-based Vertex Pharmaceuticals, the company removed blood stem cells from each patient, edited a gene in them to restore production of a fetal version of hemoglobin, then transplanted the cells back. Before the treatments, both patients required multiple infusions of red blood cells each year. Since then, they have produced high levels of fetal hemoglobin; one avoided transfusions for 4 months, the other for 9 months. “These early findings are a triumph for patients, CRISPR, and genetics,” says Bruce Conklin, who studies stem cell treatments at the University of California, San Francisco.

Easing remote access to journals

PUBLISHING | The *Nature* family of journals announced on 12 November changes intended to ease headaches for off-campus readers wanting to access its paywalled journal articles. Even though many universities allow faculty members and students to gain access by logging in through an online portal, many articles don’t clearly flag that possibility, and following the required steps can be cumbersome. Articles in the 150 journals at nature.com now prominently display a standard button for access; users can use it to identify their home institution and connect to its log-in page. A browser code will reduce the log-in steps during subsequent visits and, if other publishers team up with nature.com as expected, streamline log-ins further. But some librarians worry the changes could compromise users’ privacy.

Pakistan targets resistant typhoid

PUBLIC HEALTH | Pakistan said last week it plans widespread use of a new typhoid vaccine to help end an outbreak of an extensively drug resistant (XDR) form of

Downloaded from http://science.sciencemag.org/ on November 28, 2019

PHOTO: NASA/ESA/JPL/SSI/CASSINI IMAGING TEAM

Published by AAAS

Salmonella enterica Typhi, a sometimes deadly bacterium spread by contaminated water and food. With support from Gavi, the Vaccine Alliance, a nonprofit in Geneva, Switzerland, Pakistan will become the first country to start to routinely administer the vaccine to children. The new vaccine offers two major improvements over a predecessor: It works in very young children and triggers longer lasting immunity. Since the outbreak started in November 2016, XDR *S.* Typhi has caused more than 11,000 cases of typhoid. The XDR form is resistant to all but one oral antibiotic, and if it fails, people must take intravenous drugs that are impractically expensive for low-income countries such as Pakistan. If untreated, typhoid can cause intestinal hemorrhages; it kills about 200,000 people worldwide each year.

Far planetary object renamed

PLANETARY SCIENCE | MU69, the most distant Solar System object ever explored, has received its official name: Arrokoth, a Native American term meaning "sky" in the Powhatan/Algonquian language, NASA announced last week. On 1 January, the New Horizons spacecraft flew near Arrokoth (pronounced AR-uh-koth), revealing two primordial planetary building blocks smooshed together. The discovery team had nicknamed the object "Ultima Thule," meaning beyond the known world, but critics later noted that white supremacists had previously co-opted the term to describe a mythical Aryan homeland. The new name, used with consent of Powhatan elders, honors the original inhabitants of the Chesapeake Bay region, where the institute that operates New Horizons is located.

U.S. science advisers start work

SCIENCE POLICY | Although still below half strength, the U.S. President's Council of Advisors on Science and Technology (PCAST) hit the ground running at its first meeting this week. Science adviser Kelvin Droegemeier, its chair, told the seven current members of what's expected to be a 16-person panel that the Trump administration wants guidance on how to ensure U.S. leadership in five "industries of the future"—artificial intelligence, quantum science, 5G wireless networks, advanced manufacturing, and synthetic biology—as well as suggestions on training a robust U.S. workforce in those fields. Droegemeier is also assembling a 20-member "junior" PCAST to give university students and early-career scientists a voice in policymaking.

PHOTO: SERGEY RYUMIN/GETTY IMAGES



ANIMAL SCIENCE

Big health study examines canine longevity

One hundred and one Dalmatians won't be enough. After several years of planning, the Dog Aging Project was officially launched last week and is seeking 10,000 pets—and their owners—to participate in what may be the most ambitious canine health study ever. Using private donations and a U.S. National Institute on Aging grant, scientists and veterinarians from multiple U.S. universities will track dogs of all breeds over 10 years to identify genetic, lifestyle, and environmental factors that influence the life spans of our best friends—and, project leaders hope, also discover benefits for humans. Earlier this month, a separate research team reported on the bioRxiv preprint server that dogs and people have similar age-related buildup of methyl groups on DNA. That finding inspired them to develop a new formula for converting a dog's age to human years: 16 ln (dog age) + 31. ("ln" is the natural logarithm. Or enter a dog's age in the calculator here: https://scim.ag/DogAgeCalculator.)

Brazil's forest loss surges

CONSERVATION | Development, most of it illegal, destroyed more than 9700 square kilometers of Brazilian Amazon rainforest in the year ending in July, according to a government estimate released 18 November. That is an increase of 30% from the previous year and the highest rate of deforestation since 2007–08. The number is based on analysis of high-resolution Landsat satellite images by the Program for Monitoring Deforestation in the Amazon by Satellite, run by Brazil's National Institute for Space Research. Many scientists and environmentalists blame the deforestation spike on President Jair Bolsonaro's aggressive policies to support mining and agriculture and to dismantle environmental protections. But Minister of the Environment Ricardo Salles said the increased deforestation continues a trend that began in 2012, before Bolsonaro was elected.

BY THE NUMBERS

636

Scientists in China who published papers ranked in the top 1% by citations in their field, up 32% since 2018 and second only to the U.S. tally of 2737 (Clarivate Analytics).

$10 million

Purse in a newly announced XPRIZE competition to develop technology to help survey and protect biodiversity in tropical rainforests. The registration deadline is June 2020.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

IN DEPTH

PLANETARY SCIENCE

Bold plan to retrieve Mars samples takes shape

NASA and Europe settle on suite of missions to bring back rocks from the Red Planet

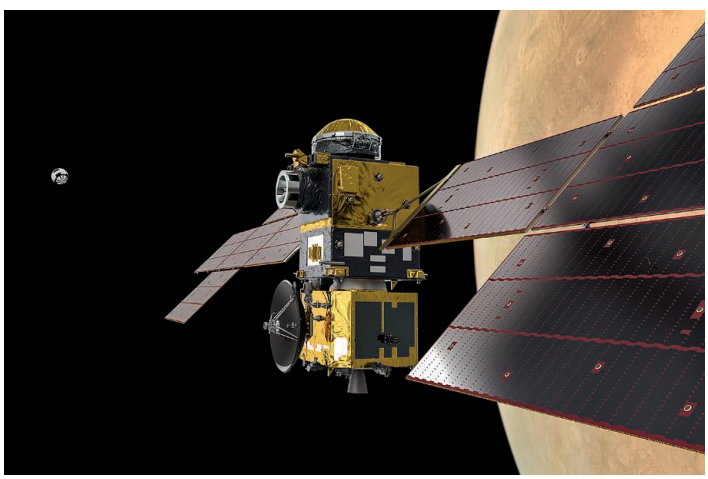
By **Daniel Clery** *and* **Paul Voosen**

In just over a decade, a small capsule shaped like a flying saucer will blaze in from space and smash into an empty Utah desert. Inside the capsule, protected by shock absorbers, will be a precious payload: about half a kilogram of rocks gathered on Mars. After years as a dream, Mars sample return (MSR) is now a $7 billion plan, devised jointly by NASA and the European Space Agency (ESA). Involving three heavy rocket launches from Earth, two rovers, the first ever rocket launch from another planet, and a daring rendezvous between the sample container and a spacecraft that would ferry it back to Earth, "It's as complicated as sending humans to the Moon," says Brian Muirhead, lead MSR planner at NASA's Jet Propulsion Laboratory in Pasadena, California.

If successful, MSR would fulfill what a 2011 report from the National Academies of Sciences, Engineering, and Medicine said was NASA's top priority in the study of the Solar System. Rovers like Curiosity, the car-size robot now exploring Mars, are increasingly capable, but their mobile labs are no match for tools available on Earth. Like the Apollo Moon rocks, Mars samples would be valuable across generations. Studies could reveal how a warm and wet Mars became cold and dry. Scientists could also get lucky and find evidence of life. "This has got to be one of the greatest missions of space exploration," says space scientist Andrew Coates of University College London.

Approval for ESA's initial contribution is expected next week, when ministers from ESA's 22 member nations gather in Seville, Spain, to approve the agency's budget for the next 3 years. David Parker, ESA's director of human and robotic exploration in Noordwijk, the Netherlands, says the total ESA contribution will be roughly €1.5 billion over 10 years, "which should be affordable to member states." At the ministerial meeting, he will ask for just under €600 million to pay for the first few years of MSR as well as the final elements of a separate project, the ExoMars rover due for launch next year. Parker is confident he'll get it. "This is the ministerial where Europe will step up and commit to this next phase of exploration," he says.

For its part, NASA will finalize its portion of the MSR architecture by January 2020, and the agency is hoping an MSR mission line will appear soon after in the Trump administration's 2021 budget request to Congress. Michael Meyer, NASA's Mars exploration scientist in Washington, D.C., says it's the right moment for the mission, which can take advantage of a cadre of engineers and scientists with rover experience and a set of still-healthy Mars orbiters that could provide communication relays. "Everything is kind of lined up right now," Meyer says.



Europe's Earth return orbiter must track and snare the orbiting sample sphere.

The first stage of MSR is nearly complete and mostly paid for: NASA's $2.5 billion Mars 2020 rover. Due to launch in July 2020, the rover will land in Jezero crater near a fossilized river delta nearly 4 billion years old. The rover will range widely, performing science experiments and occasionally drilling small cores of mudstones and other rocks that could hold signs of ancient life. Each sample, containing up to 20 grams of rock and grit, will be stored in a tube about the size of a penlight. NASA is likely to cache some tubes on the ground and keep others on the rover.

MSR's next leg would come in 2028 when another NASA lander arrives carrying both a return rocket and an ESA-built "fetch" rover. After launch in 2026, the lander would take a slow, elongated route to Mars, allowing it to deliver a hefty 2100-kilogram payload—more than the Curiosity and Mars 2020 rovers combined. It would touch down in August 2028, after Mars's dust storm season, most likely using retrorockets guided by a new technology called terrain relative navigation. While descending, the lander will use a belly-mounted camera to take stock of the terrain and avoid hazards, putting it within 10 kilometers of the cached samples.

The fetch rover, about the size of a coffee table, will trundle out of the lander to seek the sample tubes deposited years earlier. The rover will rely on solar power, which gives it roughly 6 months to get its work done before it succumbs to the waning winter sun. To meet that tight schedule, it will travel up to 200 meters per day, navigating autonomously. "It's the only way to achieve the distances we need," says Sanjay Vijendran, ESA's campaign lead for MSR in Noordwijk.

Once the fetch rover returns to the lander, an ESA-built robotic arm will remove the sample tubes from racks on the rover and place up to 30 of them, containing just over half a kilogram of martian rock and soil, in a sphere the size of a basketball. (If Mars 2020 is still operating and carrying tubes, the arm can offload them, too.) The 2020 rover has 43 sample tubes, but some will remain on Mars so the rocket can remain a reasonable size, Muirhead says. Scientists seem to be able to live with the constraint, he says. "The general consensus was, 'Bring us back 30 good tubes and we'll be happy.'"

Getting the sample sphere off the planet won't be easy, says Angie Jackman, an engineer who is leading the return rocket's de-

IMAGE: © ESA/ATG MEDIALAB

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

velopment at NASA's Marshall Space Flight Center in Huntsville, Alabama. "This will be the first rocket off another planet," she says. NASA knew the rocket would have to be short and stubby to fit in the lander's protective entry shell. It also needed to survive for months in the cold. For years the agency pursued a hybrid rocket fueled by paraffin wax and a liquid oxidizer. But the selection of Jezero, which has a mild climate for Mars, meant a conventional solid rocket less than 3 meters tall could do the job without fear of its fuel cracking or exploding.

The sample sphere will be fired into an orbit higher than 300 kilometers to meet ESA's Earth return orbiter, which is supposed to follow the lander to Mars. With no martian version of GPS to guide it, ESA's orbiter will use cameras to track and capture the sphere. "The way we've set the geometry, we have a very good chance of seeing it within the first few hours," Muirhead says.

After the rendezvous, NASA-built robotic mechanisms will shunt the sphere into a containment module, seal it, and sterilize it. Like a Russian nesting doll, the module will be placed within a third protective shell, and then finally a fourth vessel: a disk-shaped vehicle to withstand a fiery, parachuteless re-entry through Earth's atmosphere.

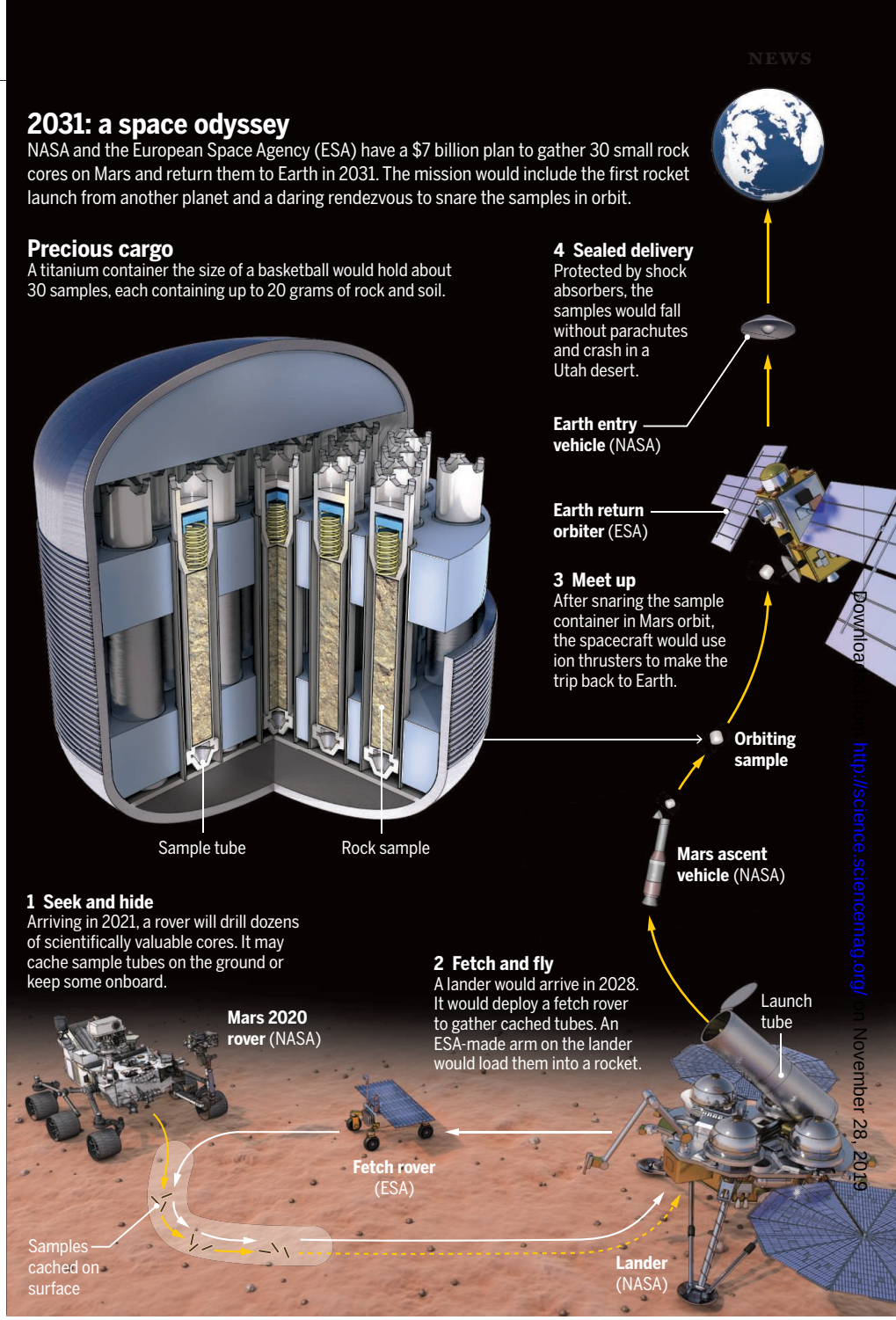
To minimize the amount of fuel needed to send the craft back to Earth, ESA will mostly rely on ion thrusters, which use electricity from solar panels to drive a stream of ions out of the spacecraft, producing a gentle impulse that can be sustained for months. The approach—which ESA's BepiColombo probe is using to get to Mercury—generates so little thrust that each leg of the journey will take 2 years instead of the usual 9 months or so. "It's a disadvantage but it means we can do this within Europe's launch capability," Vijendran says.

Once the samples arrive on Earth in 2031, they'll be handled with care: They will be quarantined in a facility meant to protect the rocks from exposure to microbes on Earth—and earthly life from any martian pathogens. Developing instruments that won't contaminate the samples will be a challenge, but a manageable one, Meyer says.

Exactly how scientists will study the samples more than a decade from now is hard to predict. But they will surely want to check for signs of past or present life. Last year, the Curiosity rover identified two sites flush with complex organic macromolecules that resemble the goopy building blocks of oil on Earth. But its instruments cannot tease out whether life or geology produced these molecules. With just a few grams of material, however, scientists on Earth will be able to analyze carbon isotopes in similar materials from Jezero crater to determine whether life created them, says Maria-Paz Zorzano of the Center for Astrobiology in Madrid, who helped define MSR science priorities and sample selection for an international panel.

And if Jezero's clays show no signs of past life, scientists will want to know why. "Was it because it wasn't habitable, or because life couldn't get started?" asks Benjamin Weiss, a planetary scientist at the Massachusetts Institute of Technology in Cambridge. Scientists believe Mars lost its magnetic field billions of years ago as its core cooled, allowing its atmosphere to escape and leaving its surface cold, dry, and exposed to radiation from space. By dating different samples and measuring their fossil magnetism and the conditions in which they formed, scientists can test whether this theory is true—and, if so, whether the martian environment grew inhospitable so early that life could never get started.

With momentum building for much more Mars exploration—including human visitors—chances are waning to study the planet while it is pristine, Zorzano says. "I hope we don't miss the chance to catch this train," she says. "If we don't step in now, we will miss this unique moment." ■



GRAPHIC: C. BICKEL/*SCIENCE*; MARS 2020 ROVER MODEL/NASA/JPL-CALTECH

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

PLANT SCIENCE

After 20 years, Golden Rice nears approval

Bangladesh may become the first country to adopt transgenic rice enriched in vitamin A

By **Erik Stokstad**

Soon. That has long been scientists' answer when asked about the approval of Golden Rice, a genetically modified (GM) crop that could help prevent childhood blindness and deaths in the developing world. Ever since Golden Rice first made headlines nearly 20 years ago, it has been a flashpoint in debates over GM crops. Advocates touted it as an example of their potential benefit to humanity, while opponents of transgenic crops criticized it as a risky and unnecessary approach to improve health in the developing world.

Now, Bangladesh appears about to become the first country to approve Golden Rice for planting. "It is really important to say we got this over the line," says Johnathan Napier, a plant biotechnologist at Rothamsted Research in Harpenden, U.K., who was not involved in the crop's development. He says approval would show that agricultural biotechnology can be successfully developed by publicly funded research centers for the public good. Still, environmental groups haven't dropped their opposition—and the first harvest isn't expected until at least 2021. And more research will be needed to show the extent of real-world benefits from Golden Rice.



A serving of Golden Rice contains half the beta-carotene children need daily.

Golden Rice was developed in the late 1990s by German plant scientists Ingo Potrykus and Peter Beyer to combat vitamin A deficiency, the leading cause of childhood blindness. Low levels of vitamin A also contribute to deaths from infectious diseases such as measles. Spinach, sweet potato, and other vegetables supply ample amounts of the vitamin, but in some countries, particularly those where rice is a major part of the diet, vitamin A deficiency is still widespread; in Bangladesh it affects about 21% of children.

To create Golden Rice, Potrykus and Beyer collaborated with agrochemical giant Syngenta to equip the plant with beta-carotene genes from maize. They donated their transgenic plants to public-sector agricultural institutes, paving the way for other researchers to breed the Golden Rice genes into varieties that suit local tastes and growing conditions.

Over the past 2 years, regulators in the United States, Canada, New Zealand, and Australia approved Golden Rice for consumption. There are no plans to grow the crop in these countries, but approval will prevent problems if Golden Rice somehow accidentally turns up in food supplies.

The Golden Rice under review in Bangladesh was created at the International Rice Research Institute (IRRI) in Los Baños, Philippines. Researchers bred the beta-carotene genes into a rice variety named dhan 29, which is grown widely during the dry season in Bangladesh and contributes about 14% of the national harvest. In tests of dhan 29 Golden Rice at multiple locations, researchers at the Bangladesh Rice Research Institute (BRRI) in Gazipur found no new farming challenges and no significant differences in quality—except for the presence of vitamin A.

BRRI submitted data to the Bangladeshi Ministry of Environment, Forest, and Climate Change in November 2017. The Biosafety Core Committee, a group of eight officials and scientists, has since been reviewing environmental risks, such as the plant's potential to become a weed, as well as food safety. The review is nearing completion; on 28 October, the *Dhaka Tribune* reported that a decision would be made by 15 November.

That date has come and gone; the holdup appears to be due to the death of a committee member. But a source familiar with the committee's deliberations says some members remain skeptical of Golden Rice, wondering for example why it is needed when people could also eat more vegetables.

Proponents are optimistic, however. The scientific evidence is strong, the committee previously approved another transgenic crop, and Golden Rice enjoys high-level political support in Bangladesh, they say. "We are hopeful that Golden Rice might get the green light soon," says Arif Hossain, director of Farming Future Bangladesh in Dhaka, which is funded by the Bill & Melinda Gates Foundation to inform policymakers and others about biotechnology.

After the environment ministry signs off, Golden Rice must be registered by a seed certification agency within the Ministry of Agriculture, which requires field trials in multiple places to test for seed quality. If all goes smoothly, farmers might have Golden Rice seed to plant by 2021.

How popular it will be is uncertain. Farmers in Bangladesh quickly adopted an eggplant variety engineered to kill certain insect pests after its 2014 introduction, but that crop offered an immediate benefit: Farmers need fewer insecticides. Golden Rice's health benefits will emerge more slowly, says agricultural economist Justus Wesseler of Wageningen University & Research in the Netherlands, so adoption may be slower as well. The government may need to promote Golden Rice and, Hossain says, even subsidize farmers to grow it.

Consumer acceptance may be another challenge, given the golden hue, says Sherry Tanumihardjo, who studies vitamin A and global health at the University of Wisconsin in Madison. "People have a difficult time changing the color of food they eat," and many people in Bangladesh prefer to eat white rice. On the other hand, cooked Golden Rice resembles khichuri, a popular dish of rice and lentils cooked with turmeric, which may increase its appeal. With Gates Foundation support, IRRI and BRRI are developing a strategy for directing farmers' harvest to rural regions and cities with high poverty and malnutrition rates.

Opposition from nongovernmental organizations could still hobble the introduction. Last month, two groups in Bangladesh—the Agricultural Farm Labour Federation and the National Women Farmers & Workers Association—called for a ban on both Golden Rice and transgenic eggplant.

If Golden Rice does make inroads in Bangladesh, additional varieties better adapted to other seasons or locations may follow. Bred at BRRI, they are now in greenhouse trials. Like all local versions of Golden Rice, these varieties were created not with genetic engineering, but by traditional backcrossing, so they will likely not need biosafety approval. "There will be no problem for a year-round supply of biofortified rice," says Partha Biswas, a plant breeder at BRRI.

But for now, all eyes are on dhan 29. "It would be great to see it approved," Napier says. "It's been a long time coming." ■

PHOTO: ISAGANI SERRANO/CPS/IRRI PHOTO/FLICKR/CC BY-NC-SA

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS



On 17 November, protesters threw Molotov cocktails at police outside Hong Kong Polytechnic University in China.

PHOTO: KYODO/AP IMAGES

CHINA

Hong Kong protests hit universities

Clashes have had little impact on research, but long-term effects on schools are unclear

By **Dennis Normile**

The once-peaceful protests in Hong Kong over the erosion of the territory's autonomous status within China and delays in expected democratic reforms escalated last week into pitched battles between protesters and police on multiple university campuses. At one site, several senior school officials who are scientists sought to negotiate a truce amid rubber bullets and the haze of tear gas.

The confrontations wound down early this week, but academics worry about the future of the city's universities. "The current situation will make it difficult for us to recruit top-quality staff in the future," says Sun Kwok, a Hong Kong–born astronomer who was dean of science at the University of Hong Kong (HKU) for 10 years. He fears it may also be tough to fill HKU's graduate schools, where most students come from mainland China or other countries. The science faculty recruits about 120 grad students each year, and filling those slots "might be quite difficult," says Matthew Evans, HKU's dean of science.

Kwok adds that authorities appear to blame university students for instigating the demonstrations. "If this results in increased control by the government on the universities, it could lead to an erosion of academic freedom," says Kwok, who is now at the University of British Columbia in Vancouver, Canada.

As *Science* went to press, an estimated 100 hard-core activists remained holed up on the campus of Hong Kong Polytechnic University (PolyU). Calm had returned elsewhere. Scientists and officials at the Chinese University of Hong Kong (CUHK), HKU, and the Hong Kong University of Science and Technology say that on their campuses, electrical power and internet service continued, and there were no fires within buildings and no losses of culture collections or lab animals. "We had zero damage to our buildings," Evans says.

Hong Kong universities had escaped disruption during the massive demonstrations in the city that started in early June, touched off by a proposed bill to ease extradition of suspected criminals to mainland China for trial. The extradition bill was withdrawn on 23 October. But by then, protesters were demanding broader democratic reforms and independent investigations of police brutality.

In the second week of November, protesters escalated tactics, using perches on hilly university campuses to hurl debris onto adjacent highways. They may have thought they could take advantage of a long-standing policy that police "do not come onto [a] campus unless they have reason to suspect some criminal act has come about or have an appropriate warrant," Evans says.

Radical activists commandeered CUHK, even demanding that visitors to the campus show identification to pass blockades, according to an open letter from the vice chancellor, Rocky Tuan, a biomedical scientist who was pictured trying to negotiate a truce on a pedestrian bridge leading to the campus last week. After masked protesters left CUHK and other schools to gather at PolyU at the end of last week, a police officer on the edge its campus was wounded by an arrow and the situation devolved into a fiery showdown. Many graduate students have fled the city.

Despite the chaos, researchers were generally able to keep their studies on track, says Chan King Ming, a CUHK aquatic toxicologist. He reached his lab to feed zebrafish. Barring further upheaval, several Hong Kong scientists told *Science* they believe research activities will resume smoothly, if not immediately. Although classes are now suspended or being taught online, most schools plan to return to a normal schedule when the next semester begins in January 2020.

Ensuring Hong Kong's long-term tranquility requires a political solution that is beyond the scope of what universities can do, school officials say. "The government must take the lead with swift and concrete action to resolve this political deadlock and to restore safety and public order now," the heads of Hong Kong's nine publicly funded universities wrote in a joint statement last week—before some of the most violent protests. ■

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

Corrected 22 November 2019. See full text.

DATA PROTECTION

Researchers sound alarm on European data law

Privacy regulations impede international studies on diabetes and Alzheimer's disease

By **Tania Rabesandratana**

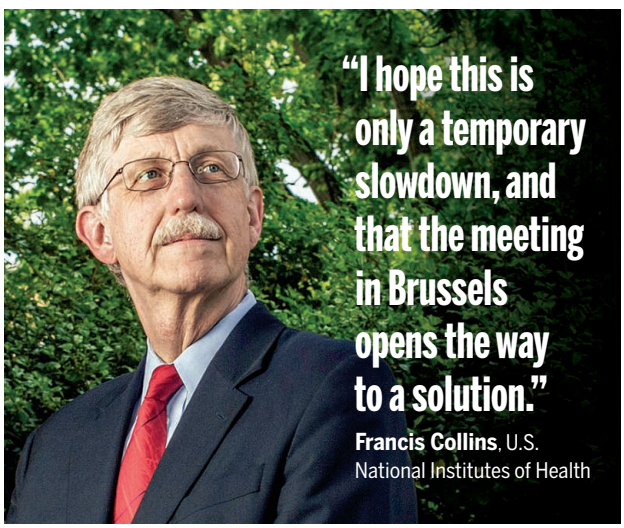
For many people, the most apparent effect of the European privacy law called the General Data Protection Regulation (GDPR) has been a flourishing of website pop-ups, demanding your consent to store browsing behavior as cookies. An annoyance, perhaps, but hardly more than an inconvenience. For Francis Collins, director of the U.S. National Institutes of Health (NIH), however, the regulation has turned out to be a serious impediment to research.

Since 1993, Collins has been principal investigator for a project studying type 2 diabetes in Finnish people, who have relatively homogenous genetics and detailed health records. Finland's National Institute for Health and Welfare has sent 32,000 DNA samples to Collins's laboratory. He and his U.S. collaborators used the data to discover more than 200 places in the genome where variants increase the risk of illness. But in May 2018, when GDPR came into force, the Finnish institute stopped all data sharing on the project, because NIH could not provide guarantees that would satisfy the institute's interpretations of the law's requirements. Progress has since "slowed to a crawl," Collins says.

This week in Brussels, representatives from NIH, academia, industry, patient advocacy groups, the European Commission, and data protection authorities met to share their GDPR frustrations. They hope to highlight the obstacles it creates for some international collaborations and explore possible responses. "I hope this is only a temporary slowdown, and that the meeting in Brussels opens the way to a solution," Collins says.

The European Union's GDPR rules, which apply to the 28 EU member states plus Iceland, Liechtenstein, and Norway, include common sense principles, such as minimizing personal data used in research and using appropriate safeguards. Because "there's now teeth and liability attached," with steep penalties for rule breakers, the regulation has "scared everyone," says Cathal Ryan, assistant commissioner at Ireland's Data Protection Commission in Dublin, leading to scrutiny of projects that rely on personal data.

The European Union recognizes some countries—Argentina, Japan, New Zealand, and Switzerland, among others—as providing adequate data protection, which frees EU scientists to share data with researchers in those nations. But not in Canada and the United States. One way for research institutions there to collaborate with EU researchers is to sign contracts that guarantee data safeguards. However, standard contracts include requirements that institutions agree to European audits of their data systems or submit to the jurisdiction of its courts—which NIH, as a U.S. government agency, cannot accept. "That was a nonstarter," Collins says.



Some researchers are finding workarounds, but they are less than ideal. Neuroscientist Sudha Seshadri of the University of Texas Health Science Center in San Antonio is one of the co-founders of the International Genomics of Alzheimer's Project, which has gathered DNA sequences from more than 90,000 people in Europe and the United States to find genetic variants associated with Alzheimer's disease. She says partners in some EU nations have restricted data sharing, so the consortium now runs separate analyses on each side of the Atlantic Ocean. But this limits analysis, particularly when searching for rare variants that require big data sets, Seshadri says.

Although GDPR lays out overarching principles, it leaves member states to spell out details of research exemptions in national laws. "There's a common joke among ourselves: If you ask 20 lawyers, they'll give you 20 different opinions" on how to comply with it, says Salvador Capella Gutierrez, who leads the Spanish National Bioinformatics Coordination Node at the Barcelona Supercomputing Center. For example, reusing data for secondary research is typically allowed in Spain without additional patient consent when this reuse is deemed in the "public interest." But in countries such as Italy, researchers often have to ask for consent again.

Another way to avoid the regulation and its uncertainties is to anonymize personal research data. But Mads Melbye, CEO of the Statens Serum Institute in Copenhagen, says countries also have different interpretations of what constitutes appropriate anonymization. Even if subjects are not identified, their biological samples contain what some officials consider personally identifiable information, he says. After GDPR was enacted, his institute, which houses the Danish National Biobank, froze data streams to important partners, including NIH and the World Health Organization's International Agency for Research on Cancer in Lyon, France. "We're talking about high-value data collections that have been costly to establish," he says. "It's a disaster for international collaboration if we can't find a solution."

NIH and its Finnish counterpart are close to resuming data transfers, under a deal defining them as "necessary for important reasons of public interest," Collins says. "We're trying to come up with a template that would be consistent with GDPR, but would resolve ambiguities that lead legal experts to take a conservative view for fear of financial penalties," he explains.

Seshadri is confident that clarity and best practices will emerge over time. "I'm hopeful that over the next few years we will find ways to do this efficiently," she says. In the meantime, researchers aren't alone in being frustrated, she says. "As a patient, you want solutions yesterday." ■

PHOTO: STEPHEN VOSS/REDUX

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS



A homeowner inspects the damage done earlier this month by one of Australia's many wildfires.

PHOTO: WILLIAM WEST/AFP/GETTY IMAGES

FIRE SCIENCE

Australian blazes will 'reframe our understanding of bushfire'

Fire scientist David Bowman discusses the unfolding crisis

By **John Pickrell**, *in Sydney, Australia*

Australia is on fire like never before—and this year's "bushfire" season, which typically peaks in January and February, has barely begun. Driven in part by a severe drought, fires have burned 1.65 million hectares in the state of New South Wales, more than the state's total in the previous 3 years combined. Six people have died and more than 500 homes have been destroyed. As *Science* went to press, some 70 uncontrolled fires were burning in adjacent Queensland, and South Australia was bracing for potentially "catastrophic" burns.

David Bowman, a fire ecologist and geographer and director of the Fire Centre at the University of Tasmania in Hobart, spoke with *Science* about the crisis. The flames have charred even moist ecosystems once thought safe, he says. And the fires have become "white-hot politically," with Prime Minister Scott Morrison's Liberal government drawing criticism for refusing to acknowledge any link to climate change. The interview has been edited for clarity and brevity.

Q: What is unusual about these fires?

A: The geographical scale and intensity—it's happening all up and down the country. The very early start to the fire season across eastern Australia. The scale of housing loss.

We're seeing recurrent fires in tall, wet eucalypt forests, which normally only burn very rarely. A swamp dried out near Port Macquarie, and organic sediments in the ground caught on fire. When you drop the water table, the soil is so rich in organic matter it will burn. We've seen swamps burning all around.

Even Australia's fire-adapted forest ecosystems are struggling because they are facing increasingly frequent events. In Tasmania, over the past few years we have seen environments burning that historically see fires very rarely, perhaps every 1000 years. The increasing tempo, spatial scale, and frequency of fires could see ecosystems extinguished.

Q: What is the role of climate change?

A: You have to ask: Has there ever been a fire event of 1.65 million hectares that's burnt a large area of what is generally considered fire-proof vegetation, and also occurred simultaneously with fires in other regions of Australia and California? What is happening is extraordinary. It would be difficult to say there wasn't a climate change dimension. We couldn't have imagined the scale of the current event before it happened. We would have been told it was hyperbole.

This is teaching us what can be true under a climate changed world. The numbers, scale, and diversity of the fires is going to reframe our understanding of bushfire in Australia. This is a major event which will have huge intellectual and policy legacies.

Q: What policy issues has the crisis raised?

A: The fires are white-hot politically. The fact that the government has questioned the appropriateness of talking about the linkage to climate has put scientists in a difficult position. There has been an unexpected alliance between 24 former state fire and emergency chiefs who are advocating an awareness of increasing bushfire risk linked to climate change. They argue the government has ignored their advice because it doesn't want to talk about climate. What they are saying is harmonizing with the warnings coming from scientists.

Q: How do the Australian fires relate to fires in other parts of the world?

A: Whilst this was unfolding in Australia, there has been a very late fire season in California. So, we had the long-dreaded overlap between the Northern and Southern Hemisphere fire seasons. If this becomes a fixture it's going to make sharing firefighting resources between nations much harder. The U.S. has a huge, specialized firefighting aviation force. Traditionally, we were able to bring U.S. aircraft and personnel to Australia, but with fire events happening simultaneously there will be strains on resource sharing.

Q: How big of a problem is the smoke?

A: There is a strong epidemiological association between smoke and respiratory and cardiovascular health. A huge fraction of the population along the east coast—maybe a third of the total Australian population of 25 million—has been exposed to dangerous smoke levels. This will be costly to the health and hospital systems and take people's lives. Smoke pollution is currently killing far more people than the flames, and it's not over yet.

Q: What research questions do the fires raise?

A: These are overwhelming in number. The priorities include: getting to grip with the human health impacts of the smoke, understanding fire severity and ecological harms, and understanding why the intensity varied between regions and how much it was related to prevailing weather. There will also be more formal government enquiries, which will draw on postfire assessments and the work of climate scientists. They will be very influential on future fire management practices.

How do you prepare and fund research that was unimaginable even a few months ago? There is a disconnect between the scale and speed of these events and the availability of funding to study them. We are lucky in that we have some existing projects and databases. I would expect a large cross-section of scientists from Australia will be entangled in the research in one way or another. ■

John Pickrell is a journalist in Sydney, Australia.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

FEATURES

THE LANDSLIDE LABORATORY

Landslides aren't just hazards. An experiment in Taiwan could reveal their outsize effect on the climate

By **Katherine Kornei**, *in Taroko National Park, Taiwan*

The frequent crackle of tumbling rocks overhead is unnerving, especially when you're picking your way through a pile of jagged debris. "I hate walking down roads like this," says Niels Hovius, a geomorphologist at the GFZ German Research Centre for Geosciences in Potsdam. "I know what can happen here."

Taroko National Park, famous for a precipitous marble gorge that cuts through it, is in a futile fight with gravity. Rockfalls litter the park's serpentine main highway. The scars of at least a dozen landslides punctuate the view in all directions. Maintenance crews are perpetually spraying concrete on slopes in a last-ditch effort to stabilize them. The park gives out safety helmets for free, and strongly encourages visitors to wear them.

For Hovius, all this moving rock and soil makes for a perfect laboratory. For the past 3 years, he and his colleagues have scrambled and rappelled across the park, installing dozens of instruments in what will end up being Taiwan's most comprehensive landscape dynamics observatory. One goal is to monitor landslides and understand their triggers. A bigger aim is to investigate their hidden impact on the climate: As massive chemical reactors, landslides draw carbon dioxide (CO_2) out of the sky and sometimes belch it out, too. Understanding their role as both carbon source and sink could help researchers better model the carbon cycle that ultimately controls our planet's climate and habitability.

The recipe for landslides requires three basic ingredients: steep hillslopes, earthquakes to weaken them, and water to make them slick. Taiwan has all three in spades, making it one of the most landslide-prone countries in the world. The island was born 6 million years ago in an ongoing collision of tectonic plates that lifts mountains and generates a drumbeat of earthquakes. And its location in the tropical western Pacific Ocean means typhoons come regularly, occasionally dumping meters of rain in just days. "You can learn a lot by looking at extremes," says Susan Brantley, a geochemist at Pennsylvania State University in State College.

Hovius first visited Taiwan to do fieldwork in the late 1990s, lured by the island's extreme climatic and tectonic forcing and its extensive scientific records. A few years ago, he and GFZ physicist Jens Turowski, working with collaborators at Taiwanese institutions, began to plan something more ambitious. By 2021, their observatory, built with roughly $600,000 in research funds from the Helmholtz Association of German Research Centres, will consist of more than 120 instruments spread across hundreds of square kilometers. Solar powered and autonomous, they will relay data once per hour to servers in Germany and Taiwan.

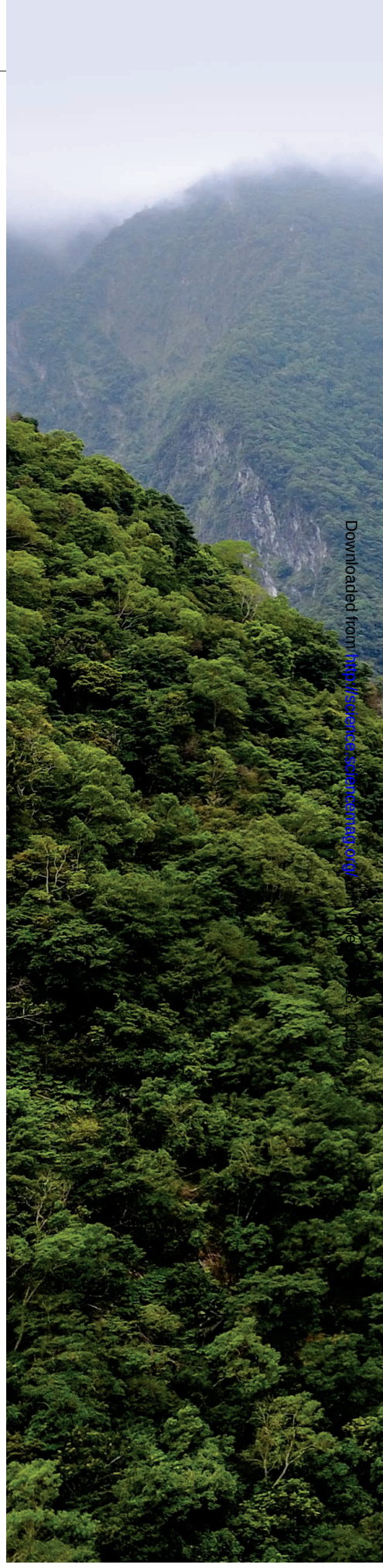


PHOTO: KRISTEN L. COOK

Downloaded from http://science.sciencemag.org/ on May 28, 2020

Published by AAAS



Steep, wet slopes and earthquakes make landslides a common feature in Taiwan's Taroko National Park.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Served by AAAS

Many of the instruments will work as sensitive landslide detectors. Seismometers will pick up ground shaking from tumbling rocks, and cameras will record fresh landslide deposits and scars. The continuous monitoring is a step up from patchy satellite observations and sporadic reports from tourists and park rangers, Hovius says. He and Turowski plan to share their data with Taroko officials, who can use the near–real-time detections to decide whether to close trails or roads—or search for stranded hikers.

Monitoring landslides will also help the team understand what triggers them and how landscapes recover. In 1999, a magnitude-7.6 earthquake struck central Taiwan, killing thousands of people. A few months later, researchers measured a roughly 20-fold increase in the landslide rate near the earthquake's epicenter. The uptick makes sense: Ground shaking presumably primes the slopes to give way. The surprise was that landslide rates returned to normal after a few years: The landscape somehow knitted itself back together. Rocks settling, fractures filling in, and plant roots binding the soil might all play a role in stabilizing the terrain, Hovius says. To test those ideas, the team might use "crack meters" to monitor how fissures open and close and laser scanners to detect small changes in topography. The observatory's 10-year life span should be long enough to record several large earthquakes, giving the researchers a chance to witness the rise and fall of landslide rates.



A technician installs a data-logging station for instruments placed in a stream. Water chemistry holds clues to the effect of landslides on the carbon cycle.

PHOTO: JENS TUROWSKI

THE TEAM WILL ALSO TACKLE a deeper mystery: the invisible influence of landslides on the atmosphere. The exchange of carbon between the atmosphere, the surface, and the oceans ultimately regulates Earth's habitability. For now, humanity—through industrial and agricultural emissions—is a dominant force in the carbon cycle. But over geologic time, the interaction of water with rocks freshly exposed by erosion—so-called chemical weathering—is another powerful player. And landslides are catalysts that speed up chemical weathering. They exhume fresh rock and grind it down into smaller pieces, creating more surface area for reactions. They also carve depressions in slopes that funnel rainwater into the rock.

Here in Taroko, Hovius and his team hope to learn whether landslide-driven weathering is releasing CO_2 into the atmosphere or drawing it down, doing a small part to counter humanity's influence.

It all depends on the rocks and acids available. When atmospheric CO_2 dissolves in rainwater, it forms carbonic acid, a weak acid that reacts with silicate rocks—the sandstones, granites, and others that form the majority of Earth's crustal rocks. The reaction liberates bicarbonate ions, which wash down rivers and into the ocean. There, the ions are taken up as calcium carbonate by shell-forming marine organisms. When these die, their skeletons sink to the sea floor, locking up carbon for millions of years in deposits of limestone. Were it not for this steady drawdown of carbon, the carbon emissions from volcanoes would, over the long term, turn Earth into a hothouse, says Jérôme Gaillardet, a geochemist at the Institute of Geophysics in Paris. "Life is possible because we have this process."

When limestone and other carbonate rocks are present, however, they can have the opposite effect, providing sulfide minerals like pyrite also exist nearby. Sulfides react with water and oxygen to form sulfuric acid, which in turn dissolves carbonate rocks to release CO_2. About one-quarter of the rocks in Taroko are carbonates.

By analyzing river water and landslide seepage for ions let loose by these two kinds of weathering, scientists can see which one dominates. Hovius and his colleagues plan to sample water from recent landslide scars, slopes that appear to be close to failing, and landscapes that haven't budged in hundreds or even thousands of years. The results, they hope, will show how chemical weathering varies across different settings.

Robert Emberson, a landslide researcher at NASA Goddard Space Flight Center in Greenbelt, Maryland, has studied fresh landslides in Taiwan and shown that they are sources of CO_2—right now. He notes that carbonate rocks and sulfuric acid react thousands of times faster than silicate rocks and carbonic acid. But over centuries the reactions can deplete the sulfide minerals or carbonates, paving the way for silicate weathering to take over, although the transition is gradual, says Kuo-Fang Huang, an isotope geochemist at Academia Sinica in Taipei who is involved in the observatory. "It's much more complicated." To see how weathering changes over time, Aaron Bufe, a GFZ earth scientist, plans to gauge the net weathering effect of landslides of different ages—in Taroko and elsewhere—that occurred in similar rock types and climates.

Whether Taroko National Park, and Taiwan as a whole, is a net source or sink of carbon hinges on other factors, too. For instance, landslides mobilize tree trunks and other plant material, kick-starting their journey toward rivers and ultimately the ocean. Rapid burial of this organic matter, before it can decay, removes carbon from the atmosphere. Hovius and his colleagues may monitor this process by recording logs washing downstream after landslides. "The full [carbon] budget remains to be determined," he says.

Hovius is eating a takeout lunch next to the Liwu River, the familiar scarred mountainsides towering overhead. He doesn't mind the prospect of more time in this grand outdoor laboratory. He points to where he lost an underwater acoustic sensor after a storm buried it in 5 meters of sediment. He runs his hand over the rocks he's sitting on and finds evenly spaced holes, relics of a study he did here 20 years ago. Studying this stretch of the river alone could yield 10 papers, he says. But he's got loftier goals. "I really want to understand the whole thing at once." ■

Katherine Kornei is a journalist in Portland, Oregon. Her trip to Taroko National Park was supported by a science journalism fellowship from the European Geosciences Union.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS



DOUBLE DIP

An NIH program pays school debt to keep scientists in academia. Many break the rules by also taking industry money

By **Charles Piller**

Joshua Bauml, a University of Pennsylvania (UPenn) oncologist, received welcome news last year: He had won a coveted spot in the National Institutes of Health (NIH) Loan Repayment Program (LRP). Since 1988, the program has aimed to keep promising young biomedical scientists in academic research by helping repay school loans that can run up to hundreds of thousands of dollars. Without it, supporters say, many of these researchers might have chosen lucrative slots at pharmaceutical companies or in private practice.

Bauml, who finished medical school in 2008, says he will get $75,000 in tax-free loan repayments during his first two LRP years and hopes a renewal will further pay down his debt. He also became a program "ambassador"—a model of success who counsels applicants.

Bauml has another distinction, too. A *Science* investigation has shown that he, like more than one-third of the 182 clinical scientist ambassadors whose years of participation could be determined, broke the program's rules against certain forms of industry funding. (NIH's own analysis finds a smaller number of violations.) The extensive rule breaking—inadvertent in Bauml's case, he says—is leading some observers to question the public purpose and fairness of a generous taxpayer subsidy for young scientists. "In this presumably exemplar program … you see widespread flouting of the rules, and no one seems to care," says Vinay Prasad, a hematologist-oncologist at Oregon Health & Science University in Portland who studies industry influence on medical research.

Since 2004, the LRP has paid nearly $1.1 billion for more than 22,000 new or renewal awards to researchers, including about $72 million in fiscal year 2018. Ericka Boone, its director for the past 4 years, said via email and in interviews that the program supports "mission-critical" NIH priorities by preventing school debt from creating "a vacuum of research innovation and discovery." Recipients, she added, have excelled in gaining funding; publishing research in basic science, clinical trials, and

ILLUSTRATION: STEPHAN SCHMITZ/FOLIO ART

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

public health; and serving in institutional leadership roles up to 15 years after receiving federal benefits.

Other taxpayer-funded student loan forgiveness plans have a similar rationale: They free recipients from debt so that they can serve the public interest. But these programs tend to be highly competitive and forcefully policed, and often require lengthy government or public service commitments. A Department of Education program for public servants requires at least 10 years in government or nonprofit jobs and awards benefits very selectively.

The LRP offers a sharp counterpoint. Applicants compete on scientific merit, but half win acceptance. The LRP generally does not consider financial need. Although many of the academic-scientist recipients earn relatively modest salaries, taxpayers also repay school loans for physician researchers who rank in the top 1% to 2% of all U.S. wage earners.

Over its history, the program has at times allowed recipients to accept industry support—either by direct payments or research funding—in unlimited amounts. As a result, the LRP in effect subsidizes some young scientists as they build strong ties to pharma. In addition, lax policing and confusing rules about acceptable forms of corporate support mean many recipients end up reaping even more industry largesse than the rules allow.

Bauml accepted more than $28,000 in pharmaceutical industry money, mostly for consulting or speaking at pharma events, during his first 6 months in the program. That was within the rules: Although the program previously barred such payments from drug companies, NIH reversed the policy in 2017. During the same period, industry also underwrote $473,000 of Bauml's research. For Bauml, this was business as usual. Drug firms had previously given at least $3.2 million for his work, in a steady flow leading up to his LRP award. The practice was allowed until 2017, but the latest program rules banned it, placing him in violation.

Bauml says he was shocked to learn from *Science* that he had broken LRP rules. He says he disclosed everything to NIH, and heard no objection. (NIH declined to verify disclosures by any recipient.)

Prasad calls the LRP's permissive approach "ethically problematic." He and others say the program should be more vigilant about industry payments to the researchers it helps. "Something designed to keep early-career investigators in academia can instead have the function of basically being a supply pipeline to industry," says Leigh Turner, a bioethicist at the University of Minnesota in Minneapolis. The result, he says, can be long-term financial dependence, regardless of whether a researcher stays in academia.

The program is a lost opportunity, critics say. "Industry influence on young scientists is vast and pervasive," Prasad says. "The intent of the LRP should be to assist people who pursue something that is very challenging in academic medicine—building a career that is truly independent of conflicted bodies."

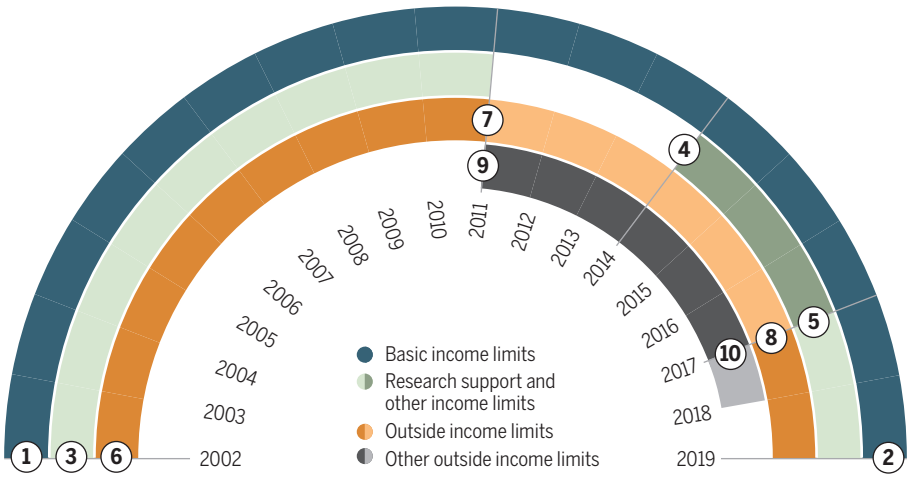
ALTHOUGH THE LRP supports a variety of biomedical fields, *Science*'s investigation focused on physicians because drug industry payments they receive must be publicly disclosed. Doctors make up more than half of recipients, and the relevant records—beginning in mid-2013—are posted on the federal Open Payments website. Disclosure statements on articles in journals, which often require authors to declare industry support, provided additional data.

Before 2011, LRP participants were free to accept all industry funding. Stephen Wiviott, a Harvard University cardiologist and an LRP awardee in the mid-2000s, wrote in an email that the LRP paid his entire $100,000 medical school debt, helping him buy a home and start a family. It "allowed me to spend more time on my research ... and to build a portfolio of publications that facilitated my first job, and ultimately set me on the path for a lifetime career," he wrote. At the same time, Wiviott earned consulting, advisory, and speaking fees from at least 12 companies, and worked on research underwritten by at least 24 firms, according to published disclosures, which did not specify amounts received.

Later, the program imposed a changing array of restrictions, whose evolution can be seen on archival versions of the program website (see graphic, above). From 2011 to 2014 it continued to permit any research funding from for-profit sources. In 2014, the Code of Federal Regulations barred such funding for LRP recipients. But the program continued

Labyrinth of rules

The National Institutes of Health Loan Repayment Program has imposed shifting restrictions on industry payments to recipients, applying to both personal income and research support. Recipients and program managers have struggled to keep track.



Basic income limits

1 Applicants must be employed for research at least half-time by a nonprofit entity, and some research funding must come from a nonprofit or government entity.

2 "Primary employment must be with a nonprofit institution. ... You do not receive salary from a for-profit institution/contractor or the federal government for engaging in the research project."

Research support and other income limits

3 None on salary or research.

4 Awardees must receive a majority of their support from nonprofit sources.

5 "Ineligible individuals" include "an individual who receives any research funding support or salary from a for-profit institution or organization."

Outside income limits

6 None on supplemental income.

7 "Any employment or income outside of the nonprofit institution must be separate from your research and cannot represent a significant amount of total income."

8 None on supplemental (nonresearch) salary or other income.

Other outside income limits

9 "Non-Eligible individuals" include "individuals who derive any income from any for-profit source or from a private practice."

10 "Any employment or income outside of the nonprofit institution must be separate from your research and cannot represent a significant amount of total income."

CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) NATIONAL INSTITUTES OF HEALTH

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

to allow corporate research support, capping it at less than half of a recipient's total, until it finally adopted the required ban in 2017. Personal income from for-profit sources followed a reverse pattern: From 2011 until 2017, it was strictly off limits, but since then, any amount has been allowed.

Boone declined to explain the policy shifts. But she said allowing awardees to accept industry money is not in conflict with the program's goal of keeping scientists in academic research. Current LRP restrictions on industry funding, Boone added, simply follow federal requirements.

She gave shifting accounts of past rules, however, insisting at first that the LRP has never permitted industry support for research or personal income. In later written comments, Boone acknowledged that assertion was wrong. But she continued to contradict the program's rules as published on its website. She said the LRP did not restrict moonlighting income from 2011 to 2017. Yet published rules for 2011 to 2017 barred income of any kind from for-profit sources or private clinical practice, and stated that income from nonprofits other than the recipient's regular employer could not be "significant."

Boone also seemed uncertain about the purpose of the program. She initially told *Science* that the LRP looks for scientists "who really had much more financial-based need—where it would really be important for them to receive this assistance or they would probably end up … getting out of research."

Yet, other than a track for individuals from disadvantaged backgrounds—less than 2% of awards since 2004—the LRP uses no means test other than total debt of at least 20% of a recipient's base salary when accepted to the program. Salaries for many recipients at public universities exceed $200,000, and elite private institutions, which employ many LRP recipients, often pay far more. In later, written comments, Boone conceded that payments from the program "are not need-based awards."

THE RULE CHANGES may explain some of the widespread violations apparent in records from mid-2013 through 2018. Among 182 LRP ambassadors, *Science* found nine violations of program rules involving research funds—from about $7000 to $473,000—and more than 75 involving consulting fees or other personal income, including a dozen from about $12,000 to $59,000. Academic papers by some LRP recipients also disclosed improper payments in unspecified amounts prior to 2013.

Boone would not comment on specific findings, but in an interview she told *Science* that violations are rare. The program relies on self-reporting by recipients, and regular validations from their research mentors and employers. "No system is foolproof. But we have a pretty good system," she said. In recent years, violations have been found only once or twice, and were due to circumstances beyond recipients' control, she said, such as a nonprofit hospital employer being taken over by a for-profit chain. Boone said the LRP has found no violators since 2017, despite the numerous cases *Science* identified.

Later, Boone reversed herself, telling *Science* that by her reckoning—which relied on a chronology of rules that is at odds with those the LRP published at the time—34 of the roughly 400 LRP ambassadors who could be tracked on Open Payments were in violation since 2014. When *Science* applied Boone's chronology to the sample of 182 ambassadors, the number of violators and the amounts accepted improperly for research actually jumped sharply. (After *Science* asked about apparent violations by some publicly posted ambassadors and "success stories," NIH protected their web pages behind passwords.)

> **"For this pattern [of violations] to emerge, many people have to be ignorant of or indifferent to the rules."**
>
> **Vinay Prasad**,
> Oregon Health & Science University

Genevieve Kanter, an economist and ethicist at UPenn's Perelman School of Medicine, whom *Science* asked to consider its findings and LRP's policies, says she isn't surprised that both administrators and awardees are confused when rules are "chaotic" and seem to lack "underlying principles."

"For this pattern [of violations] to emerge, many people have to be ignorant of or indifferent to the rules," Prasad adds. "Not only the beneficiaries. It's also the … people who are signing off on these agreements with NIH, and NIH officers who may even be aware of some of this and turn a blind eye," says Prasad, who benefited from a small, parallel LRP as an NIH employee.

Adriane Fugh-Berman of Georgetown University in Washington, D.C., who studies how the pharmaceutical industry influences medicine, says that by permitting industry funding and failing to enforce limited restrictions consistently, the LRP effectively subsidizes industry research. "When the biomedical industry controls the research agenda, it restricts the questions being asked to those with answers that are potentially profitable … excluding research on older therapies, on generic therapies, on nonpharmacologic therapies, on social influences on illness," Fugh-Berman says.

Bauml and LRP ambassador Scott Tagawa at Weill Cornell Medicine in New York City argue that industry money is crucial to certain lines of medical research, and that the LRP should actually relax its current restrictions on industry support for recipients' work. Tagawa, for example, studies cancer drugs for which industry may be the only research funder. Often, a company provides funds for a clinical study or donates a drug so study subjects don't have to pay out of pocket, Tagawa says. Keeping such support off limits to LRP recipients "doesn't make a lot of sense to me," he says.

Tagawa had robust ties to industry while he was an LRP recipient, from 2009 through fiscal year 2013. In 2011 and 2012, Sanofi U.S., Amgen, and Janssen paid him consulting or speaking fees in undisclosed amounts, and in 2013 he made more than $36,000, mostly for consulting and speeches, according to published disclosures and Open Payments. Companies "made the presentation slides" for his talks, he says, calling it "basically the same thing as being a sales rep." LRP rules at that time barred such income—something Tagawa says he didn't realize, noting that he fully disclosed his earnings. He says he stopped giving the speeches after Cornell proscribed them, but doesn't regret taking payments for those "educational" talks.

LRP alumni often retain close ties to industry later in their careers. *Science*'s investigation identified substantial industry support for dozens of LRP ambassadors, including Tagawa, after their loans were repaid—in amounts per person of up to $29 million for research and $159,000 in personal income—according to Open Payments data from mid-2013 through 2018.

Yet some LRP recipients have built robust research careers without turning to industry. Another LRP ambassador and current beneficiary, Adrienne Boire, is a biochemist and neuro-oncologist at Memorial Sloan Kettering Cancer Center in New York City who studies brain cancer cells within spinal fluid. Throughout 14 years of postgraduate training, during which she gave birth to her daughter, Boire amassed debts of $250,000. "The loan repayments made all the difference," she says.

Boire, whose current clinical trial involves an off-patent drug, says she avoided industry research funding because "I really like … the intellectual freedom you have when you're in a pure academic system." ■

This story was supported by the *Science* Fund for Investigative Reporting.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS



Downloaded from http://science.sciencemag.org/ on November 28, 2019

ETHNOMUSICOLOGY

The world in a song

Similar songs are used in similar contexts around the world

By **W. Tecumseh Fitch** and **Tudor Popescu**

Although all human cultures appear to create music, the music of different cultures is incredibly varied, leading some scholars to question whether music is really, as Henry Longfellow claimed in 1835, a universal "language" of our species. If true, it would suggest that universal cognitive mechanisms exist that can both explain the unity and allow the diversity of the world's musics. Do such universal mechanisms exist? If so, can we investigate them empirically? On page 970 of this issue, a multidisciplinary team led by Samuel A. Mehr presents a major step forward in this enterprise, combining the methods of modern data science with musical recordings and ethnographic records to provide an insightful overview of universal principles underlying sung music (*1*). Building on a new collection of song recordings and ethnographies from around the world called the Natural History of Song (NHS) database (*2*), the authors find that not only is music universal (in the sense of existing in all sampled cultures) but also that similar songs are used in similar contexts around the world.

The empirical quest for musical universals has a checkered history. It started with a bang in 1900, when after hearing a group of Thai musicians perform in Berlin, psychology professor Carl Stumpf decided to use his newly acquired phonograph to record them, and later conducted perceptual experiments with the same musicians (*3*). Stumpf was fascinated by both the differences and similarities between Thai and European music, and his key goal was to seek universal musical principles despite this diversity. His recordings provided the seeds of the Berlin Phonogramm-Archiv, a massive collection of recordings of non-Western and folk music that by 1933 had swelled to 13,300 phonograph cylinders (*4*). Stumpf hired Erich von Hornbostel and Curt Sachs to curate the collection, and together they founded a thriving but short-lived Berlin School of Comparative Musicology.

The search for "universality" in this context does not suggest that precise copies of melodies or rhythms would be shared across cultures, but rather refers to deeper cognitive principles of human "musicality" (*5*) that could explain broader patterns,

Department of Cognitive Biology, University of Vienna, Austria. Email: tecumseh.fitch@univie.ac.at

PHOTO: BRYAN AND CHERRY ALEXANDER/SCIENCE SOURCE

Published by AAAS



Timania Petaulissie and Haunaq Mikkigak are Inuit throat singers from Cape Dorset, Baffin Island, Nunavut, Canada.

just as for linguistic universals (*6*). Unfortunately, with the rise of Nazism, the key members of this school (who were mostly Jewish) were forced to flee Berlin, and this research program was destroyed (*4*).

After this ill-fated foray, the search for musical universals fell by the wayside, and by the 1970s ethnomusicologists were discouraged from even discussing musical "universals." Cultures and their musics were so diverse, it was said, that it was pointless to compare them (*7*), and both the empirical search for universal principles and the evolutionary perspective of the Berlin School were viewed with distaste by many, particularly in the English-speaking world. So with few exceptions—most prominently Alan Lomax's visionary "Cantometrics" project (*8*)—comparative musicology languished until recently, when a few brave researchers again began to attack these fundamental questions using new methods (*9*–*11*).

The new work by Mehr *et al*. analyzes songs performed by a carefully selected sample of human cultures spanning the planet, together with detailed ethnographic descriptions of the cultures and the song contexts. They analyze vocally performed songs because the voice is the most fundamental and ever-present musical "instrument" and song is a core component of human musicality (*12*). To ensure a fair and unbiased sample of cultures, the NHS builds on the well-documented Probability Sample File (https://hraf.yale.edu), which defines a stratified random sample of cultures, allowing relatively confident generalization to all of the world's cultures.

Employing the method of Bayesian principal components, they find that three main dimensions—formality, arousal, and religiosity—account for considerable variance in these contexts. They then analyzed recordings of four specific song types—lullabies, dance songs, love songs, and healing songs, selected on the basis of previous research (*2*)—finding many detailed examples of acoustic regularities.

Some of these regularities are unsurprising (for example, that dance songs are faster and more rhythmic than lullabies), and some are more intriguing (for example, that ritual healing songs are less melodically variable than dance songs). These broad, universal acoustic patterns are easily identified by naïve Western listeners, who successfully categorized the song type of sound recordings. The listeners' familiarity with world music played a minor and dispensable role in their correct classification. Furthermore, on the basis of ethnographic records, acoustically similar song types occur in certain shared contexts, and not others, across the world.

Crucially, variability of song context within cultures is much greater than that between cultures, indicating that despite the diversity of music, humans use similar music in similar ways around the world. Additionally, the authors found that the principle of tonality (building melodies from a small set of related notes, built upon a base tonic or "home" pitch) exists in all cultures. This suggests the existence of a universal cognitive bias to generate melodies based on categorical building blocks (*13*).

A critic might protest that with so much data, some apparent patterns are inevitable, but control analyses, using geographically appropriate climate data, or nonmusical ethnographic data, failed to reveal comparable results. This demonstrates that the authors have found bona fide links between musical acoustics and cultural phenomena. The authors also show that apparent deviations in the published record may result from underreporting, rather than true absence of the supposed-universal musical feature. Taken together, these new findings indicate that some basic but fundamental principles mapping musical styles onto societal functions and emotional registers exist and can be scientifically analyzed.

Despite this auspicious beginning, there is much left to do. Given the current database and analyses, considerable variance in both societal contexts and acoustic variables is left unexplained. Although some of this unexplained variance may be culture specific, some variants may reflect cultural affinities due to common descent, and other environmental contingencies [for example, the existence of yodeling in high mountain communities in both Europe and New Guinea (*14*)]. Furthermore, the NHS database only includes vocal song, but a massive playlist of instrumental and rhythmic music is still unexamined and remains to be analyzed in similar ways (*9*, *10*). There also are thousands of world cultures whose musics are not (yet) part of the database, and other musical styles to be sampled beyond the four types proposed and analyzed here. Today, with smartphones and the internet, we can easily imagine a comprehensive future database, including recordings of all cultures and styles, richly annotated with video and text, being assembled in a citizen science initiative.

Overall, the approach of Mehr *et al*., fusing data science, anthropology, and psychology, offers an exciting way to tackle age-old problems in musicology, promising deep insights into the putative psychological universals that underlie the diverse musics humans create and enjoy. A deeper understanding of these, at both cognitive and neural levels, would offer rich new insights into the cognitive biology of our species. ■

REFERENCES AND NOTES

1. S. A. Mehr *et al*., *Science* **366**, eaax0868 (2019).
2. S. A. Mehr, M. Singh, H. York, L. Glowacki, M. M. Krasnow, *Curr. Biol.* **28**, 356 (2018).
3. C. Stumpf, *Tonsystem und Musik der Siamesen* (Barth, Leipzig, 1901).
4. A. Schneider, *Ethnomusicology* **50**, 236 (2006).
5. H. Honing, C. ten Cate, I. Peretz, S. E. Trehub, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140088 (2015).
6. W. T. Fitch, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 376 (2011).
7. B. Nettl, *The Study of Ethnomusicology: Twenty-Nine Issues and Concepts* (Univ. of Illinois Press, 1983).
8. A. Lomax, *World of Music* **19**, 117 (1977).
9. S. Brown, J. Jordania, *Psychol. Music* **41**, 229 (2013).
10. P. E. Savage, S. Brown, E. Sakai, T. E. Currie, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8987 (2015).
11. S. E. Trehub, in *The Origins of Music*, N. L. Wallin, B. Merker, S. Brown, Eds. (MIT Press, 2000), pp. 427–448.
12. W. T. Fitch, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 2014.0091 (2015).
13. T. Verhoef, *Lang. Cogn.* **4**, 357 (2012).
14. E. M. von Hornbostel, in *Bericht über den musikwissenschaftlichen Kongress Basel 1924*. (Breitkopf & Haertel, Leipzig, 1925), pp. 203–210.

10.1126/science.aay2214

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

EVOLUTIONARY BIOLOGY

The multistep morphing of beetle horns

Genes that specify insect wings initiate horn development in dung beetles

By **H. Frederik Nijhout**

As with most complex problems, the simple, easy-to-understand hypothesis that novel traits arise from new genetic mutations is wrong. Although scientists have known this for decades, an alternative explanation has yet to present itself. On page 1004 of this issue, Hu *et al.* (*1*) show that deciphering the mechanisms underlying the evolution of a new biological trait is intimately intertwined with our evolving understanding of the concept of homology and its role in developmental genetics.

Darwin's postulate of descent with modification implies that all biological traits must have an evolutionary antecedent—that is, every trait is structurally related to a preceding one. This concept of homology among traits was developed more than a decade before Darwin's major works by paleontologist Richard Owen, who defined it as: "…the same organ in different animals under every variety of form and function." This seems clear enough until you parse the language and ask what exactly is meant by "the same." Owen expanded on his original definition by noting that homology among traits could be recognized by correspondences in their structure, position in the body, and embryonic development.

Paradoxically, as more has been learned about the development of biological structures, it has become increasingly difficult to recognize homologies. Part of the reason is that most structures are built from preexisting parts with different developmental and evolutionary histories. This conundrum has led to the concepts of partial homology and homology of a selected attribute of a structure—but not the structure as a whole—with a feature in an ancestor (*2*–*4*). Modern developmental genetics has made the task of identifying homologies even more fraught by revealing that a surprisingly small toolbox of regulatory genes controls the development of exceptionally diverse and seemingly unrelated structures (*5*).

Hu *et al.* investigated the developmental and genetic basis of an evolutionarily novel structure, the prothoracic horns of dung beetles (members of the family Scarabaeidae). The prothorax is the anteriormost of the three thoracic segments of insects. All three segments have a pair of ventral walking legs, and the second and third also bear a pair of dorsal wings. Among the great and continuing puzzles in insect evolution are the origin of the wings and why the prothorax does not have any.

The dorsal prothorax has, however, undergone a tremendous diversification of structure in various lineages of insects, ranging from elaborate three-dimensional helmets in treehoppers; great crests in wheel bugs; crests and long spikes in some grasshoppers; and single, double, triple, or quadruple horns of scarabaeid beetles—a remarkable example of a secondary sexual trait in the animal kingdom. Wings and legs are thought to have evolved from a single common precursor that split into a dorsal part and a ventral part for the wing and leg, respectively. Moreover, recent work has shown that the wing also has a composite origin, deriving from parts of the dorsal and lateral plates of the thorax. This dual origin of the wing, it turns out, plays an important role in the developmental origin of horns.

Hu *et al.* began their elegant series of experiments by using RNA interference (RNAi) in scarabaeid beetle larvae and pupae to inhibit the expression of several genes known to be associated with wing development. They found that the beetle's wings were indeed reduced or lost completely. More surprising, however, was the observation that the large medial prothoracic horn was also reduced in size and split into two small bilateral projections. The authors then showed that *vestigial*, a selector gene for wing development, was expressed throughout the developing wings (as expected) but also at the base of the horn precursors in the pupal stage.

Hu *et al.* also found that the medial prothoracic horn begins its development as two lateral protuberances during the pupal stage that gradually migrate to the middle of the pronotum. These findings led the authors to conclude that wing-related genes are required for initiating horn development and that, therefore, the prothoracic horns might share partial homology with wings.

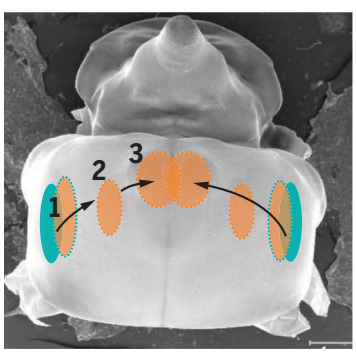
Hu *et al.* then used RNAi to inhibit expression of the *Sex combs reduced* (*Scr*) gene, whose normal activity specifies the prothoracic segment and in whose absence this segment acquires the characteristics of the second tho-

Department of Biology, Duke University, Durham, NC 27708, USA. Email: hfn@duke.edu

Dorsal wing field drives development of beetle horn

Panels show the steps in the development of a horn in dung beetles. The horn develops at the larval stage, beginning in specified regions of tissue in the insect dorsal-wing development field.

● Dorsal wing field ● Ventral wing field



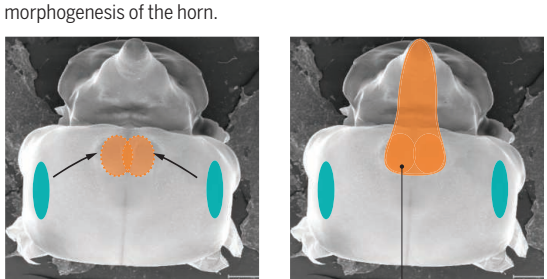
Wing-related genes initiate growth and medial migration of beetle-horn precursor structures.

1 Wing gene activation
Initial differentiation and migration

2 Wing gene inactivation

3 Fusion growth and morphogenesis

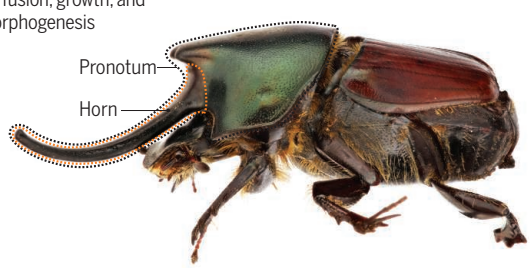
After migration, a different gene network drives fusion of beetle-horn precursor structures and subsequent growth and morphogenesis of the horn.



Wing gene inactivation

Activation of new gene network for fusion, growth, and morphogenesis

Beetle horn growth and morphogenesis



GRAPHIC: N. CARY/*SCIENCE*. (PHOTOS, TOP TO BOTTOM) Y. HU *ET AL.*, ALEXANDER WILD

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

Corrected 22 November 2019. See full text.

racic segment, which bears the forewings. This resulted in the development of structures that resembled the normal forewing (elytra), as well as a severe reduction in the size of the prothoracic horn, which was again reduced to two small lateral protuberances. Not all transformations of the thorax were equally severe. In fact, Hu *et al.* found a reciprocal relationship between the size of the wings and the size of the two horns that developed when the expression of individual wing-related genes was inhibited by RNAi; large wings appeared to develop at the expense of large horns and vice versa. Inhibiting expression of the *pannier* (*pnr*) gene, which specifies the medial region of the notum (the dorsal part of the insect's thoracic segment), confirmed the hypothesis that thoracic horn primordia provide the material from which the ectopic wings developed.

Hu *et al.* then conducted a series of RNA-sequencing studies to determine what kinds of genes were expressed in the developing horns and adjacent tissues. They found that soon after horn development started, wing-specific gene expression ceased and further development of the horns was accompanied by new gene expression patterns. The overall conclusion is that horns arise from tissues that would have formed the dorsal component of the wing and use part of the wing's gene network to get started. However, subsequent growth and morphogenesis of the horn use a diverse set of genes unrelated to wing development (see the figure).

The single medial horn of these beetles thus appears to come about by the migration and fusion of two lateral primordia derived from the dorsal wing field. It seems possible, therefore, that beetles that have two horns fail to merge their primordia, and those with four horns split the two primordia. The flexibility of using wing genes to initiate a primordium and then building on that foundation by selecting for genetic networks that promote specific morphogenesis might well have been the foundation for the elaborate and sometimes riotous diversification of form seen in the pronotum of insects. This raises a new question: What is the primitive function of what is now called the wing gene network? Is it to specify wings, or are wings latecomers in a network that evolved to regulate various outgrowths on the thorax and abdomen? Novelty may be in the eye of the beholder. ■

REFERENCES AND NOTES

1. Y. Hu, D. M. Linz, A. P. Moczek, *Science* **366**, 1004 (2019).
2. V. L. Roth, *Biol. J. Linn. Soc. Lond.* **22**, 13–29 (1984).
3. M. J. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford Univ. Press, 2003).
4. G. P. Wagner, *Homology, Genes, and Evolutionary Innovation* (Princeton Univ. Press, 2014).
5. S. B. Carroll, J. K. Grenier, S. D. Weatherbee, *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, 2004).

10.1126/science.aaz9010

NEUROSCIENCE

Compelled to drink: Why some cannot stop

Individual differences in aversive stimulus processing may explain compulsive alcohol drinking

By **Kimberly Nixon** *and* **Regina A. Mangieri**

Excessive alcohol drinking, a component cause of more than 200 diseases, is a leading cause of preventable death (*1*). The loss of control over alcohol drinking to the point of compulsion—consuming alcohol despite negative consequences—is the defining characteristic of an alcohol use disorder (AUD) (*2*). Almost 20% of adults worldwide engage in heavy alcohol drinking episodes in their lifetime (*3*), and only half of heavy drinkers in the United States are able to cut down or quit drinking when faced with adverse health consequences of their drinking (*2*). Neither the individual differences that drive compulsion nor the circuitry of compulsive alcohol intake are well understood. On page 1008 of this issue, Siciliano *et al.* (*4*) show individual differences in the activity of neurons that comprise a newly described brain circuit in mice during early alcohol experience. The activity of these neurons predicts the escalation of alcohol drinking to compulsive intake.

Why are some people able to control their alcohol drinking, but others are compelled to drink despite health, personal, and/or social consequences? People drink to excess for a variety of reasons, but as the animal model of Siciliano *et al.* demonstrates, not all heavy drinkers become compulsive. Some have suggested that a genetic predisposition underlies the likelihood to develop compulsive alcohol intake (*5*, *6*), and the authors take a major leap forward to show individual differences in compulsion, specifically the consumption of bitter and highly aversive, quinine-adulterated alcohol, at the level of neuronal activity. By using changes in intracellular calcium concentrations in neurons as a proxy for neuronal electrical activity during alcohol drinking, Siciliano *et al.* identified clusters of neuronal activity around the time when mice licked tubes of alcohol. These data provide exquisite detail on brain and behavior relationships that underlie the individual differences in compulsive alcohol drinking in mice.

The alcohol drinking–associated neuronal activity occurred in a circuit between the medial prefrontal cortex (mPFC) and dorsal periaqueductal gray (dPAG) of the brain stem. Multiple lines of evidence from animal models suggest a role of the PFC in compulsive alcohol drinking (*5*–*8*). The PFC lies just behind the forehead in humans and is responsible for executive functions, namely judgment, decision-making, and behavioral control. These functions are impaired in drug and alcohol addiction, especially in severe AUD (*9*). Siciliano *et al.* capitalize on this knowledge as a starting point for examining compulsion, led by a previous detailed anatomical study of both retrograde and anterograde tracers that revealed that this mPFC-dPAG circuit represents aversive stimuli in the brain (*10*). The PAG, best known for its role in pain, has reciprocal connections with many addiction-relevant regions, including the PFC (*11*). Recent discoveries extend the role of the PAG to punishment and aversion signals, both of which are aspects of compulsion (*11*). Thus, it is not surprising that this mPFC-dPAG projection has a role in alcohol compulsion.

Siciliano *et al.* found that the inhibition of this mPFC-dPAG circuit promotes compulsive drinking (see the figure). Prior studies posit that excitation of cortical-striatal pathways is required for compulsive drinking (*7*, *8*). The authors make a compelling case: Mimicking cell activity dynamics (lick-initiated photoinhibition of mPFC-dPAG projections) yielded a compulsive drinking phenotype. Conversely, lick-initiated photostimulation of mPFC-dPAG neurons acted as a punishment and produced lasting decreases in alcohol consumption. The authors argue that inhibition of this pathway disrupts the transmittance of an aversive signal, which would diminish punishment of responding and/or reduce the animal's avoidance behavior. Accordingly, activation of this circuit signals that something is aversive, which then promotes avoidance (*10*). Therefore, if alcohol inhibits this pathway, it should also disrupt transmittance of aversive signals. Indeed,

The University of Texas at Austin, College of Pharmacy, Division of Pharmacology & Toxicology, Austin, TX, USA. Email: kim.nixon@austin.utexas.edu

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

Corrected 22 November 2019. See full text.

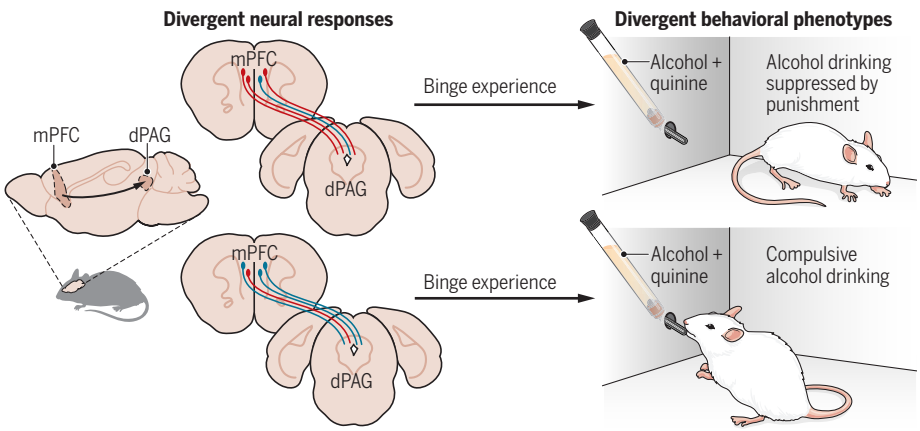
has antipunishment effects when injected into animal models just before an aversive event such as foot shock or stimulation of the PAG (*12*, *13*). Alternatively, this circuit may be underactive in compulsive mice, an effect that was not assessed directly in the study. Identifying the specific mechanisms of the reactivity of this circuit to alcohol and/or its potential constitutive underactivity in the compulsive phenotype are critical next steps to translate these findings into drug targets for AUDs.

Prior studies have examined how manipulating neural activity affects behavior once compulsive drinking has emerged, whereas Siciliano *et al.* observed individual differences in mPFC-dPAG neuron activity before compulsive drinking occurred. Notably, greater inhibitory activity in mPFC-dPAG neurons during early alcohol experience was only predictive of compulsive drinking after—and not before—a period of binge-like alcohol consumption. Thus, compulsive drinking may result from an interaction between a predisposition for alcohol-induced inhibition of cortical-to-brainstem signaling (reduced aversive processing) and experience-dependent enhancement of cortical-to-accumbal signaling (reward processing) (*5*). This point leads to the question of whether the inhibition of mPFC-dPAG neurons during alcohol consumption strengthens with repeated experience and/or dysfunction from alcohol-induced toxicity (*9*).

Despite the exciting results of Siciliano *et al.*, there are important caveats. As with any animal model, will the work translate to the human condition? Some researchers question whether quinine- or foot shock–resistant drinking in rodents actually models compulsion observed in AUDs (*14*). In humans, the negative consequences of drinking are rarely concurrent with drinking nor purely sensory but are dissociated by hours, days, or weeks from the drinking episode. Nevertheless, what matters is whether these models have predictive validity. Time will tell whether these "compulsive" alcohol drinking models are useful in identifying drug targets that are effective in humans with AUDs. With only three modestly effective drugs approved by the U.S. Food and Drug Administration to treat AUDs, this discovery is a needed leap forward—a newly-discovered circuit behind the essence of alcoholism in which to direct drug discovery efforts. ■

Activity of a neural circuit predicts compulsion to drink alcohol

A greater proportion of inhibitory (blue) versus excitatory (red) responses of the dorsal periaqueductal gray (dPAG)–projecting medial prefrontal cortex (mPFC) neurons during early alcohol experience corresponds to binge drinking–induced compulsive alcohol consumption.



REFERENCES AND NOTES

1. World Health Organization (WHO), Global status report on alcohol and health 2018 (WHO, 2018); www.who.int/substance_abuse/publications/global_alcohol_report/en.
2. National Institute on Alcohol Abuse and Alcoholism (NIAAA), Rethinking drinking: Alcohol and your health (NIAAA, 2016); https://pubs.niaaa.nih.gov/publications/RethinkingDrinking/Rethinking_Drinking.pdf
3. B. F. Grant *et al.*, *JAMA Psychiatry* **72**, 757 (2015).
4. C. A. Siciliano *et al.*, *Science* **366**, 1008 (2019).
5. V. Vengeliene *et al.*, *Addict. Biol.* **14**, 384 (2009).
6. C. Giuliano *et al.*, *J. Neurosci.* **39**, 1744 (2019).
7. T. Seif *et al.*, *Nat. Neurosci.* **16**, 1094 (2013).
8. E. N. Grodin *et al.*, *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 1022 (2018).
9. F. T. Crews, K. Nixon, *Alcohol Alcohol.* **44**, 115 (2009).
10. C. M. Vander Weele *et al.*, *Nature* **563**, 397 (2018).
11. D. T. George *et al.*, *Trends Neurosci.* **42**, 349 (2019).
12. J. R. Glowa, J. E. Barrett, *Pharmacol. Biochem. Behav.* **4**, 169 (1976).
13. P. Bovier *et al.*, *Pharmacol. Biochem. Behav.* **21**, 353 (1984).
14. F. W. Hopf, H. M. Lesscher, *Alcohol* **48**, 253 (2014).

ACKNOWLEDGMENTS

The authors are funded by the National Institute on Alcohol Abuse and Alcoholism: R01AA016959 (K.N.), R01AA025591 (K.N.), and U24AA016651 (R.A.M.).

10.1126/science.aaz7357

ORGANIC CHEMISTRY

A catalytic one-two punch

Cooperative catalysis enables the controlled synthesis of unnatural amino acids

By **John M. Ovian** *and* **Eric N. Jacobsen**

Amino acids are the building blocks of proteins, and chemists use these difunctional molecules for the laboratory synthesis of a wide variety of useful chiral frameworks (*1*). New and efficient methods for the synthesis of unnatural amino acids are always in demand to augment the numerous approaches that have been reported (*2*). On page 990 of this issue, Li *et al.* (*3*) report a strategy for the enantioselective synthesis of amino acids based on the cooperative action of an achiral transition-metal complex with a chiral hydrogen-bond donor to catalyze carbenoid insertions into the N–H bonds of aliphatic amines.

Cooperativity in asymmetric catalysis can be achieved either with multifunctional catalysts or through the synergistic operation of two different catalysts to access reactivity modes that are difficult to achieve or control otherwise (*4*). The principle has been applied with particular success in catalyst systems that can activate and position two different reacting partners simultaneously (see the figure, top). For example, an impressive system developed by Krautwald *et al.* (*5*) uses a chiral organocatalyst together with a chiral transition-metal complex to activate separate nucleophilic and electrophilic reacting partners simultaneously and generate coupling products with full control over multiple stereocenters. Other examples pair an achiral transition-metal catalyst with a chiral organocatalyst, inducing enantioselectivity through covalent or noncovalent interactions between the organocatalyst and either the reacting substrate or the cocatalyst.

A common feature in these transformations is that the key bond-forming step controlled by the cooperative catalyst system is also the enantiodetermining step, with both catalytic functions operating to construct

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. Email: jacobsen@chemistry.harvard.edu

Downloaded from http://science.sciencemag.org/ on November 28, 2019

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Published by AAAS

the one or more new stereocenters. By contrast, the key bond-forming step in the system developed by Li *et al.* is decoupled from the enantiodetermining step with an achiral transition metal promoting a C–N bond formation and a chiral organocatalyst subsequently effecting an enantioselective proton transfer to generate the desired stereodefined product.

The authors have applied the new tandem catalysis strategy to the enantioselective insertion of carbenes into the N–H bond of aliphatic amines (see the figure, bottom). The convergent coupling of an amine with carbenoids stabilized by an adjacent carboxyl group provides an attractive route to unnatural α–amino acid derivatives (*6*). Metal-catalyzed carbenoid insertion into N–H bonds was known but had not been extended successfully to aliphatic amines, which tend to bind unproductively and inhibit transition-metal catalysts (*7*). This reactivity problem was overcome through judicious selection of the ligand on a copper complex.

In particular, achiral hydrotris (pyrazolyl) borate (Tp) complexes promoted carbenoid insertions into aliphatic amines in high yields. Stereocontrol was achieved through the inclusion of a chiral amino thiourea derivative. An impressive variety of aliphatic amines engaged successfully in this transformation and furnished secondary and tertiary amino acid derivatives with high enantioselectivity. The protocol was leveraged to prepare intermediates in the synthesis of bioactive compounds and to perform late-stage derivatization of known drug structures.

On the basis of a kinetic analysis of the catalytic reaction, Li *et al.* surmise that the hydrogen-bond donor coordinates unproductively to the copper catalyst in the resting state. Dissociation is necessary in order to achieve rate-determining carbenoid generation at the metal. After insertion of the carbenoid into the N–H bond of the amine, they proposed that the copper complex dissociates to produce an achiral enol intermediate, which is then intercepted by the organocatalyst to promote the enantioselectivity-determining proton transfer. In this scheme, each catalyst promotes a distinct step in the overall transformation, with the transition metal catalyzing the construction of the C–N bond and the organocatalyst controlling formation of the stereocenter.

Computational analysis of the major and minor transition states, in concert with the observed influence of the copper-ligand complex on enantioselectivity, support the authors' hypothesis that the copper complex coordinates to the hydrogen-bond donor, enhancing its acidity and helping to promote the proton-transfer pathway. This reactivity principle inverts the more common scenario of a chiral organic ligand influencing reactivity at a transition-metal center.

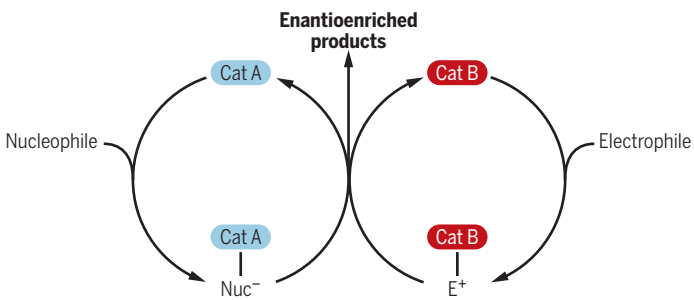
The cooperative action of achiral transition metal complexes with chiral hydrogen-bond donors holds enormous potential for achieving new asymmetric transformations. Organotransition-metal chemistry provides access to a wealth of reactivity modes inaccessible to organocatalysts, and chiral hydrogen-bond-donor catalysts have been found to promote enantiocontrol through a rich variety of noncovalent mechanisms. The system developed by Li *et al.* represents a compelling glimpse into some of the possibilities. ■

From amines to unnatural amino acids

A cooperative catalytic process developed by Li *et al.* converts aliphatic amines into unnatural amino acids.

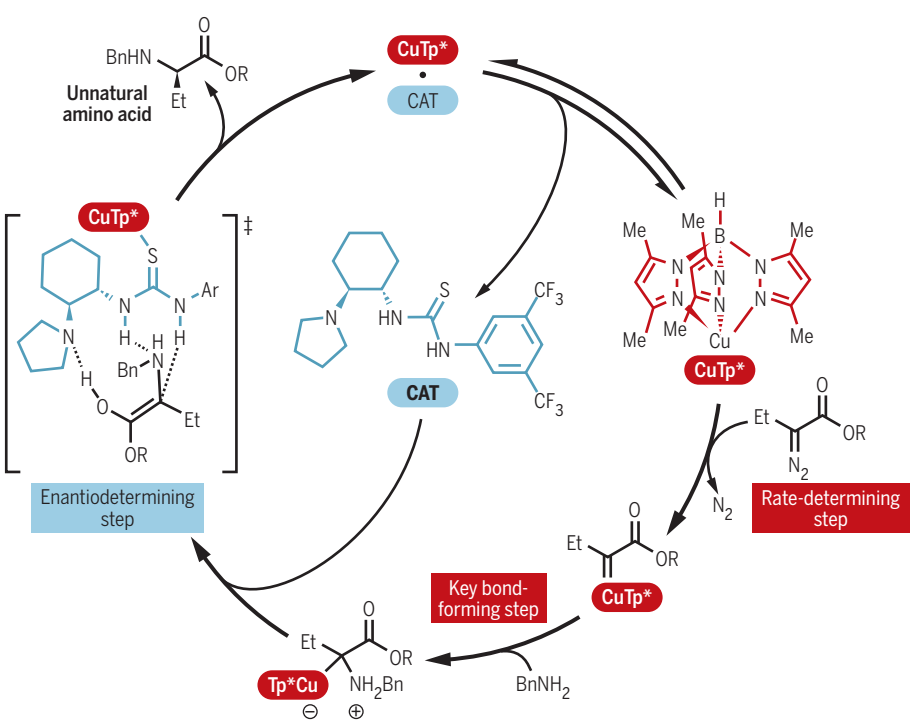
Dual activation

Activation of a nucleophile (Nuc) by one catalyst (Cat) and of an electrophile (E) partnered by a second catalyst can furnish desired enantiomers.



Coupling catalytic cycles

Distinct portions of a cycle for synthesizing unnatural amino acids from aliphatic amines (here, $BnNH_2$) are activated by two catalysts, CuTp* (shown in red) and chiral amino thiourea (CAT, shown in blue).



Ar, 3,5-bis(trifluoromethyl)phenyl; Bn, benzyl; Et, ethyl; Me, methyl; R, $C(Me)_2Ph$; Tp*, hydrotris(3,5-dimethylpyrazolyl)borate.

GRAPHIC: A. KITTERMAN/*SCIENCE*

REFERENCES AND NOTES

1. E. R. Jarvo, S. J. Miller, *Tetrahedron* **58**, 2481 (2002).
2. C. Nájera, J. M. Sansano, *Chem. Rev.* **107**, 4584 (2007).
3. M.-L. Li, J.-H. Yu, Y.-H. Li, S.-F. Zhu, Q.-L. Zhou, *Science* **366**, 990 (2019).
4. S. M. Banik, A. Levina, A. M. Hyde, E. N. Jacobsen, *Science* **358**, 761 (2017).
5. S. Krautwald, D. Sarlah, M. A. Schafroth, E. M. Carreira, *Science* **340**, 1065 (2013).
6. A. Ford *et al.*, *Chem. Rev.* **115**, 9981 (2015).
7. S. R. Hansen, J. E. Spangler, J. H. Hansen, H. M. L. Davies, *Org. Lett.* **14**, 4626 (2012).

10.1126/science.aaz6166

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

ECONOMICS

Unraveling the claims for (and against) green growth

Can the global economy grow indefinitely, decoupled from Earth's limitations?

By **Tim Jackson**[1] *and* **Peter A. Victor**[2]

American economist Kenneth Boulding famously quipped, "Anyone who believes that exponential growth can go on forever in a finite world is either a madman or an economist" (*1*). He was giving evidence to the U.S. Congress in 1973, in the wake of the Club of Rome's first, enormously influential and provocative report, *The Limits to Growth* (*2*). The remark has survived to this day as a somewhat satirical comment on the economics profession, but it also has a certain internal logic and provides a useful starting point for thinking about the "decoupling wars" that tend to be fought around the compatibility between economic growth and environmental limits (*3*).

When economists contend that growth can continue indefinitely, it is because in their view, growth is something measured in terms of economic value rather than material throughput. The preferred measure of output for economists—the gross domestic product (GDP)—is denominated in monetary value rather than in material weight. These things, they argue, are separable: By decoupling one from the other, economies ought to be able to escape the dominion of finite limits at least to any relevant degree (if not literally forever) (*4*).

Nobel laureate Paul Krugman goes as far as to suggest that physical scientists simply have a false conception of economic growth: "They think of it as a crude, physical thing, a matter simply of producing more stuff, and don't take into account the many choices—about what to consume, about which technologies to use—that go into producing a dollar's worth of GDP" (*5*). His conviction that these "many choices" will allow for even the most stringent ecological goals to be achieved without ever compromising economic growth leads him to denounce growth skeptics as "prophets of despair."

Krugman's argument is essentially an appeal to technology: more efficient processes, lighter and less polluting products, or a structural shift from materially intensive goods to materially light services, for instance. The importance of this kind of decoupling is not disputed, even by those who maintain that growth may not be feasible nor even necessarily desirable on a finite planet (*6*, *7*). What divides opinion, rather, is the question of whether a continuous decoupling might allow economic expansion to go on indefinitely.

It is useful to clearly distinguish between relative decoupling and absolute decoupling. The former refers to a decline in resource (or environmental) intensities, whereas the latter refers to an absolute fall in consumption or emissions (*8*). Put very simply, relative decoupling is about doing things more efficiently; because efficiency is one of the things that modern economies are supposed to be good at, decoupling has a familiar logic and a clear appeal to those who hope that growth can continue indefinitely. It isn't hard to find evidence for relative decoupling, even at the global level. For example, the carbon dioxide intensity of the global economy fell from about 760 g of carbon dioxide per dollar (g CO_2/$) in 1965 to less than 500 g CO_2/$ today, a decline of almost 35% in half a century (*6*).



TOMORROW'S EARTH
Read more articles online at scim.ag/TomorrowsEarth

But relative decoupling is barely half the story. An improvement in the emissions intensity of economic output does not necessarily mean that emissions themselves are falling. For this, absolute decoupling needs to occur, where emissions fall over time, even as economic output continues to rise. For relative decoupling to lead to absolute decoupling, the emissions (or resource) intensity must decline at least as fast as economic output rises. If the rate of decline in emissions intensity is greater than the rate of economic growth, then the level of emissions will decline. If not, then it won't.

It is not impossible to find some partial evidence for absolute decoupling over specific time periods, particularly when looking at data on a national or regional level. For instance, across the European Union, between 1990 and 2017, carbon emissions fell by 22% even as the economy grew by 58%, as measured on a territorial basis (*9*). Similar evidence can be found of both relative and absolute declines at the regional level in relation to material resource consumption (*10*).

One problem with this "partial" evidence is the porous nature of national and regional trade boundaries. In a globalized economy, territorial accounts of production-based emissions fail to take adequate account of a region's "footprint"—that is, the carbon emissions associated with a region's consumption patterns. The carbon footprint of the EU, for example, has fallen considerably more slowly, and remains 20% higher, than territorial emissions (*11*).

Such findings emphasize that, for a pollutant like carbon and for resources generally, it is what happens at the global level that counts; and here, there is no evidence of absolute decoupling at all. The amount of carbon dioxide going into the atmosphere today is more than 60% greater than the amount in 1990, despite the best efforts of the United Nations Framework Convention on Climate Change (*7*). Since the global financial crisis of 2008, the rate of growth in carbon dioxide emissions worldwide has slowed somewhat. Between 2014 and 2016, total global emissions seemed momentarily to have stabilized. But they rose again by 1.6% in 2017 and are estimated to have risen by a further 2.7% in 2018 (*12*).

There is another crucially important point here: Even absolute decoupling is not enough to ensure sustainability. Neither a stabilization in carbon emissions, nor a moderate decline in emissions, is enough to avoid a climate breakdown. The Intergovernmental Panel on Climate Change has estimated that to have a 66% chance of remaining below a 1.5°C temperature rise, there is a maximum available global carbon budget of 420 Gt of CO_2 that can be emitted into the atmosphere (*13*). At the current rate of emissions, this carbon budget would be exhausted within a decade.

In other words, decoupling GDP from the flow of emissions is not the same as decoupling economic activity from the stocks of environmental and material resources on which future prosperity depends. To achieve the latter, sufficient (*14*) or strong (*7*) absolute decoupling is needed. At the global level, sufficient absolute decoupling to prevent climate breakdown would require an average annual decline in the carbon intensity of global economic output of around 14% every year for the next three

[1]Centre for the Understanding of Sustainable Prosperity, University of Surrey, Guildford, Surrey, UK. [2]Environmental Studies, York University, Toronto, Ontario, Canada. Email: t.jackson@surrey.ac.uk

Downloaded from http://science.sciencemag.org/ on November 28, 2019

ILLUSTRATION: ADAM SIMPSON/HEART AGENCY

Published by AAAS



ILLUSTRATION: MICHAEL J. DIMOTTA

decades (*6*). The highest rate of decoupling ever achieved by the world's advanced economies was a little under 3%, in the years immediately following the oil crises of the 1970s. The average rate of decline across the world at the moment is less than 1%. In the case of a rich country like the United Kingdom, sufficient absolute decoupling would mean a decline in the nation's carbon footprint at a rate in excess of 20% each year, with a net zero target that might need to be as early as 2030 (*15*).

Proponents of so-called green growth—economic growth that uses natural resources in a sustainable manner—must show that it is possible to effectively eliminate carbon emissions from developed economies in the space of little more than a decade with no impact at all on economic expansion. This challenge cannot be answered solely by an appeal to technology. The question is not whether technological measures such as energy efficiency and solar power are possible (they clearly are); nor whether, in the past, countries have managed to harness these technologies sufficiently (they clearly haven't); but rather, whether countries can now achieve sufficient gains in a short enough time to allow the pursuit of economic growth indefinitely, while still remaining within the safe operating space of the planet.

In a sense, this once again raises the question of whether economic value is something completely separate from—or at least separable from—physical and material flows. Certainly, in the past, the two things have gone hand in hand. According to economics, monetary value surely has something to do with activity. According to physics, activity is impossible without the expenditure of energy. There may well be efficiencies to be had, but these will ultimately be constrained by thermodynamic limits, as all activity is. Those who believe that this is not a constraint on expansion typically appeal to the massive quantities of solar energy that flood Earth. But it remains true that these flows are diffuse (rather than concentrated, as fossil fuels are) and must be captured using material devices.

It still is not clear that this immediately rules out some form of growth. But it is clear that the larger the economy becomes, the more difficult it is to decouple that growth from its material impacts. One doesn't need thermodynamics to make this point. A bigger economy implies a bigger capital stock. A bigger capital stock means higher depreciation. An infinite economy (the ultimate outcome from eternal growth) means infinite depreciation and infinite maintenance costs. The only alternative would seem to be to begin assigning economic values to increasingly immaterial exchanges—love, friendship, the spoken word, perhaps—which seems both abusive and inflationary.

None of this is to suggest that decoupling itself is either unnecessary or impossible. On the contrary, decoupling well-being from material throughput is vital if societies are to deliver a more sustainable prosperity—for people and for the planet. ■

REFERENCES AND NOTES

1. *U.S. Congress, Energy Reorganization Act of 1973: Hearings, Ninety-third Congress, First Session, on H.R. 11510* (U.S. Government Printing Office, 1973), p. 248.
2. D. H. Meadows, D. L. Meadows, J. Randers, W. W. Behrens, *The Limits to Growth* (Universe, 1972).
3. European Environment Bureau, *Decoupling Debunked: Evidence and Arguments Against Green Growth as a Sole Strategy For Sustainability* (2019); https://eeb.org/library/decoupling-debunked/.
4. United Nations Environment Programme, *Decoupling 2: Technologies, Opportunities and Policy Options* (2014); www.resourcepanel.org/reports/decoupling-2.
5. P. Krugman, "Errors and emissions." *New York Times*, 18 September 2014; www.nytimes.com/2014/09/19/opinion/paul-krugman-could-fighting-global-warming-be-cheap-and-free.html?_r=1.
6. T. Jackson, *Prosperity Without Growth: Foundations for the Economy of Tomorrow* (Routledge, 2017).
7. P. A. Victor, *Managing Without Growth: Slower by Design, Not Disaster* (Elgar, ed. 2, 2019).
8. Organization for Economic Cooperation and Development, *Resource Productivity in the G8 and the OECD*; www.oecd.org/env/waste/47944428.pdf.
9. European Commission, *EU and the Paris Climate Agreement: Taking Stock of Progress at Katowice COP* (2018); https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0716&from=EN.
10. F. Krausmann, H. Schandl, N. Eisenmenger, S. Giljum, T. Jackson, *Annu. Rev. Environ. Resour.* **42**, 647 (2017).
11. G. P. Peters, S. J. Davis, R. Andrew, *Biogeosciences* **9**, 3247 (2012). Data available online from the Global Carbon Atlas, www.globalcarbonatlas.org/en/CO2-emissions.
12. R. B. Jackson *et al.*, *Environ. Res. Lett.* **13**, 120401 (2018).
13. Intergovernmental Panel on Climate Change, *Special Report: Global Warming of 1.5°C* (2018); www.ipcc.ch/sr15/.
14. K. Raworth, "Will these Sustainable Development Goals get us into the doughnut (aka a safe and just space for humanity)?" (2014); https://oxfamblogs.org/fp2p/will-these-sustainable-development-goals-get-us-into-the-doughnut-aka-a-safe-and-just-space-for-humanity-guest-post-from-kate-raworth/.
15. T. Jackson, *Zero Carbon Sooner: The Case for an Early Zero Carbon Target for the UK* (Centre for the Understanding of Sustainable Prosperity, 2019); www.cusp.ac.uk/themes/aetw/zero-carbon-sooner/.

10.1126/science.aay0749

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

SYNTHETIC BIOLOGY

Designer sense-response systems

Artificial chemically induced protein dimerization yields biological sensor-actuator devices

By **Roberto A. Chica**

Cells adapt to their changing environment to survive and grow. This requires the ability to sense and respond to chemical cues, such as changes in nutrient availability or signals from other cells. Biological sensing and response systems often involve networks of protein-protein interactions that transmit a signal triggered by molecular recognition of a specific chemical, ultimately resulting in adjustments to the cellular biochemistry. If it were possible to design protein pairs the interaction of which is mediated by the binding of a target small molecule, in turn resulting in a downstream cellular response, this would open the door to the creation of customizable biological sensors and actuators for a variety of medical and biotechnological applications. On page 1024 of this issue, Glasgow *et al.* (*1*) report a computational procedure for the design of modular sense-response systems based on protein complexes that can detect and respond to a user-defined small molecule.

Biological sense-response systems consist of an input module that interacts specifically with the target signal molecule (the sensor) and an output module that converts molecular recognition into a downstream effect (the actuator). Protein-based sensor-actuator devices can be created by genetically fusing preexisting proteins acting as input or output modules into a single polypeptide chain (*2*) or by engineering the input module of natural sense-response systems to detect different chemical signals (*3*). Although these approaches have been successful, the types of molecules that can be sensed and responses that can be obtained are limited. An alternative strategy is to exploit chemically induced dimers (CIDs), which are protein pairs that bind to each other in the presence of a specific ligand molecule. By fusing CIDs to split protein reporters—fragmented proteins that can reassemble and regain function when attached to interacting protein partners—a wide range of output signals can be achieved (*4*). However, as is the case with all protein-based sensor-actuator systems, the specificity of the CID input module must be engineered every time a new chemical signal is to be detected.

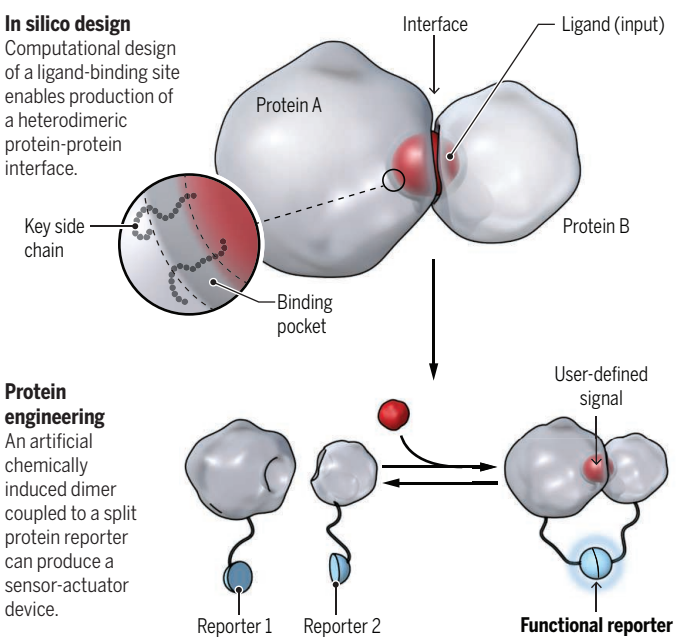
To facilitate the development of new and programmable sense-response systems, Glasgow *et al.* developed a generalizable computational procedure to create artificial CIDs that can respond to user-defined small-molecule inputs (see the figure). Their approach involves the de novo design of a binding cavity that is specific to a target small molecule at the interface of preexisting heterodimeric protein complexes. This objective is challenging to achieve because of the large number of positions and orientations in space that the small molecule could adopt relative to the protein dimer, and the huge number of possible amino acid combinations in the binding pocket that must be evaluated to identify those that can make the molecular interactions required for specific binding of the target molecule. To further complicate matters, the designed cavity must sufficiently destabilize the protein complex to disfavor heterodimerization in the absence of the signal molecule without causing unfolding of the individual protein monomers. Therefore, the computational design of artificial CIDs requires a careful balance of protein stability, dimerization ability, and ligand-binding activity.

As a case study, the authors chose to design a CID that can dimerize in the presence of farnesyl pyrophosphate, an intermediate in the biosynthesis of valuable terpenoid compounds that are used commercially as antimicrobials, fragrances, and flavors. The computational methodology for CID design consisted of four general steps: (i) geometrical definition of key interactions between protein side chains and the target molecule necessary for its specific binding, (ii) modeling of these geometries on a set of heterodimeric protein-protein interfaces to identify compatible scaffolds, (iii) optimization of binding pockets to further stabilize the interaction with the target molecule, and (iv) ranking of designs according to predicted ligand-binding affinity.

A total of nine designs generated with this methodology were tested experimentally. To do so, Glasgow *et al.* genetically fused the designed CIDs to a split mouse dihydrofolate reductase (*5*), an enzyme that is essential for cell survival, which served as the output module. By screening for survival of the bacterial *Escherichia coli* cells engineered to express the designed CIDs in the presence of an inhibitor specific to the bacterial dihydrofolate reductase, Glasgow *et al.* were able to evaluate changes in bacterial cell growth in the presence and absence of the farnesyl pyrophosphate ligand. If cells expressing the designed CID grew faster in the presence of farnesyl pyrophosphate than in its absence,

Design of a protein-based sensor-actuator device

Modular sense-response systems can be designed based on protein complexes that can detect and respond to a user-defined small molecule (ligand).



GRAPHIC: C. BICKEL/*SCIENCE*

Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada.
Email: rchica@uottawa.ca

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

the design was determined to be functional, as the target molecule would have induced dimerization of the CID and thereby restored the dihydrofolate reductase activity. Of the nine designs, two showed robust signal response to farnesyl pyrophosphate, and these were further improved using a combination of random and targeted mutagenesis. The final engineered CID bound specifically to farnesyl pyrophosphate in a concentration-dependent manner, and its experimentally derived three-dimensional structure in the presence of bound ligand was in very good agreement with the computational model. The engineered CID was also shown to work with a variety of output modules that produced enzymatic, fluorescence, or luminescence signals upon dimerization of the protein complex, demonstrating the modularity of these artificial sense-response systems.

The methodology described by Glasgow *et al.* provides a roadmap for the design of artificial CIDs for various biotechnological applications. Hitherto, the computational design of de novo ligand-binding sites into protein-protein interfaces had not been realized. Compared with previous sensor-actuator systems designed to be unstable and thereby degraded by the cell in the absence of the target ligand, and which respond by activating transcription of reporter genes (*6*, *7*), these CIDs have the benefit of not requiring the eukaryotic transcriptional activation and protein degradation machinery for actuation. This advantage, coupled with the modularity of their output signals, should make these artificial CIDs amenable to a broad range of applications.

Although successful, the design procedure of Glasgow *et al.* optimized protein sequences for stability of the ligand-bound dimer complex but did not simultaneously optimize for destabilization of the ligand-free dimer. Application of such a negative design approach (*8*) could potentially yield better CIDs that display enhanced dynamic range. Nonetheless, the ability to design artificial CIDs for user-defined molecules should open the door to the development of cell-based therapies in which sensor-actuator devices are used to detect disease markers and respond by activating the production and secretion of drugs. ■

REFERENCES AND NOTES

1. A. A. Glasgow *et al.*, *Science* **366**, 1024 (2019).
2. Z. Guo *et al.*, *J. Am. Chem. Soc.* **141**, 8128 (2019).
3. N. D. Taylor *et al.*, *Nat. Methods* **13**, 177 (2016).
4. A. Fegan *et al.*, *Chem. Rev.* **110**, 3315 (2010).
5. J. N. Pelletier, F.-X. Campbell-Valois, S. W. Michnick, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12141 (1998).
6. M. J. Bick *et al.*, *eLife* **6**, e28909 (2017).
7. J. Feng *et al.*, *eLife* **4**, e10606 (2015).
8. J. A. Davey, R. A. Chica, *Protein Sci.* **21**, 1241 (2012).

10.1126/science.aaz8085



PHOTO: INGA SPENCE/SCIENCE SOURCE

Root nodules harbor nitrogen-fixing bacteria. Engineering plants to form root nodules could reduce fertlizer use.

PLANT BIOLOGY

Turning lateral roots into nodules

The evolutionary origin of legume root nodules that help them grow is revealed

By **Anthony Bishopp** *and* **Malcolm J. Bennett**

Nitrogen is critical for plant growth but must be acquired from the soil in reduced forms such as nitrate. Often this occurs through biological fixation whereby nitrogen-fixing bacteria, such as rhizobia, live symbiotically in root nodules of legumes. Theories concerning the evolutionary origins of legume root nodules range from them being highly modified stems or lateral roots (*1*, *2*). The latter origin is supported by rhizobia inducing lateral root-like nodules in nonleguminous *Parasponia* by triggering pericycle cell divisions, which normally produce lateral roots (*3*). However, in legume roots, rhizobia also induce cortical cell divisions, triggering many morphological differences to lateral root organogenesis (*2*). Two studies—by Soyano *et al.* (*4*) on page 1021 of this issue and by Schiessl *et al.* (*5*)—report key regulatory components of lateral root organogenesis to have been hijacked in legumes to coordinate nodule formation. This could open avenues to engineer nitrogen fixation in nonlegume crops.

Root nodule organogenesis in legumes is initiated when rhizobia first colonize young root hairs (*2*). Nodulation factors released by rhizobia trigger root hair curling, cortical cell divisions that produce the nodule primordia, and infection thread formation (see the figure). The infection thread is a root-derived conduit that enables rhizobia to colonize dividing cortical cells, where they differentiate into nitrogen-fixing bacteroids and provide ammonium to the plant host. Root nodule organogenesis therefore represents a complex interplay between plant and bacterial symbiotic partners. Genetic studies have proved instrumental in dissecting the underlying regulatory mechanisms and signals controlling this important developmental program.

One of the first nodule regulatory genes identified in legumes was *NODULE INCEPTION* (*NIN*) (*6*), which encodes a transcription factor that is required for rhizobia infection and nodule organogenesis (*7*). NIN promotes the expression of two *NUCLEAR FACTOR-Y* (*NF-Y*) genes which, when co-expressed with NIN, can activate cortical cell division during nodule development and also alter development of lateral roots. This provided a tantalizing link between lateral root and nodule organogenesis, but it remained unclear whether these root processes shared a common regulatory program. An observation that suggested that lateral roots and nodules develop through distinct regulatory programs was the contrasting roles of the phytohormones auxin and cytokinin. It was well documented that auxin accumulated at the sites of both lateral root and nodule primordia (*8*). By contrast, cytokinin promotes nodule formation in a NIN-dependent manner (*7*, *9*) but antagonizes lateral root formation (*10*).

Soyano *et al.* and Schiessl *et al.* dissected the nodule gene regulatory networks operating in two model legume plants, *Lotus japonicus* and *Medicago truncatula*, using distinct approaches. Soyano *et al.* adopted a chromatin immunoprecipitation–based approach to pinpoint NIN target genes that are expressed following rhizobial infection of lotus roots. Each NIN target was overex-

Plant and Crop Sciences, School of Biosciences, University of Nottingham, Nottingham LE12 5RD, UK. Email: anthony.bishopp@nottingham.ac.uk; malcolm.bennett@nottingham.ac.uk

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

pressed with NF-Y and only one gene, *LATERAL ORGAN BOUNDARIES DOMAIN 16* (*LBD16*), was able to promote cortical cell divisions. Schiessl *et al.* probed gene regulatory networks operating during lateral root initiation and nodule formation. This revealed that lateral root and nodule initiation programs overlap considerably. Common genes included many associated with auxin and cytokinin action, genes known to be involved in nodule formation, such as the *ARF* family of auxin response factors and the *YUCCA* family of auxin biosynthesis genes, as well as genes previously associated with lateral root development. Soyano *et al.* and Schiessl *et al.* analyzed mutant alleles in these common genes and found that only *LBD16* is required for both lateral root and nodule formation.

LBD16 encodes a transcription factor, first identified in the commonly used model plant *Arabidopsis thaliana*, in which it is expressed in the pericycle and promotes asymmetric cell divisions in which daughter cells gain distinct identities that are required for lateral root morphogenesis (*11*). Soyano *et al.* revealed that LBD16 expression in legume roots induced the expression of reporter genes in dividing cortical and pericycle cells. This observation raises a critical question: Does *LBD16* expression in dividing cortical cells provide the key step in creating populations of dividing pericycle and cortical root cells that form root nodules? Soyano *et al.* showed that the presence of a NIN-binding site in an *LBD16* intron led to its expression in nodule primordia. Moreover, these authors observed that the NIN-binding site was conserved in many species of legumes, but absent in nonlegume species. This raises the fascinating possibility that evolution of this NIN-binding site in an ancestral legume genome caused *LBD16* to be ectopically expressed in cortical cells. This event would have caused a gene regulatory network, normally restricted to pericycle cells and controlling lateral root development, to also be expressed in the cortex and induced by cytokinin (through NIN), leading to nodule formation (see the figure).

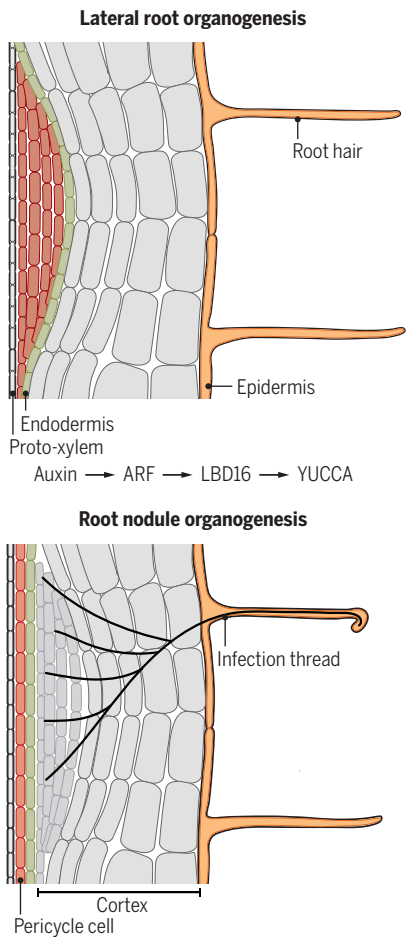
This "co-option" model helps answer a long-standing question relating to the morphological differences between lateral root-like nodules in nonlegumes versus root nodules in legumes. Lateral root-like nodules originate from dividing pericycle cells, whereas root nodules arise from dividing pericycle and cortical cells. This morphological difference may simply reflect the evolution of NIN-binding sites in the *LBD16* intron sequence of legumes that are missing in nonlegumes. It would be intriguing to test whether introducing NIN-binding sites in the *Parasponia LBD16* sequence is able to convert their lateral root-like nodules to true root nodules.

Stepping beyond nodule formation and even plant science, the question of how new structures emerge is an enduring conundrum that has intrigued evolutionary biologists. It is known from several model systems that the expression of key transcription factors has been fine-tuned through evolution to favor the development of one organ over another. For example, in blind cavefish, sonic hedgehog genes *ShhA* and *ShhB* are amplified locally to negatively control eye development, but their overexpression increases jaw and taste bud development (*12*), providing an elegant example of how gene regulatory networks can be modified to promote the development of one organ rather than another. The discovery of enhanced LBD16 activity in cortical cells takes this phenomenon to a different level: Whether there is another example by which misexpression of a transcription factor in another tissue has hijacked an existing gene regulatory network to create an entirely new organ is unknown.

The apparent co-option of the master regulator LBD16 raises the intriguing possibility that the number of genetic changes required for nodule formation in legumes may be relatively small. If so, it might be feasible to engineer legume-style nodules into other crops. Nodule organogenesis is also widely spread across four orders of flowering plants, and comparative genome sequencing studies recently revealed that inactivation of core nodulation genes, including *NIN* and *NF-Y*, have likely led to the loss of nodule-forming ability in different species (*13*, *14*). In principle, this is good news because crops in these orders may require limited gene editing to reengineer their nodule-forming abilities. If either reengineering strategy could be achieved, it promises a revolution in agriculture, with symbiotic bacteria potentially lessening the hundreds of millions of tonnes of nitrogen fertilizer applied to crops annually (*15*), generating major energy savings (more than 3% of the global carbon budget is required for nonbiological nitrogen fixation) and aiding food security in countries with limited access to fertilizers. ■

Lateral roots and nodules share common regulators

Lateral roots are derived primarily from dividing pericycle cells. Root nodule organogenesis initiates after rhizobia trigger root hair curling, cortical cell divisions, and infection thread formation. Despite being regulated by distinct hormone signals, lateral root and root nodule regulatory networks have common components.



ARF, auxin response factor; LBD16, lateral organ boundaries domain 16; NF-Y, nuclear factor-Y; NIN, nodule inception.

GRAPHIC: N. DESAI/*SCIENCE*

REFERENCES AND NOTES

1. J. I. Sprent, *New Phytol.* **111**, 129 (1989).
2. A. M. Hirsch, T. A. LaRue, J. Doyle, *Crit. Rev. Plant Sci.* **16**, 361 (1997).
3. J. H. Becking, *Development and Function of Roots* (Academic Press, London) (1975).
4. T. Soyano *et al.*, *Science* **366**, 1021 (2019).
5. K. Schiessl *et al.*, *Curr. Biol.* S0960-9822(19)31165-0 (2019). 10.1016/j.cub.2019.09.005
6. L. Schauser, A. Roussis, J. Stiller, J. Stougaard, *Nature* **402**, 191 (1999).
7. J. F. Marsh *et al.*, *Plant Physiol.* **144**, 324 (2007).
8. T. T. Xiao *et al.*, *Development* **141**, 3517 (2014).
9. J. D. Murray *et al.*, *Science* **315**, 101 (2007).
10. L. Laplaze *et al.*, *Plant Cell* **19**, 3889 (2007).
11. T. Goh, S. Joi, T. Mimura, H. Fukaki, *Development* **139**, 883 (2012).
12. Y. Yamamoto, M. S. Byerly, W. R. Jackman, W. R. Jeffery, *Dev. Biol.* **330**, 200 (2009).
13. M. Griesmann *et al.*, *Science* **361**, 144 (2018).
14. R. van Velzen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4700 (2018).
15. Food and Agriculture Organization of the United Nations, *World Fertilizer Trends and Outlook to 2018* (2015); www.fao.org/3/a-i4324e.pdf.

ACKNOWLEDGMENTS

A.B. and M.J.B. are supported by Royal Society, Leverhulme Trust and the Biotechnology and Biological Sciences Research Council.

10.1126/science.aay8620

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

GLOBAL HEALTH

Artificial intelligence for global health

Socially responsible technologies promise to help address health care inequalities

By **Ahmed Hosny**[1,2] *and* **Hugo J. W. L. Aerts**[1,2,3]

Artificial intelligence (AI) has demonstrated great progress in the detection, diagnosis, and treatment of diseases. Deep learning, a subset of machine learning based on artificial neural networks, has enabled applications with performance levels approaching those of trained professionals in tasks including the interpretation of medical images and discovery of drug compounds (*1*). Not surprisingly, most AI developments in health care cater to the needs of high-income countries (HICs), where the majority of research is conducted. Conversely, little is discussed about what AI can bring to medical practice in low- and middle-income countries (LMICs), where workforce shortages and limited resources constrain the access to and quality of care. AI could play an important role in addressing global health care inequities at the individual patient, health system, and population levels. However, challenges in developing and implementing AI applications must be addressed ahead of widespread adoption and measurable impact.

Health conditions in LMICs and HICs are rapidly converging, as indicated by the recent shift of the global disease burden from infectious diseases to chronic noncommunicable diseases (NCDs, including cancer, cardiovascular disease, and diabetes) (*2*). Both contexts also face similar challenges, such as physician burnout due to work-related stress (*3*), inefficiencies in clinical workflows, inaccuracies in diagnostic tests, and increases in hospital-acquired infections. Despite these similarities, more basic needs remain unmet in LMICs, including health care workforce shortages, particularly specialist medical professionals such as surgical oncologists and cardiac care nurses. Patients often face limited access to drugs, diagnostic imaging hardware (ultrasound, x-ray), and surgical infrastructures (operating theaters, devices, anesthesia). When equipment is available, LMICs often lack the technical expertise needed to operate, maintain, and repair it. As a result, 40% of medical equipment in LMICs is out of service (*4*). Conditions are exacerbated in fields that require both specialized workforce and equipment. For example, delivering radiotherapy requires a team of radiation oncologists, medical physicists, dosimetrists, and radiation therapists—together with sophisticated particle accelerator equipment. Consequently, 50 to 90% of cancer patients requiring radiotherapy in LMICs lack access to this relatively affordable and effective treatment modality (*5*).

LMICs have undertaken substantial health care spending, saving millions of lives by improving access to clean water, vaccinations, and HIV treatments. However, changes in health care needs owing to increased mortality from complex NCDs require high-quality, longitudinal, and integrated care (*6*). These emerging challenges have been central to the United Nations' Sustainable Development Goals, including the aim to reduce by one-third premature mortality from NCDs by 2030. AI has the potential to fuel and sustain efforts toward these ambitious goals.

Health care–related AI interventions in LMICs can be broadly divided into three application areas (see the figure). The first includes AI-powered low-cost tools running on smartphones or portable instruments. These mainly address common diseases and are operated by nonspecialist community health workers (CHWs) in off-site locations, including local centers and households. CHWs may use AI recommendations to triage patients and identify those requiring close follow-up. Applications include diagnosing skin cancer from photographic images and analyzing peripheral blood samples to diagnose malaria (*7*); more are expected given the emergence of pocket diagnostic hardware, including ultrasound probes and microscopes. With increasing smartphone penetration, patient-facing AI applications may guide lifestyle and nutrition, allow symptom self-assessment, and provide advice during pregnancy or recovery periods—ultimately allowing patients to take control of their health and reducing the burden on limited health systems.

The second application area focuses on more specialized medical needs, with the goal of supporting clinical decision-making. AI may allow nonspecialized primary care physicians to perform specialized tasks including reading diagnostic radiology and pathology images, only referring to specialists if necessary. AI tools may also help provide specialists with expert knowledge across multiple subspecialties. This is particularly important in oncology, where lack of subspecialists may force an oncologist to manage tumors across multiple anatomical sites, and thus deliver care of inferior quality owing to the constantly varying scope of services. In radiotherapy, for example, semi-automation of the treatment planning process may speed up treatment delivery, increase patient intake, and allow greater focus on the clinical nuances of patient management—all without requiring additional personnel. Although AI may not directly address diagnostic and therapeutic equipment shortages, AI integration into equipment design may help nontechnical operators troubleshoot issues when technicians are scarce. By analyzing historical maintenance data, AI may also help sustain long-term operations, predict failures, and avoid delay on parts and consumables.

The third application area relates to population health and allows public agencies to realize cause-and-effect relationships, appropriately allocate the often limited resources, and ultimately mitigate the progression of epidemics (*8*). Improving data collection in LMICs is central to these applications. For example, AI may help maintain up-to-date national cancer registries. Automated registry curation, by extracting standard data from free-form text found in radiology and pathology reports, may help reduce labor costs that account for more than 50% of all registry activity expenses (*9*). Other applications include identifying hotspots for potential disease outbreaks in unmapped rural areas by utilizing AI-powered analysis of aerial photography and weather patterns, as well as planning and optimizing CHWs' household visiting schedules. Although these applications may prompt immediate actionable interventions, their translation into effective long-term health policies remains unclear.

HIC-based AI applications in health care are far from perfect. Most are at the proof-of-concept stage and require further demonstration of utility through clinical validation in prospective trials. The underlying methods are often uninterpretable, making it difficult to predict failures and critically assess results. Data used to train AI models are almost entirely collected within HICs, and models are hence skewed toward certain diseases, demographics, and geogra-

[1]Artificial Intelligence in Medicine (AIM) Program, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [2]Radiaton Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [3]Radiology, Maastricht University, Maastricht, Netherlands. Email: haerts@bwh.harvard.edu

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

phies. With varying degrees of statistical data analysis and quality control, errors and systematic biases are introduced into models, thereby limiting their generalizability, especially when deployed in different contexts. Ethical concerns about the use of AI in health care include undermining patient data privacy protections and exacerbating the existing tension between providing care and generating profit, as well as introducing a third party into the patient-doctor relationship, which changes expectations of confidentiality and responsibility (*10*). From a regulatory perspective, medical malpractice and liabilities in health-related algorithmic decision-making are yet to be formulated. Nearly all AI tools in health care are single-task applications, and so they are incapable of fully substituting for health professionals. Understanding these limitations may help avoid hype and inflated expectations.

Introducing AI tools in resource-constrained settings presents additional challenges. The distinct needs, diseases, demographics, and standards of care in LMICs must be acknowledged through identifying specific use cases where AI involvement would have the greatest impact. Data for AI training and validation must be context specific: Computer vision systems may be required to work with legacy data formats (e.g., film versus digital x-ray), whereas developing chatbots will require compiling corpora in local languages. Solutions must also be context specific. For example, an automated system should not recommend treatments that are unavailable locally or are prohibitively expensive. Moreover, human factors should be considered: What levels of skill, education, and computer literacy are required of end users? The amount of behavioral change needed to raise awareness and confidence in AI systems should also be addressed, enabling users to recognize limitations and accurately interpret results. Infrastructure constraints should be assessed, including the availability of devices for serving AI applications, reliability of internet connectivity and bandwidth, electricity, and the amount and quality of existing digital data, as well as future digitization efforts.

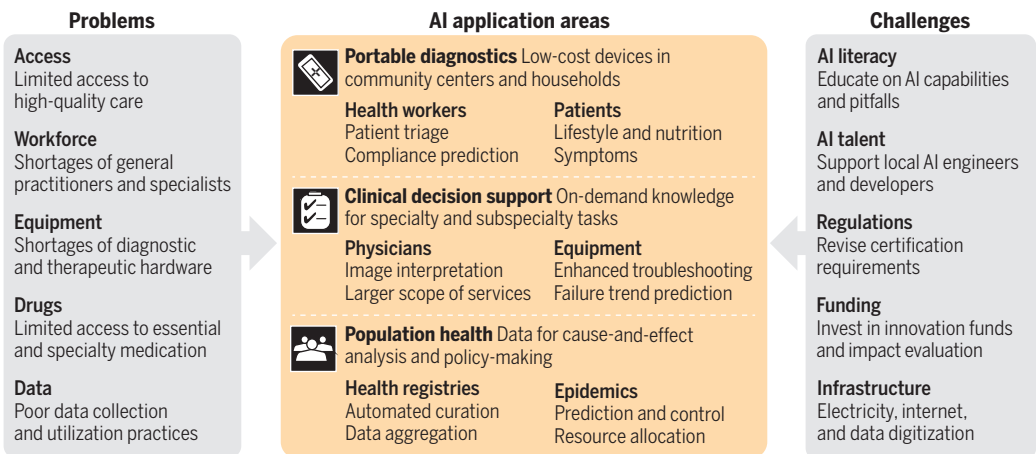
Multiple digital initiatives have been proposed to enhance access to and quality of health care in LMICs. These include technologies to support health care practices using electronic processes (eHealth) and remote telecommunications (Telehealth), an example of which is mobile health (mHealth) using mobile phones and tablets. Best practices for scaling these initiatives in LMICs have been established on the basis of real-world experiences, including the World Health Organization's mHealth Assessment and Planning for Scale (MAPS) Toolkit (*11*). These efforts could provide learning opportunities for similar digital AI applications. Many of the challenges faced by integrating electronic medical records in LMICs, for example, are likely to also impede AI applications, including limited funding, poor infrastructure for reliably delivering technologies, and discontinuous participation from users (*12*). Integration opportunities could also be considered: An existing mHealth application for patient-physician remote communication can be enhanced with an AI chatbot to triage patients prior to the consultation.

There is skepticism about the value of introducing AI in LMICs given the need to prioritize investments in basic infrastructure (*13*). AI-driven interventions should not be evaluated in isolation, nor should they be regarded as a universal panacea: Although sizable initial investments may be required, the marginal cost of providing an existing AI software service to one more user is minuscule, giving it economical scalability. An AI application may also use the deployment channels of existing digital technologies, making it almost readily deployable.

Ultimately, AI interventions in LMICs should be initiated, owned, and administered by local stakeholders—with HICs providing funding, expertise, and advice when needed. AI literacy may be included in existing global health educational programs to raise awareness about its capabilities and pitfalls. Empowering local technical AI talent will also be crucial, and may be accelerated through high-quality free educational online resources. AI implementation will require rethinking existing regulatory frameworks. For example, the training and scope of practice of CHWs may be expanded to include screening and diagnosing NCDs (*14*). Investment areas critical to bringing AI into LMICs must also be identified, as well as gathering evidence on the impact of AI solutions (*15*). Uneven distribution of access to technologies has created a digital divide between the rich and poor, while contributing to existing global inequalities. AI could emerge as a socially responsible technology with inherent equity. ■

Artificial intelligence in resource-poor health care settings

The rise in incidences of cancer and other noncommunicable diseases is straining the limited resources and infrastructure in low- and middle-income countries. Artificial intelligence (AI) applications aimed at individual patient, health system, and population levels promise to enhance the access to and quality of care.



GRAPHIC: H. BISHOP/*SCIENCE*

REFERENCES AND NOTES

1. E. J. Topol, *Nat. Med.* **25**, 44 (2019).
2. S. Mayor, *BMJ* **355**, i5456 (2016).
3. L. S. Rotenstein *et al.*, *JAMA* **320**, 1131 (2018).
4. L. Perry, R. Malkin, *Med. Biol. Eng. Comput.* **49**, 719 (2011).
5. E. H. Zubizarreta *et al.*, *Clin. Oncol. (R. Coll. Radiol.)* **27**, 107 (2015).
6. M. E. Kruk *et al.*, *Lancet Glob. Health* **6**, e1196 (2018).
7. A. D. Oliveira *et al.*, *JMIR Res. Protoc.* **6**, e70 (2017).
8. B. Wahl *et al.*, *BMJ Glob. Health* **3**, e000798 (2018).
9. F. K. L. Tangka *et al.*, *Cancer Epidemiol.* **45** (Suppl 1), S50 (2016).
10. D. S. Char *et al.*, *N. Engl. J. Med.* **378**, 981 (2018).
11. A. B. Labrique *et al.*, *Global. Health* **14**, 103 (2018).
12. B. Jawhari *et al.*, *BMC Med. Inform. Decis. Mak.* **16**, 116 (2016).
13. B. Gyawali, *Lancet Oncol.* **19**, 599 (2018).
14. S. R. Mishra *et al.*, *Global. Health* **11**, 43 (2015).
15. T. Lancet, *Lancet* **393**, 1478 (2019).

ACKNOWLEDGMENTS

Thanks to R. H. Mak and W. Ngwa for their valuable input. Supported by the U.S. National Institutes of Health (U24CA194354 and U01CA190234). A.H. is a shareholder of and receives consulting fees from Altis Labs. H.J.W.L.A. is a shareholder of Genospace and Sphera.

10.1126/science.aay5189

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS



Pedestrians avoid a crack that formed in Trona, California after the Ridgecrest earthquake.

PHOTO: MARCIO JOSE SANCHEZ/AP PHOTO

POLICY FORUM

NATURAL HAZARDS

How low should we go when warning for earthquakes?

Social responses to alerts are critical but understudied

By **Elizabeth S. Cochran**[1] *and* **Allen L. Husker**[2]

A key goal of earthquake early warning (EEW) systems is to alert populations who may be affected by a particular level of ground shaking so that they can take action to reduce impacts of that shaking, such as injuries, damages to physical infrastructure, or emotional distress. Most EEW systems work by rapidly determining the location and size of an earthquake, estimating shaking levels, and then distributing an alert to potentially affected populations. But EEW systems are limited by how rapidly the size of an earthquake can be determined as well as the details of the earthquake rupture process, the path of the seismic waves, and the alert distribution mechanism. And we are just beginning to understand how people respond to earthquake alerts, often relying on anecdotes. Determining the appropriate shaking intensities for public warnings requires understanding the range of individual and societal responses to earthquake alerts. The decision on when to issue earthquake alerts must balance the technical capabilities and potential outcomes, both desired and undesired, when choosing a ground-motion alerting threshold. Only when benefits outweigh the risks and users are prepared for alerts should they be used to warn the public about the possibility of earthquake shaking.

[1]U.S. Geological Survey, Earthquake Science Center, Pasadena, CA, USA. [2]Instituto de Geofísica, Universidad Nacional Autónoma de México, Mexico City, Mexico. Email: ecochran@usgs.gov

The city of Los Angeles released a public smartphone app in late 2018 to deliver alerts generated by the ShakeAlert EEW system operated by the U.S. Geological Survey (*1*). Alert levels of the app had been set to warn residents if potentially damaging shaking was expected, although the conditions for issuing alerts were not clearly conveyed to residents. After the 5 July 2019 moment magnitude (M_w) 7.1 Ridgecrest earthquake, many residents of Los Angeles were unhappy that an alert was not sent for shaking that was felt across the city. The original alert threshold for ShakeAlert was 4 on the modified Mercalli intensity (MMI) shaking severity scale, which is generally felt but not dangerous. The threshold was chosen to be somewhat lower than the shaking intensity where it was possible that some buildings might incur slight damage (MMI 6). The M_w 7.1 Ridgecrest earthquake produced only light shaking (MMI 3 and 4) across Los Angeles. Pressure from residents after that earthquake resulted in the city lowering the app's alert threshold to MMI 3 (*2*). Shaking at this level may not be recognized as an earthquake because it is similar to vibrations from a passing truck.

EEW is characterized by the Goldilocks principle: If too far from the earthquake rupture, the shaking is not dangerous; if too close, there will be little to no advanced warning for strong shaking. There is a donut-shaped zone around the earthquake rupture where EEW can outrun the seismic waves yet shaking may still be strong enough to cause impacts (see the figure). The size of the donut generally increases with magnitude, but for moderate earthquakes (magnitude less than M_w 5), the zone that is alerted in advance of damaging shaking may become very small (*3*, *4*).

How quickly an EEW system can issue an alert to a location depends on the distance to and size of an evolving earthquake rupture (*5*). But earthquakes are not instantaneous: The smallest earthquakes will grow to their final size in a fraction of a second, but it can take more than a minute for the largest ruptures. Despite the initial hope that the final size of an earthquake could be predicted from how an earthquake starts (*6*), there is growing evidence that small and large earthquakes start identically (*7*, *8*).

Accurately measuring the earthquake size and location is only part of the challenge—the EEW system must also characterize the distribution of the expected shaking before an alert can be issued. The ability to predict shaking levels is fraught with uncertainty, and the actual shaking that occurs can vary substantially from predictions, depending on whether the earthquake is more or less energetic for its size, the direction a fault breaks, the path the seismic waves take through sometimes-complex subsurface structures, and the particular geology under a location. This means that even when the system correctly estimates the earthquake location and final magnitude, it is still not possible to precisely predict the shaking intensities that are used to issue alerts (*3*, *4*). Furthermore, many users likely conflate magnitude, M_w, with ground shaking, MMI.

One might consider that a best-case scenario for EEW could be in Mexico City, which is located in a country with a high seismicity rate but far from the plate boundary where the largest earthquakes occur and built atop a drained lake bed filled with soft sediments where shaking intensities can be amplified by a factor of 500 (*9*). In this case, high-intensity shaking can result from remote earthquakes but provide potentially long warning times (>1 min) (*10*). Typically, populations far from an earthquake can receive long warnings for shaking, but at those distances, the shaking intensities are often low, making an alert unnecessary. The large distance of Los Angeles from the Ridgecrest earthquake, for example, allowed some ShakeAlert technical users to receive up to a ~30-s warning, but only for light shaking.

Yet, despite Mexico City being distinctly suited to benefit from EEW, none of the more than 40 public alerts issued to Mexico City in the almost 30-year history of the alert system have been shown to have clearly saved a life. The only earthquake since the advent of the alert system in 1993 to have caused substan-

Downloaded from http://science.sciencemag.org/ on November 28, 2019

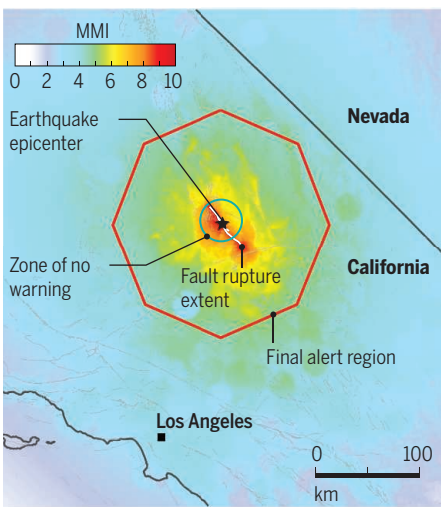
Published by AAAS

tial shaking in Mexico City, resulting in damage and deaths, was the 19 September 2017 M_w 7.1 Puebla earthquake that occurred inland, close to the city (~120 km). The EEW system provided no advance warning for that quake, with alerts sounding after the first, strong seismic waves reached Mexico City. This was largely due to it being a very deep earthquake (~60 km), and the waves were more than halfway to Mexico City when they reached the surface. The since-updated algorithm would have provided a short (~5 s) warning before seismic waves reached the city (*11*). An earthquake alert the week after the Puebla earthquake led to the injury of at least one person and possibly to the deaths of two others (*12*). The alert was issued for a moderate-sized earthquake, not associated with the Puebla earthquake, located ~600 km from the city that was not felt by most residents of Mexico City. The deaths were reported in the news as the result of heart attacks, which are hard to definitively link to the alert but which occurred in the individuals following evacuation from their respective buildings after the alert. The injury resulted from a person jumping from a window because they feared their building might collapse when shaking arrived. This is noteworthy because social scientists have typically found that a panic-like response is rare during emergencies (*13*). Though overreactions may be uncommon, even a very small percentage of a population responding imperfectly to an alert can have grave consequences.

In Los Angeles, public alerts will now be sent to areas even when shaking is not expected to be widely felt. It is worth considering the case of an alert being issued to thousands (or millions) of residents who will ultimately experience nondamaging shaking. On one hand, the majority of people who receive the alert may take a recommended protective action (such as drop, cover, and hold on) and benefit from practicing their response to earthquakes even though they are actually not at risk during this event. On the other hand, some fraction of those receiving an alert will be driving on one of the many freeways of Los Angeles. If just one driver of the thousands that receive an alert becomes distracted, an accident could occur.

We highlight this example in an attempt to bring balance to the discussion of EEW that has generally touted the potential to save lives. We are not arguing against the potential benefits of issuing earthquake alerts. The earthquakes that these EEW systems have been designed for will eventually occur, and lives will probably be saved. And EEW is particularly well suited for automated systems that incur only minor costs when an action is taken in response to an alert and where the savings from reduced damage or downtime can be substantial (*4*). For example, automated applications already slow trains, such as the Shinkansen in Japan and the Bay Area Rapid Transit system in California. The public is hungry for rapid information about earthquakes. In Japan, EEW is viewed favorably as a way to mentally prepare for shaking even though the majority of people take no protective action in response to the alerts (*14*). Similarly, residents of Mexico City are accepting of the technical limitations of EEW and view the system as necessary and valuable (*15*).

Some level of risk associated with issuing an alert is acceptable as long as the potential benefits of an alert are likely to outweigh the negative consequences. Studies of expected EEW system performance show that for many users, taking action at somewhat lower shaking intensities than would cause damage can increase warning times and provide warnings for ground motions that are unexpectedly strong (*4*, *8*). The expected technical performance of EEW systems must be paired with robust social science research into the needs and likely responses of users. Research into user expectations has begun following the Ridgecrest earthquake, but it will take additional studies by the disaster communications and emergency response communities to inform the optimal alerting strategy. Decisions on appropriate alerting thresholds may require understanding broader societal questions, such as the level of trust in both the warning system itself and in organizations responsible for issuing warnings. As with weather or other natural-hazard warning systems, alerting thresholds will be refined over the years as the public gains experience with alerts and becomes familiar with appropriate response actions.



Shaking intensity near the moment magnitude (M_w) 7.1 Ridgecrest earthquake. The no-warning zone is estimated from the time of the first alert after the earthquake began (7.0 s), using a nominal *S*-wave speed of 3.0 km/s, The final alert region is where ShakeAlert-estimated shaking would exceed 4 on the modified Mercalli intensity (MMI) scale. The alert region expanded as the estimated magnitude grew from M_w 5.5 to 6.3 from the time of the first alert (7.0 s after the earthquake began) until 14.3 s later (21.3 s after the earthquake began). Light gray lines show other faults in California.

Earthquake alerts have the potential to be a valuable tool, as long as the societal readiness, impacts, and technical capabilities of the EEW system are considered in the decision to send an alert or for users to opt in to receiving one. We have focused here on earthquake alerts that are used to initiate protective actions. Earthquake notifications that provide rapid information but do not prompt a user to take protective action require less scrutiny. To maximize the benefits of alerts, they need to be paired with robust messaging before, during, and after an earthquake. We expect the best outcomes from EEW if the public is broadly aware of when they should or should not expect to receive an alert and what actions they should take when an alert is received. ■

REFERENCES AND NOTES

1. D. D. Given *et al.*, "Revised technical implementation plan for the ShakeAlert system—An earthquake early warning system for the West Coast of the United States" (U.S. Geological Survey Open-File Report 2018–1155, U.S. Geological Survey, 2018).
2. R.-G. Lin II, "2 quakes in 2 days, no warning from ShakeAlertLA. Now the app is getting reworked," *Los Angeles Times*, 14 August 2019; www.latimes.com/california/story/2019-08-14/earthquake-early-warning-app-shakealertla-released.
3. M. Meier, *J. Geophys. Res. Solid Earth* **122**, 5561 (2017).
4. S. E. Minson *et al.*, *Sci. Rep.* **9**, 2478 (2019).
5. S. E. Minson, M.-A. Meier, A. S. Baltay, T. C. Hanks, E. S. Cochran, *Sci. Adv.* **4**, eaaq0504 (2018).
6. E. L. Olson, R. M. Allen, *Nature* **438**, 212 (2005).
7. M.-A. Meier, T. Heaton, J. Clinton, *Geophys. Res. Lett.* **43**, 7991 (2016).
8. D. T. Trugman, M. T. Page, S. E. Minson, E. S. Cochran, *J. Geophys. Res. Solid Earth* **124**, 4642 (2019).
9. H. Flores-Estrella, S. Yussim, C. Lomnitz, *Nat. Hazards* **40**, 357 (2007).
10. A. Cuéllar, G. Suárez, J. M. Espinosa-Aranda, *Bull. Seismol. Soc. Am.* **108**, 238 (2018).
11. A. Cuéllar, G. Suárez, J. M. Espinosa-Aranda, *Bull. Seismol. Soc. Am.* **108**, 2068 (2018).
12. J. O. Fierro, "Sismo en CDMX: 2 muertas por infarto y un lesionado por lanzarse de un segundo piso," *Aristegui Noticias*, 23 September 2017; https://aristeguinoticias.com/2309/mexico/sismo-en-cdmx-2-muertas-por-infarto-y-un-lesionado-por-lanzarse-de-un-segundo-piso/.
13. E. Auf der Heide, in *The First 72 Hours: A Community Approach to Disaster Preparedness* (iUniverse Publishing, Lincoln, 2004), pp. 337–380.
14. K. Nakayachi, J. S. Becker, S. H. Potter, M. Dixon, *Risk Anal.* **39**, 1723 (2019).
15. R. M. Allen, E. S. Cochran, T. J. Huggins, S. Miles, D. Oteguí, *Eos* **99**, EO105095 (2018).

ACKNOWLEDGMENTS

E.S.C. served as Chief Scientist of the U.S. ShakeAlert system. A.L.H. has not been involved in any of the EEW efforts described in the article but has consulted for Grillo, an EEW start-up in Mexico. The authors thank J. McGuire, A. Wein, K. Knudsen, S. Hickman, and three anonymous reviewers for suggestions that improved the article.

10.1126/science.aaz6601

GRAPHIC: N. DESAI/*SCIENCE*, (DATA) U.S. GEOLOGICAL SURVEY SHAKEMAP

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS



ARCHAEOLOGY

Excavating Uruk

A richly illustrated volume reveals what life was like in the world's first city

By **Andrew Robinson**

The world's oldest known mathematical problem appears on a cuneiform tablet dated to the late 4th millennium BCE. It was discovered in Uruk, the world's first city, located on the Euphrates River between Ur and Babylon. The calculation concerns the surface area of a roughly rectangular field and uses a sexagesimal system comparable to our present-day system for measuring time (60 min in an hour) and angle (360° in a circle).

Excavation at Uruk began in 1912–1913, conducted by a scholarly society based in Germany known as the Deutsche Orient-Gesellschaft. It continued under German archaeologists until its disruption in the 1980s by Iraq's wars and political troubles.

In 2013, to mark the centenary of this research, the Vorderasiatisches Museum staged an exhibition, *Uruk: 5,000 Years of the Megacity*—the first devoted exclusively to Uruk. Its accompanying book (written in German) contained more than 60 essays on subjects ranging from satellite maps and geoarchaeology to laborers' ration bowls and state administration to cuneiform and religious thought.

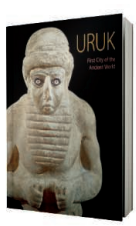
This new book is an English translation of that 2013 volume, edited by Timothy Potts, an archaeologist of the ancient Near East who is now director of the J. Paul Getty Museum. It is lavishly illustrated with objects from collections in Berlin, Dresden, and Heidelberg, as well as Brussels, London, Oxford, and Paris, and includes fascinating reconstructions of life in Uruk generated using scientific data.

One such illustration shows a cult scene in Uruk's Eanna Sanctuary with the king and priests from the shrine celebrating a festival in honor of the goddess Inanna, or Ishtar. In the distance, one sees the impressive brick wall surrounding the city—9 km long with roughly 900 semicircular bastions—built by the legendary king Gilgamesh. Margarete van Ess, one of two directors of research on Uruk at the Deutsches Archäologisches Institut and an editor of the original book, estimates that its construction would have required a thousand laborers working 6 months a year for 13.5 years.

As Potts acknowledges, Uruk is neither the most famous nor the most spectacular of ancient Mesopotamian cities. "But Uruk has a unique and in many ways wider significance," he writes. First, it is "the 'mother city' of ancient Sumer…the earliest site where the key social, cultural, economic, and technological elements that come to characterize Mesopotamian civilization for the next three millennia appear." Second, it is "the key point of reference for understanding the origins and dynamics of early urbanization around the world." Furthermore, Uruk has long been believed to be the site where writing originated, as cuneiform script, although this long-standing claim has been challenged by recent Egyptian finds, as Egyptologist Jochem Kahl discusses in his essay.

A virtual reconstruction depicts a festival honoring the goddess Inanna/Ishtar in Uruk's Eanna Sanctuary.

The area within Uruk's city wall is 5.5 km², yet less than 5% has been dug—and only 0.1% of the ancient levels. In the center, some 35 layers of building lie on top of one another, spanning more than 4500 years of urban history and a depth of 25 m. Given this degree of complexity, "It would not be sensible, either scientifically or from the point of view of international standards for the preservation of archaeological heritage, to excavate the entire city," argue van Ess and geophysicists Helmut Becker and Jörg Fassbinder.



Uruk: First City of the Ancient World
Timothy Potts, editor
J. Paul Getty Museum, 2019. 404 pp.

Such layering has created difficulties in dating objects from Uruk, notes archaeologist Hans Nissen, especially cuneiform tablets, which have been found in layers of rubble and fill between building phases, making it impossible to determine their age before disposal. Still, from various discoveries it is clear that the oldest Uruk cuneiform—and writing's appearance in Mesopotamia—dates to around 3300 BCE.

With so many contributors (some four dozen) and so much uncertainty about Uruk, there is inevitably considerable overlap in the book's essays, especially those concerning cuneiform. They are also universally academic in tone, with only occasional attempts to appeal to nonspecialists. That said, this book is a worthy and welcome expansion of the literature on Uruk in English, with a cornucopia of images—including a cuneiform tablet listing 58 different names for pigs—sure to intrigue almost any 21st-century city dweller. ■

10.1126/science.aaz7969

The reviewer is the author of *The Story of Writing: Alphabets, Hieroglyphs and Pictograms* (Thames & Hudson, 1995) and *The Indus: Lost Civilizations* (Reaktion Books, 2015).
Email: andrew@andrew-robinson.org

IMAGE: ARTEFACTS-BERLIN.DE; WISSENSCHAFTLICHES MATERIAL: DEUTSCHES ARCHÄOLOGISCHES INSTITUT, ORIENT-ABTEILUNG

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

ECONOMICS

Trash, treasure, and everything in between

A journalist traces the evolution of the large, but largely unknown, global secondhand market

By **Kevin Dooley**

In *Secondhand*, Adam Minter—building on his acclaimed first book, *Junkyard Planet*, which examined the global material recycling market—takes readers around the world to discover the people and companies that make up the global secondhand market. Along the way, he meets consumers who donate and sell their used goods; representatives from the companies that buy, aggregate, sort, refurbish, redistribute, and resell these items; and the customers who eventually buy them.

The son of a junkyard owner, and a journalist who often covers issues related to sustainability, Minter tells stories and offers insight suffused with legitimacy, pragmatism, and optimism. His narrative style is documentary, mixing context with quotes from interviews.

The book begins by contrasting how aging populations in the United States and Japan are grappling with the need to downsize their material belongings. At the root of the burgeoning need to declutter, Minter writes, is the rapid growth in the production of consumer goods since World War II.

The secondhand market has many of the same characteristics and is subject to many of the same challenges as the material recycling market. Both operate as global trades, where asymmetries in value between one country or region and another create economic opportunities. In some countries, for example, a better-made used product is preferable to a low-quality new product.

Both recycling and secondhand markets depend on sorting and aggregation and have a formal and an informal component in which entrepreneurs have discovered niche opportunities. Both spur opportunities for jobs and companies that create economic and social value.

Both markets have also been constrained by the poor quality of many recovered goods. Despite the growing amount of clothing and textiles available for secondhand use, for example, many end up in landfills, either because the quality of these items is so poor that they are unfit for further use or because there are few recycling options available.

There are ways, however, in which the secondhand market is different. There is an emotional connection between people and many of the goods they own, notes Minter. Nostalgia can drive hoarding of goods but can also drive demand for items in particular markets. Polaroid cameras, for example, which had virtually no value a decade ago, now sell for $200 because of people's nostalgia for instant film.

The secondhand market is also more dependent on global trade than the recycling market, because of the presence of value asymmetries. Secondhand goods tend to move from richer people to poorer people, which makes it more likely that they will cross borders. If logistics costs or tariffs were to increase greatly, the secondhand market would be hit hard.

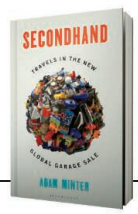


Women sort wool blankets for recycling in Panipat, India.

Consumers also play a more active role in determining the potential value of a secondhand item. Minter shares the story of Book Off, a used bookstore chain in Japan that has trained customers to take better care of their books, ensuring that those books will be purchased back after use.

The takeaway from these stories is that the secondhand market creates value for many. Customers get the value of reselling items, other customers get the benefit of buying products at a lower price point, and service organizations enable the trading, providing jobs and economic development, to say nothing of the environmental benefits associated with reusing goods instead of manufacturing new ones.

Secondhand: Travels in the New Global Garage Sale
Adam Minter
Bloomsbury, 2019. 320 pp.



Minter offers several solutions for maintaining and improving the global secondhand market. The first centers on the need to increase product longevity. For almost 100 years, many companies and industries have employed a "design for obsolescence" mind-set to promote continued sales, despite consumers showing a preference for products with longer life spans and lower cost of ownership in certain categories, including appliances. Minter argues that the value of making a product last longer is not only that it provides more value to the initial customer, but that it increases the likelihood that the product will be reused in a secondhand market.

Minter also advocates for "right to repair" laws. Many companies have made it difficult for independent repair services to service their products, forcing consumers to use those brands' repair services. This, he argues, constrains the development of a competitive market for repair services.

Minter also believes that people living in developed countries must stop making judgments about how people in the developing world should or should not make a living, as it relates to recycling or secondhand markets. He argues that certain prejudices lead to burdensome regulations that prevent used goods and materials from being shipped abroad.

In the book's afterword, Minter shares a story of how his wife, Christine—inspired by his research and reporting—has started a secondhand book service. Now that I am done with *Secondhand*, perhaps I will track her down and ask her to pass it on to another reader, who will no doubt enjoy it as much as I did. ■

10.1126/science.aaz5198

The reviewer is at the Sustainability Consortium and Department of Supply Chain Management, Arizona State University, Tempe, AZ 85287, USA. Email: kevin.dooley@asu.edu

Downloaded from http://science.sciencemag.org/ on November 28, 2019

PHOTO: SECONDHAND

Published by AAAS



LETTERS

PHOTO: STEFAN HUWILER/ROLFNP/ALAMY STOCK PHOTO

Edited by **Jennifer Sills**

EU Court: Science must justify future hunting

For strictly protected species in Europe, the 1992 Habitats Directive requires EU Member States to implement conservation actions that include a ban on their capture and killing (*1*). Several Member States have creatively evaded this requirement to allow annual hunting of some of these species, including wolf (*Canis lupus*), bear (*Ursus arctos*), and lynx (*Lynx lynx*), by exploiting provisions that allow exceptions to strict protection (*2*). The Directive allows limited exceptions to achieve particular goals when there is no satisfactory alternative and making the exception would not harm the conservation status of the species' populations. A recent decision by the Court of Justice of the European Union (CJEU) (*3*) makes it much harder for Member States to interpret these provisions to allow hunting and rightly centers future policy decisions on the results of scientific research.

The case, initiated by the nongovernmental nature-protection organization Tapiola, challenged Finland's justification of wolf hunting as a conservation measure needed to prevent poaching (*4*). The CJEU ruled that the prevention of poaching is a legitimate conservation goal that might justify exceptions from strict protection. However, it also interpreted the associated conditions in such a strict manner that in practice it will be difficult to justify hunting for this purpose (*3*).

This ruling lays out important limitations on hunting strictly protected species throughout the EU (*3*). First, Member States cannot allow hunting for conservation purposes unless rigorous scientific studies indicate that hunting would have a positive net impact on the strictly protected population. Second, exceptions from strict protection may be used only as a last resort for achieving their claimed purposes. The Member State must be able to demonstrate, with reference to scientific sources, that there is no other satisfactory alternative. Third, the CJEU emphasized that the precautionary principle prevents Member States from making exceptions to strict protection if the best available science leaves uncertainty as to whether the conservation status of populations involved would be negatively affected.

This decision makes explicit the need for good science to inform environmental protection laws. Examples of how conservation scientists and others can contribute include modeling the demographic and ecological impacts of exemptions and identifying scientifically grounded alternative solutions to hunting. A greater awareness of the legal questions that require the help of scientists to answer could result in more policy-relevant research agendas and improved environmental decision-making.

Yaffa Epstein[1]*, José Vicente López-Bao[2], Arie Trouwborst[3], Guillaume Chapron[4]
[1]Faculty of Law, Uppsala University, 753 09 Uppsala, Sweden. [2]Research Unit of Biodiversity (UO/CSIC/PA), Oviedo University, E-33600 Mieres, Spain. [3]Department of European and International Law, Tilburg University, 5000 LE Tilburg, The Netherlands. [4]Department of Ecology, Swedish University of Agricultural Sciences, 73091 Riddarhyttan, Sweden.
*Corresponding author.
Email: yaffa.epstein@jur.uu.se

An EU decision will make hunting strictly protected species such as the lynx more difficult to justify.

REFERENCES AND NOTES

1. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora (1992); https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31992L0043.
2. A. Christiernsson, *J. Eur. Environ. Plan. Law* **16**, 237 (2019).
3. Court of Justice of the European Union, Luonnonsuojeluyhdistys Tapiola Pohjois-Savo—Kainuu ry, case C-674/17 (2019); http://curia.europa.eu/juris/celex.jsf?celex=62017CJ0674.
4. J. Darpö, *J. Eur. Environ. Plan. Law* **16**, 305 (2019).

COMPETING INTERESTS

Y.E. and G.C. receive funding for a project that touches on hunting and conservation issues from Riksbankens jubileumsfond and have previously (within the past 3 years) received funding for species conservation research from the Swedish Environmental Protection Agency. J.V.L.-B. is a member of the IUCN Canid Specialist Group. A.T. is a member of the IUCN Large Carnivore Initiative for Europe. G.C. is a member of the IUCN Large Carnivore Initiative for Europe, IUCN Canid Specialist Group, and IUCN Cat Specialist Group.

10.1126/science.aaz8424

Time to ban lead hunting ammunition

Despite evidence that lead is an extremely neurotoxic and persistent element (*1*), its use in hunting ammunition continues. The European Chemicals Agency (ECHA) is conducting an investigation into ammunition-derived lead's risk to wildlife and humans, but its results will take time (*2*). Individuals and organizations must take immediate action—independent of governmental legislation—to stop the use of lead in hunting ammunition.

ECHA estimates that 35,000 tons of lead is released into Europe's environment each year, including 5000 tons dispersed into wetlands (*3*). Ammunition-derived lead has caused suffering and population declines in the region's birds (*4*, *5*). Losses due to lead ammunition cost USD1.1 billion per year in terms of lost wildlife and biodiversity, environmental health, and socio-economy as measured by hospitalizations and loss of IQ (*6*). Yet, EU legislation is rare, and only Denmark and the Netherlands have enacted total bans on lead shot (*7*).

In the United States, documentation of the adverse effects of ammunition-derived lead on wildlife dates back to the 1870s (*8*). Evidence of millions of water bird deaths annually (*9*) resulted in a phase-out of lead

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

Where Science Gets Social.

AAAS.ORG/COMMUNITY

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

Member COMMUNITY AAAS

shot for hunting in 1991 (*10*). By 2017, a federal ban on all lead ammunition use on federal land was declared; however, the Trump administration revoked the ban shortly afterward (*11*). Encouragingly, a total ban on all lead hunting ammunition took effect on 1 July 2019 in California (*12*).

We urge national hunter's organizations across Europe, the United States, and worldwide to collaborate and persuade members and industry to phase out the use and production of lead hunting ammunition. This approach would promote the interests of hunters through the survival of more quarry animals and through the stimulation of a more positive public perception of hunting. Moreover, it would protect the environment and prevent further harm to wildlife.

Christian Sonne[1,2]* Aage K. O. Alstrup[3], Yong Sik Ok[2,4], Rune Dietz[1], Niels Kanstrup[5]
[1]Aarhus University, Roskilde, Denmark. [2]School of Forestry, Henan Agricultural University, Zhengzhou, China. [3]Aarhus University, Aarhus, Denmark. [4]Korea University, Seoul, Korea. [5]Aarhus University, Kalø, Denmark.
*Corresponding author. Email: cs@bios.au.dk

REFERENCES AND NOTES

1. World Health Organization, "Preventing disease through healthy environments," (2010); www.who.int/ipcs/features/lead.pdf.
2. European Chemicals Agency, "Registry of restriction intentions until outcome" (2019); https://echa.europa.eu/registry-of-restriction-intentions/-/dislist/details/0b0236e1840159e6.
3. European Chemicals Agency, "ECHA identifies risks to terrestrial environment from lead ammunition" (2018); https://echa.europa.eu/-/echa-identifies-risks-to-terrestrial-environment-from-lead-ammunition.
4. R. J. Williams *et al.*, *Rev. Environ. Contam. Toxicol.* **245**, 157 (2018).
5. A. H. Boesen *et al. Front. Vet. Sci.* **6**, 285 (2019).
6. D. J. Pain *et al.*, *Ambio* **48**, 969 (2019).
7. R. Mateo *et al.*, *Ambio* **48**, 989 (2019).
8. M. Friend, J. C. Franson, W. L. Anderson, in *Ingestion of Lead from Spent Ammunition: Implications for Wildlife and Humans*, R. T. Watson, M. Fuller, M. Pokras, W. G. Hunt, Eds. (The Peregrine Fund, Boise, ID, 2009), pp. 34–60.
9. C. F. Bellrose, *Illinois Natural Hist. Survey Bull.* **27**, 235 (1959).
10. D. Avery, R. T. Watson, in *Ingestion of Lead from Spent Ammunition: Implications for Wildlife and Humans*, R. T. Watson, M. Fuller, M. Pokras, W. G. Hunt, Eds. (The Peregrine Fund, Boise, ID, 2009), pp. 161–168.
11. V. Volcovici, "New Interior head lifts lead ammunition ban in nod to hunters," *Reuters* (2017); www.reuters.com/article/us-usa-interior-zinke/new-interior-head-lifts-lead-ammunition-ban-in-nod-to-hunters-idUSKBN16930Z.
12. California Department of Fish and Wildlife (2019); www.wildlife.ca.gov/Hunting/Nonlead-Ammunition.

10.1126/science.aaz8150

Beyond meat: Ecological functions of livestock

Livestock production and meat consumption are major drivers of biodiversity loss and carbon emissions globally (*1*, *2*). Governments and civil society will have to prioritize the reduction of livestock numbers and meat consumption [e.g., (*2*, *3*)] to mitigate impacts and achieve international sustainability goals. However, traditional livestock systems also play a role in biodiversity conservation, climate adaptation, and socioecological resilience at regional and local scales.

In Europe, traditional breeds of free-range livestock are fulfilling conservation goals by securing the ecological role of wild large herbivores that are long absent or in low abundance (*4*). These livestock breeds may include traditional breeds of cattle, often grazing unattended, and small herds of domestic goats or sheep that contribute to the maintenance of high-nature-value habitats and diverse landscape mosaics, the regulation of vegetation growth and structure (also linked to fire prevention, especially in southern Europe), and the maintenance of genetic diversity, local identity, and knowledge (*5*, *6*).

As the 2021–2030 UN Decade on Ecosystem Restoration approaches (*7*), agri-environmental schemes and labeling and certification schemes (*8*) should adopt proactive measures that go beyond impact mitigation. Sustainable use of traditional livestock systems can help restore and manage biodiversity and ecosystem services where their maintenance contributes to local, regional, and ultimately global conservation goals.

Vânia Proença* and Carlos M. G. L. Teixeira
MARETEC, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal.
*Corresponding author.
Email: vania.proenca@tecnico.ulisboa.pt

REFERENCES AND NOTES

1. A. Marques *et al.*, *Nat. Ecol. Evol.* **3**, 628 (2019).
2. H. C. J. Godfray *et al.*, *Science* **361**, eaam5324 (2018).
3. R. Harrabin, "Plant-based diet can fight climate change—UN," *BBC* (2019); www.bbc.com/news/science-environment-49238749.
4. S. J. G. Hall, R. G. H. Bunce, *Ecol. Evol.* **9**, 5859 (2019).
5. S. J. G. Hall, *People Nat.* **1**, 284 (2019).
6. R. Lovreglio, O. Meddour-Sahar, V. Leone, *iForest* **7**, 260 (2014).
7. UNEP, "New UN Decade on Ecosystem Restoration offers unparalleled opportunity for job creation, food security and addressing climate change," UNEP (2019); www.unenvironment.org/news-and-stories/press-release/new-un-decade-ecosystem-restoration-offers-unparalleled-opportunity.
8. Global Nature Fund, Lake Constance Foundation, "Baseline Report : Biodiversity in Standards and Labels for the Food Sector," LIFE (2017); www.business-biodiversity.eu/en/baseline-report.

COMPETING INTERESTS

V.P. is the scientific coordinator of a project conducted by Terraprima as part of her research activity at MARETEC. V.P. and C.M.G.L.T. have been financially supported in the past through short-term contracts by Terraprima.

10.1126/science.aaz7084

Published by AAAS

Downloaded from http://science.sciencemag.org/ on November 28, 2019

PRIZE ESSAY

GENOMICS, PROTEOMICS, AND SYSTEMS BIOLOGY

Three-dimensional genome structure of a single cell

Chromosome organization in diploid cells reveals a structural basis for smell

By **Longzhi Tan**

Since the 1880s, scientists such as Carl Rabl have been looking at cell nuclei under a microscope and speculating about their three-dimensional (3D) structure. We now know that each nucleus in our body carries 6 billion base pairs (bp) of DNA, which would be 2 m long if fully stretched. The linear sequence of this DNA was determined by the Human Genome Project in 2003; however, its 3D structure remains elusive.

In recent decades, scientists began to uncover how the 2-m-long DNA folds into the 10-μm cell nucleus, taking advantage of next-generation DNA sequencing. This method, termed chromosome conformation capture [3C (*1*) or Hi-C (*2*)], provides a genome-wide view of DNA folding by averaging a large ensemble of cells. However, ensemble averaging cannot capture the relative positioning of different chromosomes and their interactions, because each chromosome, which occupies a contiguous territory in a cell nucleus, only contacts a few other chromosomes (*3*). Single-cell 3C has been achieved (*4*) but has limited spatial resolution.

Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. Email: tttt@stanford.edu

HIGH-PRECISION STRUCTURES WITH ALLELIC RESOLUTION

In Hi-C, a cell nucleus is chemically fixed and its genome cut into small pieces and then ligated on the basis of 3D proximity. This procedure creates numerous artificial linkages—termed chromosome "contacts"—between DNA loci that are distant on the linear genome but nearby in the 3D space.

Working in the laboratory of Xiaoliang Sunney Xie at Harvard University, I improved the resolution of single-cell 3C by inventing new whole-genome amplification methods. Through a highly uniform, transposon-based (*5*) amplification method termed multiplex end-tagging amplification (META) (*6*) and other improvements, I obtained an average of ~1 million contacts per cell—five times as many as before. (High numbers of contacts are necessary for acquiring high-resolution 3D structures.)

The 23 chromosomes inherited from one's mother and the 23 from one's father differ by less than 0.1%, making it difficult to distinguish the parent of origin for each observed contact. Previous studies (*7*) have focused on haploid mouse cell lines, which are not applicable to diploid cells. I developed an algorithm to infer the parent of origin of each contact. The algorithm aggregates the minute (0.1%) parental difference from many neighboring contacts to accurately infer their common parent of origin. With these biochemical and computational advances—together termed diploid chromosome conformation capture (Dip-C)—we can now look at almost any cell in the human body (see the figure, left).

DIPLOID HUMAN AND MOUSE GENOMES IN 3D

We achieved a resolution of 20,000 bp, or 100 nm, in the first 3D structure of a diploid human genome (*6*). That is, we simultaneously localized ~300,000 DNA loci in each cell, without using a microscope. This capability is far beyond state-of-the-art optical imaging of chromosomes, which can achieve comparable resolution but can only measure ~100 loci per cell, meaning it cannot resolve the whole genome.

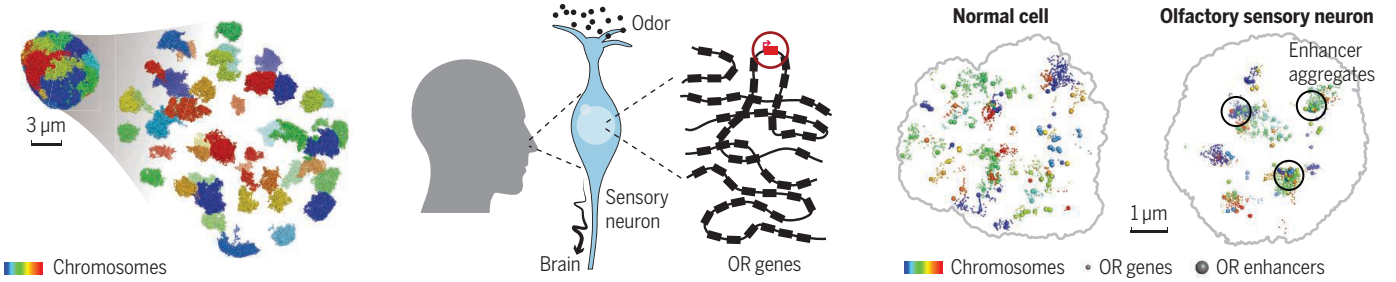
Our structures provide a holistic 3D view of the human genome with unprecedented details. Zooming in from the whole nucleus to individual chromatin loops, we have visualized the fractal organization of DNA that was originally theorized from ensemble-averaged Hi-C measurements (*2*). In imprinted genes and female X chromosomes, where two alleles differ dramatically in gene expression, we observed clear structural differences as well. This highlights an intimate relationship between structure and function.

In each tissue, different types of cells

Science & SciLifeLab PRIZE For Young Scientists

Downloaded from http://science.sciencemag.org/ on November 28, 2019

New technology enables a holistic view of cell type–specific three-dimensional genome organization



Diploid chromosome conformation capture (Dip-C)
A new method, termed diploid chromosome conformation capture" (Dip-C), resolves the 3D structure of all 46 chromosomes in a single human cell with 20-kb and 100-nm resolution.

One neuron–one receptor expression
Our sense of smell relies on each sensory neuron to express one and only one olfactory receptor (OR) gene (red circle) out of ~400 (human) or ~1100 (mouse). This one neuron–one receptor phenomenon may arise from a special 3D organization of OR genes.

Dip-C reveals interchromosomal aggregation
Aggregates (black circles) of OR genes and their enhancers in olfactory sensory neurons (right) but not in normal cells of the same tissue (left) are shown as z-projections. Note that on the left, ORs from different chromosomes may appear to be interacting because of 2D projection.

GRAPHIC: N. DESAI/*SCIENCE*, FROM L. TAN

Published by AAAS

perform different functions. We found that knowing the 3D genome structure alone allowed investigators to distinguish between the cell types in the absence of any other information. This "structure typing" not only puts Dip-C on the growing list of single-cell omics tools—such as transcriptomics (*8*) and epigenomics—for charting a complete cellular atlas of our body, it also provides a structural basis for the diverse functions of human cell types.



An electron micrograph depicts human chromosomes and nucleus.

PHOTOS: (CLOCKWISE FROM LEFT) POWER AND SYRED/SCIENCE SOURCE; LONGZHI TAN; GUSTAVO ALARCÓN-NIETO; COURTESY OF HUMSA VENKATESH; SHUTING SHAO

STRUCTURAL BASIS FOR SMELL

To demonstrate the power of Dip-C, I turned to the mouse olfactory system. In the nose, neurons sense odors by expressing olfactory receptors (ORs) (*9*). ORs constitute the largest gene family in the mouse genome, consisting of ~1100 G protein–coupled receptor (GPCR) genes from different chromosomes. Each neuron, however, only expresses a single OR, silencing all ~1100 others (*10*) after a brief developmental period of multi-OR expression (*11*, *12*). This "one neuron–one receptor" rule ensures that each neuron senses its own subset of odors and projects to the brain to form a map (*13*) of smell (see the figure, center).

Previous studies (*14*) have proposed interchromosomal interaction as a mechanism to achieve one neuron–one receptor expression. In this model, OR genes interact to ensure mutual silencing, while their enhancers jointly activate the one OR that will be expressed. However, older methods of fluorescence imaging cannot resolve all ORs at once.

Using Dip-C, I mapped all ~1100 OR genes and their ~60 enhancers in single cells throughout mouse development (*15*). What I learned is that, in a normal cell, where ORs are not expressed, most OR genes reside on the surface of the nucleus, and ORs from different chromosomes barely interact (see the figure, right). Only olfactory neurons bring together OR genes to form large, multichromosome aggregates at the nuclear center, which presumably silence all but one OR. Each cell harbors a few such aggregates, the largest of which summons nearly a dozen enhancers from different chromosomes. The formation of this enhancer hub coincides with the developmental transition in OR expression (*11*, *12*). This distinct structure is the basis for one neuron–one receptor expression.

The importance of 3D genome organization goes beyond olfaction. It may enhance night vision by inverting euchromatin and heterochromatin to form a microlens (*16*), the 3D structure of which has also been determined by Dip-C (*15*). Chromosome conformation also helps our immune system perform V(D)J recombination, coordinates the formation of neural circuits through protocadherin genes, and holds therapeutics value for cancer epigenomics and genome editing. The ability to measure 3D genomes of single cells, then, will have far-reaching impacts on fundamental biology as well as on human health. ■

REFERENCES AND NOTES

1. J. Dekker, K. Rippe, M. Dekker, N. Kleckner, *Science* **295**, 1306 (2002).
2. E. Lieberman-Aiden *et al.*, *Science* **326**, 289 (2009).
3. A. Bolzer *et al.*, *PLOS Biol.* **3**, e157 (2005).
4. T. Nagano *et al.*, *Nature* **502**, 59 (2013).
5. C. Chen *et al.*, *Science* **356**, 189 (2017).
6. L. Tan, D. Xing, C.-H. Chang, H. Li, X. S. Xie, *Science* **361**, 924 (2018).
7. T. J. Stevens *et al.*, *Nature* **544**, 59 (2017).
8. A. R. Chapman *et al.*, *PLOS ONE* **10**, e0120889 (2015).
9. L. Buck, R. Axel, *Cell* **65**, 175 (1991).
10. L. Tan, C. Zong, X. S. Xie, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 21148 (2013).
11. L. Tan, Q. Li, X. S. Xie, *Mol. Syst. Biol.* **11**, 844 (2015).
12. N. K. Hanchate *et al.*, *Science* **350**, 1251 (2015).
13. L. Tan, X. S. Xie, *Chem. Senses* **43**, 427 (2018).
14. S. Lomvardas *et al.*, *Cell* **126**, 403 (2006).
15. L. Tan, D. Xing, N. Daley, X. S. Xie, *Nat. Struct. Mol. Biol.* **26**, 297 (2019).
16. I. Solovei *et al.*, *Cell* **137**, 356 (2009).

ACKNOWLEDGMENTS

I thank my thesis adviser, X. S. Xie, for being a fantastic mentor, a visionary scientist, and a wonderful human being. I am also grateful to all members of the Xie laboratory and to my collaborators.

10.1126/science.aaz7774



GRAND PRIZE WINNER

Longzhi Tan

Longzhi Tan received his undergraduate degree from Massachusetts Institute of Technology and a Ph.D. from Harvard University. He started his postdoctoral fellowship at Stanford University in 2019. His research combines single-cell genomics and optogenetics to study the 3D chromatin basis of neurodevelopment and behaviors and to develop new tools for chromatin biology.



CATEGORY WINNER: ECOLOGY AND ENVIRONMENT

Barbara C. Klump

Barbara Klump received her master's degree from Heidelberg University, Germany, and a Ph.D. from the University of St Andrews, UK. After completing a postdoctoral fellowship at the University of St Andrews, she moved to the Max Planck Institute of Animal Behavior in Radolfzell, Germany, where she is currently a postdoc in the Cognitive and Cultural Ecology Lab. Her research explores how a species's ecology shapes its behavioral repertoire.

www.sciencemag.org/content/366/6468/965.1

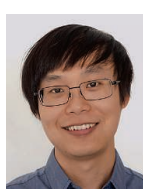


CATEGORY WINNER: MOLECULAR MEDICINE

Humsa Venkatesh

Humsa Venkatesh received her undergraduate degree from the University of California, Berkeley, and her Ph.D. from Stanford University. She is currently completing her postdoctoral fellowship at Stanford University. Her research combines principles of neuroscience and cancer biology to understand the electrical components of cancer pathophysiology and harness these malignant dependencies for therapeutic intervention.

www.sciencemag.org/content/366/6468/965.2



CATEGORY WINNER: CELL AND MOLECULAR BIOLOGY

Zibo Chen

Zibo Chen received his undergraduate degree from the National University of Singapore and his Ph.D. in biochemistry from the University of Washington. He is currently a postdoctoral scholar at the California Institute of Technology, where he is programming mammalian cells using proteins designed from scratch.

www.sciencemag.org/content/366/6468/965.3

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

PRIZE ESSAY



CATEGORY WINNER: ECOLOGY AND ENVIRONMENT

Barbara C. Klump

Barbara Klump received her master's degree from Heidelberg University, Germany, and a Ph.D. from the University of St Andrews, UK. After completing a postdoctoral fellowship at the University of St Andrews, she moved to the Max Planck Institute of Animal Behavior in Radolfzell, Germany, where she is currently a postdoc in the Cognitive and Cultural Ecology Lab. Her research explores how a species's ecology shapes its behavioral repertoire.

www.sciencemag.org/content/366/6468/965.1

ECOLOGY AND ENVIRONMENT

Of crows and tools

Tool-using crows, culture, and what it means to be human

By **Barbara C. Klump**

Hammers and chisels, pens and smartphones: Human life is built on tool use. Indeed, each of us likely uses tools every day. For a long time, crafting tools was thought to be uniquely human. And although we know that many species occasionally use tools, everyday tool use is extremely rare in the animal kingdom (*1*).

I study those rare animals that routinely use tools. My Ph.D. research took me to a small archipelago in the South Pacific called New Caledonia, the home of the New Caledonian (NC) crow, *Corvus moneduloides* (see the figure). Here, I explored the species's tool-oriented behavior using a series of noninvasive experiments with wild-caught birds.

NC crows are the only nonhuman species known to craft hooks in the wild (*1*). The invention of the hook was one of the key technological advances in our own evolutionary history in the middle stone age, allowing us to develop productive fishing technologies and weapons with enhanced killing power (*2*–*4*).

For the NC crow, a hook—delicately crafted from a forked branch (*5*, *6*)—is also used for "fishing." NC crows feed on insects and grubs, which they extract from deadwood. Using a hook tool allows them to get the prey several times faster than using a simple stick (*7*).

Like humans, NC crows must find suitable materials for their hook manufacture. They show a strong site-specific preference for the material they use to make these tools (*8*), but how do they identify their preferred plant species in the densely vegetated forest, where there might be hundreds of suitable options?

To assess which features of a plant crows select for, I presented wild-caught birds with experimental leaf-stem combinations and found that they can identify their preferred species from its stems alone. Some birds can even identify it from only the leaves (*9*). Intriguingly, this suggests that NC crows may have the equivalent of a "search image" in mind when looking for good tool material that includes features of the tool they intend to make.

Keeping already made tools safe when you don't need them can allow you to avoid having to make new tools. I discovered that NC crows look after their tools in between uses, especially when the cost or likelihood of losing the tool is high (*10*). Strategies to prevent tool loss can increase foraging efficiency and help remove some of the constraints experienced by arboreal tool users.

Human tools show cultural variation: Different means of achieving the same end are preserved for generations through social learning. NC crows vary in their hook tool manufacture (*11*), which raises the question of whether this variation is perhaps "cultural," as has been suggested for another NC crow tool type (*12*). But could the variation be explained by the properties of the available raw materials?

By giving crows the choice of material varying in thickness and rigidity, I showed that some of the observed variation in hook tool manufacture is indeed due to the material properties of the plants (*11*). My findings don't necessarily show that NC crows can't have cultural variation in their behavior, but they provide an alternative hypothesis for explaining regional differences observed in NC crows' tool making and, more generally, suggest another potential explanation for variation in animal tool behavior that appears to be culturally transmitted.

Despite the progress in understanding the NC crows' tool-oriented behavior, one mystery remained. Why do they, but no other crow species, use tools (*6*)?

A striking morphological feature of the NC crow is its unusually straight bill, which may reflect a specific adaptation for holding tools (*6*), as it allows the crow to position the tool tip in the center of binocular overlap (*13*). When checking the bill shapes of other living *Corvus* species, one with a relatively straight bill stood out: the Hawaiian crow or 'Alalā, *Corvus hawaiiensis* (*6*).

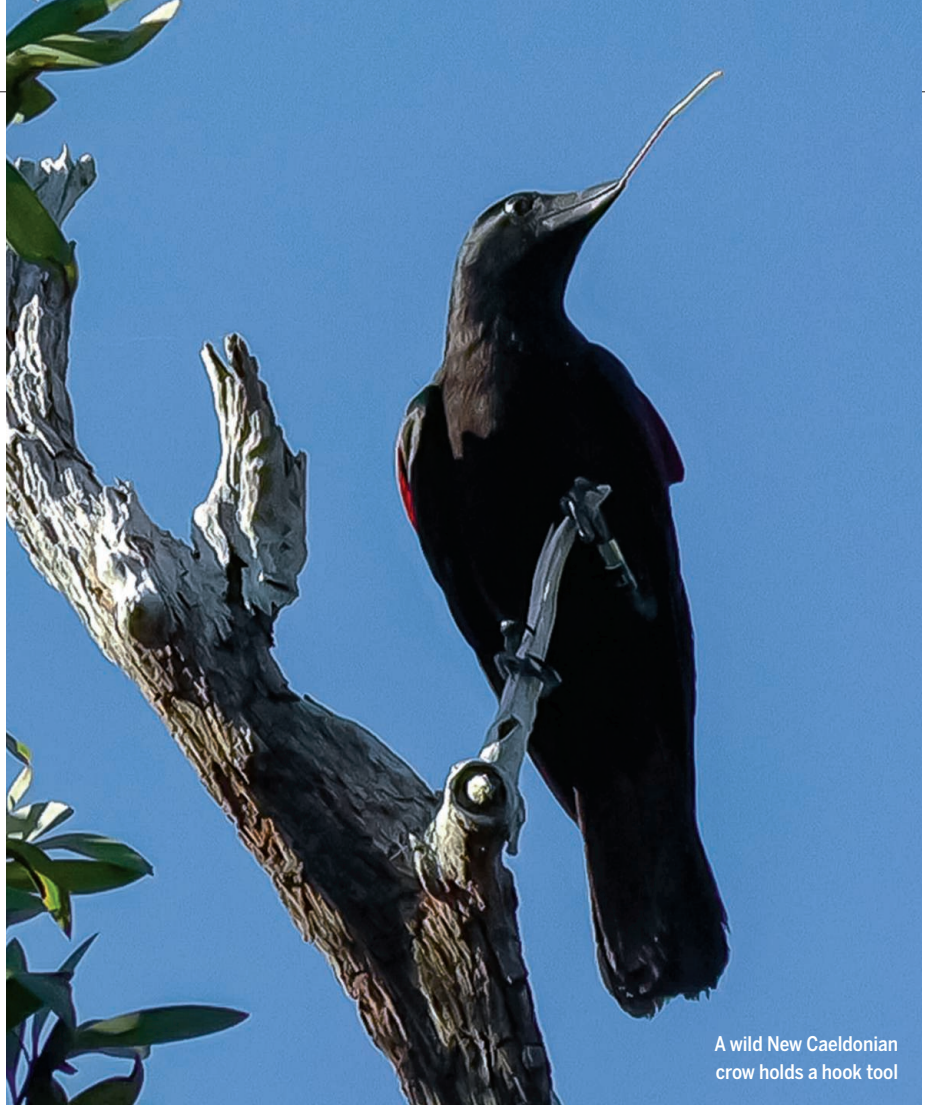
The 'Alalā was already extinct in the wild when we began our study—captive-reared birds have since been released—but San Diego Zoo Global has two breeding facilities in Hawai'i, where we tested all healthy

Cognitive and Cultural Ecology Laboratory, Max Planck Institute of Animal Behavior, 78315 Radolfzell, Germany, and Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, UK.
Email: bklump@ab.mpg.de

PHOTO: GUSTAVO ALARCÓN-NIETO

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS



A wild New Caeldonian crow holds a hook tool

PHOTO: PEDRO BARROS DA COSTA/RUTZ GROUP

individuals of the species (n = 104). We presented the birds with sticks and naturalistic extraction tasks and found that they are indeed very dexterous tool users: 93% of all adults and 47% of the younger birds spontaneously used sticks to probe for hidden food (*14*). They are not only very good tool users, they also exhibit an extensive repertoire of tool modification and manufacture skills (*15*). It is possible, then, that the marked population decline suffered by the ʻAlalā in the late 20th century (*16*) might have been the result of lost opportunities to forage with tools as human-caused habitat loss spread across Hawai'i (*14*). Since 2016, cohorts of captive-reared birds are being released in habitats that contain raw materials the ʻAlalā could use for tool building, and the future will show whether and how these charismatic birds will use tools in the wild.

It is probably no coincidence that these two distantly related (*17*) tool-using crow species evolved on remote tropical islands. With no woodpeckers, there is little competition for embedded food sources, and in the absence of big predators, the crows can spend less time maintaining vigilance and use more time to develop tool use (*6*). Remote islands, therefore, combine rare ecological conditions that seem to aid the evolution of foraging tool use in birds (*14*). The evolution of human tool use might have been similarly facilitated by the reduction in predation risk that accompanied the transition from solitary life to living in groups.

Our own tool-making journey is unparalleled, having enabled us to walk on the Moon in what amounts to an evolutionary eye-blink after the first fishing hook was crafted. But two tool-using crow species on remote tropical islands can still offer insight into how we became the master tool users we are today.

REFERENCES AND NOTES

1. R. W. Shumaker, K. R. Walkup, B. B. Beck, *Animal Tool Behavior: The Use and Manufacture of Tools by Animals* (Johns Hopkins Univ. Press, 2011).
2. J. E. Yellen, A. S. Brooks, E. Cornelissen, M. J. Mehlman, K. Stewart, *Science* **268**, 553 (1995).
3. S. O'Connor, R. Ono, C. Clarkson, *Science* **334**, 1117 (2011).
4. B. Barnes, B. J. N. Edwards, J. S. Hallam, A. J. Stuart, *Nature* **232**, 488 (1971).
5. G. R. Hunt, *Nature* **379**, 249 (1996).
6. C. Rutz, J. J. H. St. Clair, *Behav. Processes* **89**, 153 (2012).
7. J. J. H. St. Clair *et al*., *Nat. Ecol. Evol.* **2**, 441 (2018).
8. J. J. H. St. Clair, B. C. Klump, J. E. M. van der Wal, S. Sugasawa, C. Rutz, *Biol. J. Linn. Soc. London* **118**, 226 (2016).
9. B. C. Klump, M. Cantat, C. Rutz, *Biol. Lett.* **15**, 20180836 (2019).
10. B. C. Klump, J. E. M. van der Wal, J. J. H. St. Clair, C. Rutz, *Proc. R. Soc. London Ser. B* **282**, 20150278 (2015).
11. B. C. Klump, S. Sugasawa, J. J. St. Clair, C. Rutz, *BMC Biol.* **13**, 97 (2015).
12. G. R. Hunt, R. D. Gray, *Proc. R. Soc. London Ser. B* **270**, 867 (2003).
13. J. Troscianko, A. M. P. von Bayern, J. Chappell, C. Rutz, G. R. Martin, *Nat. Commun.* **3**, 1110 (2012).
14. C. Rutz *et al*., *Nature* **537**, 403 (2016).
15. B. C. Klump, B. M. Masuda, J. J. H. St. Clair, C. Rutz, *Commun. Integr. Biol.* **11**, e1509637 (2018).
16. P. C. Banko, D. L. Ball, W. E. Banko, in *The Birds of North America no. 648*, A. Poole, F. Gill, Eds. (Birds of North America Inc., 2002), pp. 1–27.
17. E. Haring, B. Däubl, W. Pinsker, A. Kryukov, A. Gamauf, *J. Zool. Syst. Evol. Res.* **50**, 230 (2012).

ACKNOWLEDGMENTS

I am extremely grateful to my Ph.D. supervisor, C. Rutz, for inspiration, professional guidance, and scientific advice. This research would not have been possible without his enthusiasm, interest in my work, and high degree of confidence in me. I also thank K. Cross for advice, as well as the Rutz group and everyone who supported me throughout my Ph.D. I thank our local partners in New Caledonia and Hawai'i: Province Sud, DENV, SEM Mwe-Ara, and San Diego Zoo Global. These studies were funded through scholarships to B.C.K. from the BBSRC and from the School of Biology, University of St Andrews, and a BBSRC David Phillips Fellowship (BB/G023913/1 and /2) to C. Rutz.

10.1126/science.aaz7775

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

PRIZE ESSAY



CATEGORY WINNER: MOLECULAR MEDICINE

Humsa Venkatesh

Humsa Venkatesh received her undergraduate degree from the University of California, Berkeley, and her Ph.D. from Stanford University. She is currently completing her postdoctoral fellowship at Stanford University. Her research combines principles of neuroscience and cancer biology to understand the electrical components of cancer pathophysiology and harness these malignant dependencies for therapeutic intervention.

www.sciencemag.org/content/366/6468/965.2

MOLECULAR MEDICINE

The neural regulation of cancer

Cancers hijack mechanisms of neural plasticity to promote malignant disease progression

By **Humsa S. Venkatesh**

Cancer has been called the "emperor of all maladies" (*1*). An estimated 40% of us will be diagnosed with it at some point in our lives (*2*). We have all witnessed the irreverent brutality of this disease that seems to know no bounds. Like weeds in a garden, cancers emerge suddenly and proliferate aggressively by co-opting vital systems necessary for survival. Although current treatment approaches have targeted the vascular and immune systems as well as cell-intrinsic growth mechanisms, the nervous system's role in cancer growth has been largely underrecognized. My research aims to combine principles of neuroscience and cancer biology to harness microenvironmental dependencies of tumors and disable hijacked mechanisms of neural plasticity.

Science & SciLifeLab PRIZE For Young Scientists

Cancers, at their core, survive by recapitulating mechanisms of normal development. Activity of the nervous system has been well established as a powerful modulator of neurodevelopment and plasticity. Electrical activity shapes brain organogenesis as well as the behavior of persistent populations of neural precursor cells (NPCs) in pediatric and adult brains (*3*). In the context of cancer, we hypothesized that this neurodevelopmental principle may similarly inform tumor progression.

Gliomas provide a compelling case for studying these microenvironmental cues because these cancers arise in striking spatiotemporal patterns. Given the robust influence of active neurons on NPC populations, many of which are candidate cells of origin for gliomas, we sought to uncover a relationship between neuronal activity and glioma growth.

UNDERSTANDING THE ROLE OF THE NERVOUS SYSTEM

In the Monje laboratory at Stanford University, we created a first-of-its-kind mouse model that enables both transplantation of patient-derived glioma cells and control of neuronal activity through optogenetic stimulation. This platform facilitated manipulation of the neuronal component of the tumor microenvironment and isolation of its direct effect on tumor biology. Using this model, we found that neuronal activity robustly promotes circuit-specific glioma growth and progression (*4*). These findings demonstrated, for the first time, the critical role of neurons in the brain tumor microenvironment.

To study the molecular mechanism mediating this effect, we developed an acute optogenetic slice culture model in combination with biochemical and proteomic assays to isolate a small list of specific activity-dependent secreted proteins. Soluble neuroligin-3 (NLGN3; a cleaved fragment of the full-length version normally involved in synaptic adhesion) was identified as a differentially secreted protein. It had the most robust effect on glioma proliferation through phosphatidylinositol 3-kinase (PI3K) pathway activation and, interestingly, induced an up-regulation of synapse-related genes in gliomas (*4*, *5*). Thus, soluble extracellular NLGN3, previously unknown to be secreted or involved in growth, was identified as a potent activity-dependent mitogen.

IDENTIFYING A THERAPEUTIC TARGET

Clinical analysis of hundreds of adult glioma patient samples revealed that higher *NLGN3* expression levels predicted a substantially worse prognosis, indicating an important role for NLGN3 in glioma pathogenesis. To verify this necessity, we examined multiple patient-derived pediatric and adult gliomas in mice genetically deficient of NLGN3. Strikingly, gliomas failed to grow in the absence of NLGN3, emphasizing its vital requirement in the tumor neuromicroenvironment and its promise as a therapeutic target (*5*). Our focus subsequently shifted to studying this unknown mechanism of NLGN3 secretion.

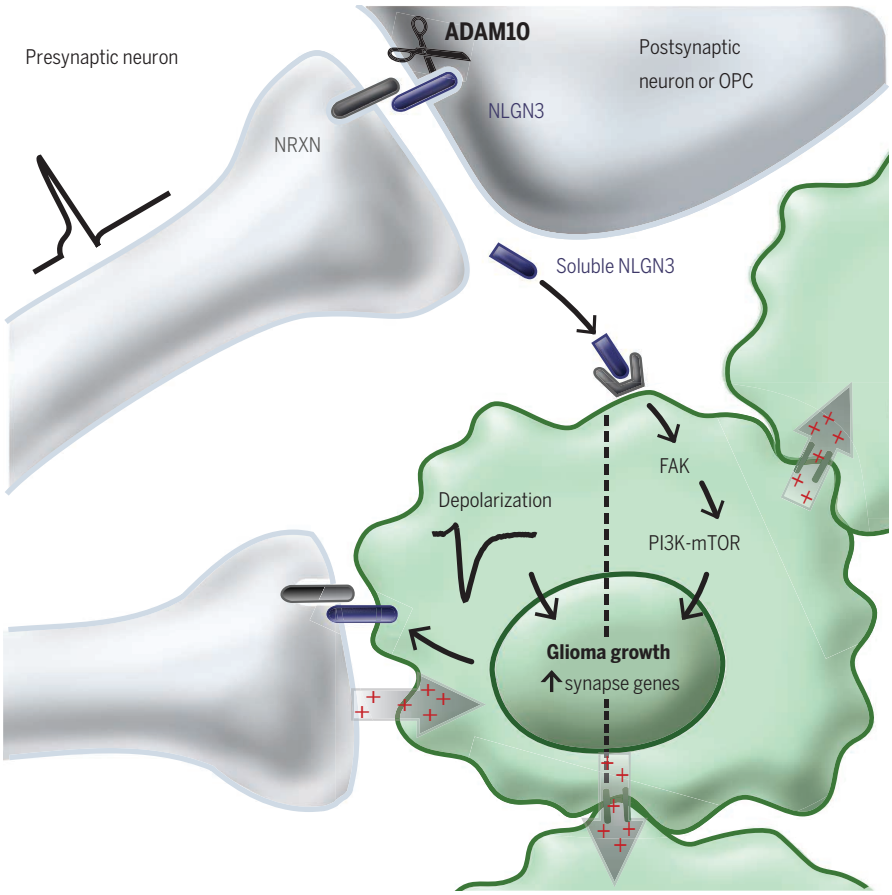
Using a series of genetic mouse models and protease inhibitor analyses, we identified ADAM10 (A disintegrin and metalloproteinase domain-containing protein 10) as the enzyme responsible for neuronal activity-dependent cleavage and secretion of NLGN3 from the membrane of neurons and oligodendrocyte precursor cells (OPCs) as well as from the glioma cells themselves. The enzyme, we surmised, was thus an attractive therapeutic target. When tumor-bearing

Neurology and Neurological Sciences, Stanford University, Stanford, CA 94305. Email: humsav@stanford.edu

Downloaded from http://science.sciencemag.org/ on November 28, 2019

PHOTO: COURTESY OF HUMSA VENKATESH

Published by AAAS



Neuronal activity drives cancer growth and progression through direct signaling mechanisms and the functional integration of glioma into neural circuitry. An important mechanism that mediates the neural regulation of brain cancer is activity-dependent cleavage and secretion of the synaptic adhesion molecule neuroligin-3 (NLGN3), which promotes glioma proliferation through the PI3K–mammalian target of rapamycin (mTOR) pathway and induces up-regulation of several synapse-related genes. NLGN3 is cleaved through the ADAM10 sheddase. ADAM10 inhibitors prevent the release of NLGN3 into the tumor microenvironment and robustly block cancer growth. Additional neuron-glioma interactions include electrical and synaptic integration of glioma into neural circuitry through direct electrochemical synaptic transmission. Synaptic activity depolarizes electrically coupled glioma cell networks, and this depolarization in turn induces proliferation. These findings suggest that neuron-glioma circuit dynamics are integral to tumor progression. [Figure adapted from (*5*)]

mice were treated with an ADAM10 inhibitor, glioma growth stagnated as it had in the genetically deficient NLGN3 models (*5*). These results formed the rationale for upcoming clinical trials for the treatment of both pediatric and adult high-grade glioma patients with an ADAM10 inhibitor.

ADDITIONAL NEURON-GLIOMA SYNAPTIC COMMUNICATION

The stark reduction in glioma growth after NLGN3 removal suggests that the protein plays a role more fundamental to glioma progression, well beyond its role as a mitogen. Given its function in the postsynaptic assembly in normal cells and its involvement in the up-regulation of synaptic genes in gliomas, we hypothesized that NLGN3 may be facilitating direct glioma engagement with neurons.

Neurons form functional synapses with OPCs, and many gliomas molecularly and phenotypically resemble these OPCs. Do malignant glioma cells, similar to their glial counterparts, form functional synapses with neurons to further hijack mechanisms of normal neural plasticity, we wondered?

To answer this question, we used single-cell transcriptomics and immuno-electron microscopy techniques to identify synaptic interactions between neurons and malignant cells. Whole-cell patch clamp recordings of xenografted glioma cells with simultaneous stimulation of afferent neurons subsequently revealed AMPA receptor–dependent postsynaptic currents in the glioma cells (*6*). Further, using real-time two-photon calcium imaging, we uncovered rapid depolarization and ion flow through gap-junction–connected tumor cell networks in response to neuronal stimulation. These data suggest that gliomas functionally integrate into electrically active neural circuits through neuron-to-glioma synapses and that the effects of neuron-glioma signaling are amplified throughout the tumor network by means of gap junction–mediated connections (*6*).

Membrane depolarization of normal neural precursor cells can regulate proliferation, differentiation, and survival. Thus, we hypothesized that depolarization similarly promotes glioma growth. Using in vivo optogenetic techniques to depolarize xenografted glioma cells, we found that glioma membrane depolarization robustly promoted proliferation, whereas pharmacologically or genetically blocking electrochemical signaling inhibited glioma xenograft growth and extended mouse survival (*6*).

Together, these results demonstrate, for the first time, that electrical circuit integration of glioma promotes its progression (see the figure).

A NOVEL CANCER TREATMENT STRATEGY

The study of the neural regulation of cancers is a burgeoning field that highlights the nervous system's central role in facilitating tumor progression. Unveiling other critical mechanistic details of the neurobiology of cancers will help us further understand how cancers exploit these powerful microenvironmental interactions. My work identifies activity-dependent mitogen secretion, synaptic neurotransmission, and gap junction–mediated electrical coupling as previously unknown mechanisms controlling glioma progression.

This demonstration of the critical role of neural elements in the microenvironment has been extended to prostate (*7*), pancreatic (*8*), skin (*9*), and gastric cancers (*10*, *11*), illustrating its expansive applicability. Taken together, we, and others, demonstrate that targeting neuron-cancer circuit dynamics represents a promising therapeutic intervention.

REFERENCES AND NOTES

1. S. Mukherjee, *The Emperor of All Maladies: A Biography of Cancer* (Scribner, 2010).
2. National Cancer Institute, data from 2013–2015; https://www.cancer.gov/about-cancer/understanding/statistics.
3. H. S. Venkatesh, M. Monje, *Trends in Cancer* **3**, 89 (2017).
4. H. S. Venkatesh *et al.*, *Cell* **161**, 803 (2015).
5. H. S. Venkatesh *et al.*, *Nature* **549**, 533 (2017).
6. H. S. Venkatesh *et al.*, *Nature* **573**, 539 (2019).
7. C. Magnon *et al.*, *Science* **341**, 1236361 (2013).
8. B. W. Renz *et al.*, *Cancer Cell* **33**, 75 (2018).
9. S. C. Peterson *et al.*, *Cell Stem Cell* **16**, 400 (2015).
10. C.-M. Zhao *et al.*, *Sci. Transl. Med.* **6**, 250ra115 (2014).
11. Y. Hayakawa *et al.*, *Cancer Cell* **31**, 21 (2017).

ACKNOWLEDGMENTS

I am very grateful to M. Monje, whose mentorship has been instrumental in shaping me as a scientist. I also thank R. Malenka, W. Morishita, the members of the Monje Laboratory, and my collaborators for their guidance and friendship.

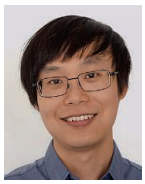
10.1126/science.aaz7776

GRAPHIC: N. DESAI/*SCIENCE* FROM H. VENKATESH

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS

PRIZE ESSAY



CATEGORY WINNER: CELL AND MOLECULAR BIOLOGY

Zibo Chen

Zibo Chen received his undergraduate degree from the National University of Singapore and his Ph.D. in biochemistry from the University of Washington. He is currently a postdoctoral scholar at the California Institute of Technology, where he is programming mammalian cells using proteins designed from scratch.

www.sciencemag.org/content/366/6461/965.3

CELL AND MOLECULAR BIOLOGY

Creating the protein version of DNA base pairing

Programmable and modular protein-protein interactions designed from scratch

By **Zibo Chen**

Specific interactions between molecules give rise to life. Nucleic acid is a good example: The pairing of adenine (A) to thymine (T) and cytosine (C) to guanine (G) ensures accurate passage of genetic materials over generations. Thanks to the programmable and modular nature of A-T/C-G base pairing, DNA has been repurposed to build nanoscopic smiley faces (*1*) and robots that sort molecular cargoes (*2*), giving rise to the field of DNA nanotechnology.

As an undergraduate student, I became intrigued by the possibility of engineering protein-protein interactions (PPIs) in a similarly modular manner. However, I found that there was no straightforward way to artificially encode specificity in PPIs, which are often mediated by multiple side chains in a structurally nonmodular fashion. [Two features contribute to the modular specificity of DNA base pairing: buried hydrogen bonds between the bases (specificity), and the repetitive double helical backbone (modularity).]

As a first-year graduate student working in the labs of David Baker and Frank DiMaio at the University of Washington, I contributed to a project that aimed to design specific hydrogen bonds (mediated by polar amino acid side chains) at the binding interfaces between de novo designed protein α-helical bundles. We computationally designed and experimentally validated homodimers, trimers, and tetramers with such structural features (*3*). This work showcased the engineering of a new interaction modality—the protein equivalent of DNA base pairing. I went on to pursue three projects in parallel. Each project (described below) represented a key aspect of programmable protein interactions.

PROGRAMMABLE DESIGN OF ORTHOGONAL PROTEIN HETERODIMERS

To push the limit of designed specificity in proteins, I decided to design a big set of mutually orthogonal protein heterodimers (see the figure). Heterodimers are more difficult to create, as the designer needs to not only account for structurally accurate heterodimerization, but also disfavor each monomer's tendency to homodimerize. I began this project with the hypothesis that two monomers could only heterodimerize if their hydrogen-bonding patterns match exactly at the binding interface.

Using large-scale computational sampling, I demonstrated that protein heterodimeric interaction specificity can be achieved using extensive hydrogen-bond networks (see the figure). The specificity and modularity of these networks recapitulates those of Watson-Crick base pairing, but in the context of proteins. Furthermore, I mixed 14 pairs of heterodimers in a single test tube and found that only the designed pairs associated, with almost no spurious cross-talk—a property that is taken for granted with DNA but had been missing in designed proteins (*4*).

DE NOVO DESIGN OF PROTEIN LOGIC GATES

Until now, the lack of available orthogonal protein pairs has made it difficult to control cell signaling at the posttranslational level (*5*, *6*) and has limited the field of synthetic biology to mainly transcriptional regulation (*7*–*9*). Once we have a big set of mutually orthogonal proteins at hand, as *Star Trek*'s Spock would no doubt agree, "it was only logical" to design protein-based logic circuits in cells.

By modularly combining the set of orthogonal protein heterodimers, I designed two-input AND, OR, NAND, NOR, XNOR, and NOT logic gates that could regulate the association of arbitrary protein units ranging from split enzymes to transcriptional machinery both in vitro and in vivo. I further demonstrated the modularity of this approach by extending it to various three-input logic gates.

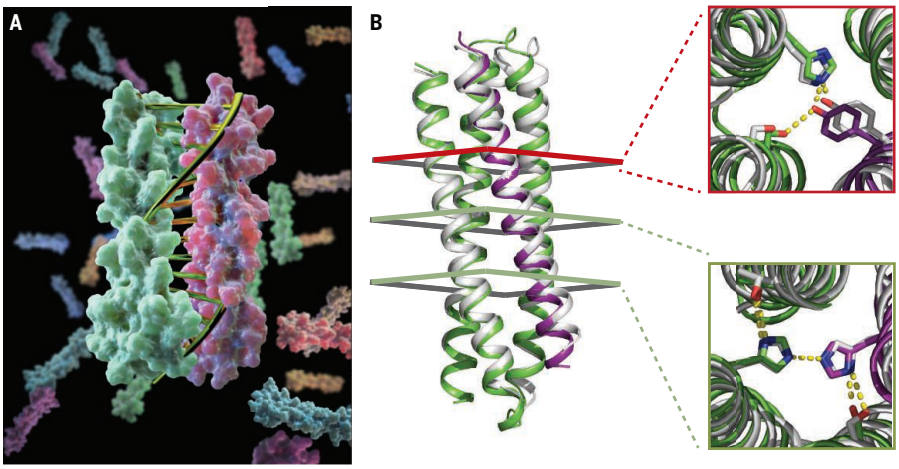
While experimentally characterizing these logic gates, I discovered that they operate in a cooperative manner: One protein does not bind another without a third protein that

Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. Email: zibochen@caltech.edu

PHOTO: SHUTING SHAO

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS



(A) An illustration of the protein version of DNA base pairing. Interactions among color-coded protein monomers are mediated by hydrogen bonds much like those found in DNA. Two monomers bind only if there is perfect pairing of hydrogen bonding patterns at the binding interface. (B) A heterodimer design model with monomers colored green and purple, superimposed on its crystal structure (white); colored cross-sections of backbones (red and green) indicate the locations of designed hydrogen-bond networks.

opens up a binding interface. Binding interaction cooperativity makes the gates largely insensitive to stoichiometric imbalance in the inputs, making them ideal for cellular applications. The modularity and cooperativity of these control elements, coupled with the ability to de novo design an essentially unlimited number of protein components, should enable the creation of sophisticated posttranslational control logic over a wide range of biological functions.

SELF-ASSEMBLING 2D MATERIALS WITH DE NOVO PROTEIN BUILDING BLOCKS

Creating two-dimensional (2D) materials is straightforward with DNA but has been difficult to achieve using proteins, owing to the lack of modular specificity in their interactions. I developed a general computational algorithm to design 2D materials using de novo designed protein building blocks.

By optimizing the sequences at binding interfaces, I showed that the same protein building block can be reconfigured into two different array geometries. The single-layered 2D arrays assembled to micrometer scale under electron microscopy and atomic force microscopy, and displayed perfect agreement with the designed lattice geometry (*10*).

This work, describing how a de novo designed protein is used as a building block to construct 2D materials, opens the door to the creation of programmable protein-based materials.

TOWARD A FULLY SYNTHETIC FUTURE

More than 60 years have passed since the discovery of Watson-Crick base pairing, and we are beginning to engineer such features into proteins. Now that we have proteins that behave in a manner similar to that of DNA molecules, the timing could not be better to start exploring synthetic biology using de novo designed proteins.

As a start, we designed proteins that are sensitive to pH changes and capable of endosomal escape (*11*); proteins that insert into cell membranes and can potentially become synthetic ion channels (*12*); and proteins that change conformation to send out a customizable signal upon binding to small peptides (*13*). I'm tremendously excited to see what comes next.

REFERENCES AND NOTES

1. P. W. K. Rothemund, *Nature* **440**, 297 (2006).
2. A. J. Thubagere *et al.*, *Science* **357**, eaan6558 (2017).
3. S. E. Boyken *et al.*, *Science* **352**, 680 (2016).
4. Z. Chen *et al.*, *Nature* **565**, 106 (2019).
5. X. J. Gao, L. S. Chong, M. S. Kim, M. B. Elowitz, *Science* **361**, 1252 (2018).
6. T. Fink *et al.*, *Nat. Chem. Biol.* **15**, 115 (2019).
7. K. Yehl, T. Lu, *Curr. Opin. Biomed. Eng* **4**, 143 (2017).
8. M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
9. T. S. Gardner, C. R. Cantor, J. J. Collins, *Nature* **403**, 339 (2000).
10. Z. Chen *et al.*, *J. Am. Chem. Soc.* **141**, 8891 (2019).
11. S. E. Boyken *et al.*, *Science* **364**, 658 (2019).
12. P. Lu *et al.*, *Science* **359**, 1042 (2018).
13. R. A. Langan *et al.*, *Nature* **572**, 205 (2019).

ACKNOWLEDGMENTS

I am grateful to my Ph.D. advisers D. Baker and F. DiMaio. I also thank S. Boyken for his mentorship.

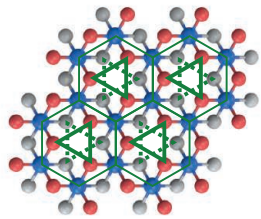
10.1126/science.aaz7777

Downloaded from http://science.sciencemag.org/ on November 28, 2019

GRAPHIC: N. DESAI/*SCIENCE* FROM DEMIN LIU, MOLGRAPHICS

Published by AAAS

RESEARCH



Magnetism in a bilayer material

Chen et al., *p. 983*

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



A split concretion reveals fossil remains of a vertebrate skull.

PALEONTOLOGY

Terrestrial record of recovery

The extinction that occurred at the end of the Cretaceous period is best known as the end of the nonavian dinosaurs. In theory, this paved the way for the expansion of mammals as well as other taxa, including plants. However, there are very few direct records of loss and recovery of biotic diversity across this event. Lyson *et al*. describe a new record from the Cretaceous-Paleogene in Colorado that includes unusually complete vertebrate and plant fossils that describe this event in detail, including the recovery and expansion of mammalian body size and increasing plant and animal biotic diversity within the first million years. —SNV

Science, this issue p. 977

SIGNAL TRANSDUCTION

Cellular response to nutrient status

An intricate regulatory mechanism is taking shape around control of the mechanistic target of rapamycin complex 1 (mTORC1) protein kinase complex. Physiological responses of cells to nutrient abundance is regulated by signaling through mTORC1, which interacts with a complex of other proteins at the lysosome that includes small guanosine triphosphatases (GTPases), GTPase-activating proteins, and others. Lawrence *et al*. present a cryo–electron microscopy structure and biochemical experiments that reveal that the small GTPases act in unanticipated ways in the complex. The tumor suppressor folliculin (FLCN), along with FLCN-interacting protein 2, interacts with its partner GTPase RagC in both active and inactive states, suggesting that a particularly elaborate and stringent regulatory mechanism is at work. —LBR

Science, this issue p. 971

PLANT SCIENCE

Common gene yields different structures

Nodules form on legume roots to house symbiotic, nitrogen-fixing bacteria. Lateral roots, characteristic of a much broader range of plants, extend to take up nutrients and water from the soil. Soyano *et al*. found common ground in the developmental pathways that build nodules and lateral roots (see the Perspective by Bishopp and Bennett). Evidence from *Lotus japonicus*, a legume that can fix atmospheric nitrogen, shows that the nodule-forming pathway shares components with the lateral root pathway. —PJH

Science, this issue p. 1021; see also p. 953

IMMUNOMETABOLISM

A fresh look at glutamine targeting

Glutamine is essential for tumor growth and has long been an attractive therapeutic target for cancer researchers. Some attempts at blocking glutamine metabolism in cancer patients resulted in toxicity, prompting Leone *et al*. to develop an innovative approach to reduce general side effects. They designed a prodrug form (JHU083) of the glutamine antagonist 6-diazo-5-oxo-L-norleucine (DON), which is administered in an inert state but then preferentially activated by enzymes enriched in the tumor microenvironment. JHU083 simultaneously shut down glycolysis and oxidative phosphorylation in mouse cancer cells while enhancing T cell oxidative phosphorylation and anticancer immune responses. —PNK

Science, this issue p. 1013

CREDITS: (GRAPHIC) CHEN *ET AL.*; (PHOTO) HHMI TANGLED BANK STUDIOS

Published by AAAS

IN OTHER JOURNALS

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

COMPUTER SCIENCE

Making well-behaved algorithms

Machine learning algorithms are being used in an ever-increasing number of applications, and many of these applications affect quality of life. Yet such algorithms often exhibit undesirable behavior, from various types of bias to causing financial loss or delaying medical diagnoses. In standard machine learning approaches, the burden of avoiding this harmful behavior is placed on the user of the algorithm, who most often is not a computer scientist. Thomas *et al*. introduce a general framework for algorithm design in which this burden is shifted from the user to the designer of the algorithm. The researchers illustrate the benefits of their approach using examples in gender fairness and diabetes management. —JS

Science, this issue p. 999

LEISHMANIASIS

Calling healing lesions

Cutaneous leishmaniasis results in skin sores that can be difficult to treat, resulting in further complications. Farias Amorim *et al*. performed dual RNA-sequencing of host and parasite gene expression in pretreatment lesional skin biopsies from two cohorts of patients infected with the parasite *Leishmania braziliensis*. A prognostic signature comprising expression of three cytolytic genes plus pathogen load predicted treatment response in both cohorts and could potentially be used to triage patients who are unlikely to respond to conventional treatment as candidates for alternate therapies. This study also provides evidence that inflammasome-mediated interleukin-1β production influences treatment outcome in patients, strengthening the argument that therapies targeting these components could treat cutaneous leishmaniasis. —CAC

Sci. Transl. Med. **11**, eaax4204 (2019).

CORRELATED ELECTRONS

Looking for a strange metal

In many materials, charge carriers are well described as noninteracting quasiparticles. However, in materials with strong correlations, this approximation can break down, leading to anomalous transport properties at high temperatures. Huang *et al*. used quantum Monte Carlo calculations to look for this so-called strange metal phase in the simplest two-dimensional model of interacting electrons, the Hubbard model. They found that the calculated resistivity had a linear temperature dependence when hole doping was introduced, as expected in the strange metal phase. This observation provides confidence that simplified models can be used to describe and understand the behavior of real materials, such as cuprate high-temperature superconductors. —JS

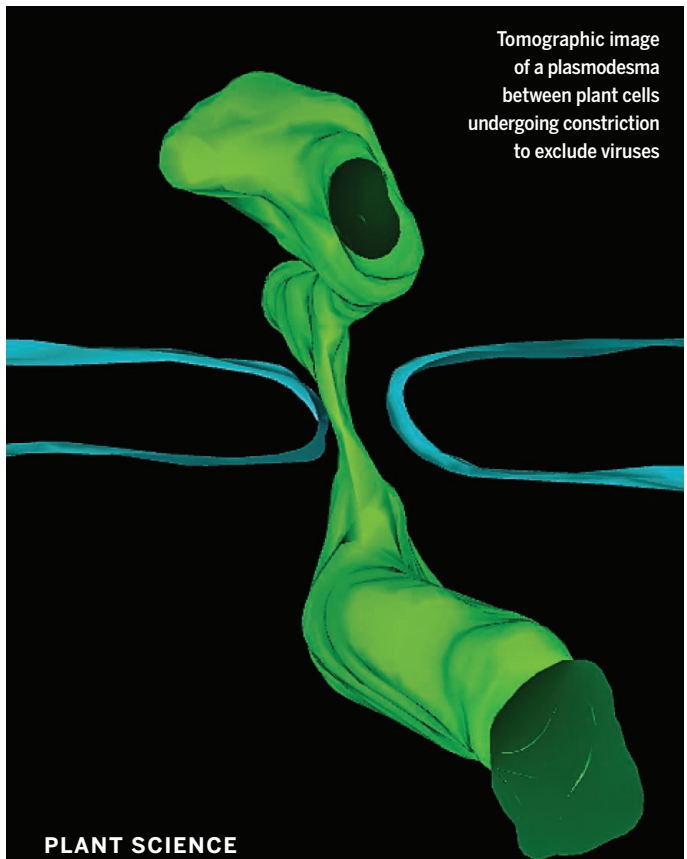
Science, this issue p. 987

CANCER

Mixed signals at tumor margins

The Hippo signaling pathway has been implicated in tumor growth, sparking interest in the pathway as a potential therapeutic target. In a study of liver cancer in genetically manipulated mice, Moya *et al*. discovered that the role of this pathway in tumorigenesis is more complex than previously appreciated. They confirmed that activation of the Hippo pathway within tumor cells drives tumor growth; however, they also found that activation of the pathway in adjacent healthy cells has the opposite effect, suppressing tumor growth. Whether tumor cells survive or are eliminated thus appears to depend on competing signals produced by the tumor and surrounding tissue. —PAK

Science, this issue p. 1029



Tomographic image of a plasmodesma between plant cells undergoing constriction to exclude viruses

CREDIT: HUANG *ET AL*., *PROC. NATL. ACAD. SCI. U.S.A.* **116**, 42 2019

PLANT SCIENCE

Shut the door on virus spread

Neighboring cells in a plant are linked by plasmodesmata that form tunnels through which the good, the bad, and the ugly can flow from one cell to another. During viral infection, defensive signals and viruses race through the resulting network of linked cells. The plant uses salicylic acid (SA) to signal the plasmodesmata to close off routes for infecting viruses. Huang *et al*. show that SA drives shifts in patches of lipid in the membranes of plasmodesmata through a mechanism regulated by the remorin protein family. The resulting lipid rafts stiffen the plasma membrane, and the plasmodesmata freeze into dysfunction, thus shutting the door on marauding viruses. —PJH

Proc. Natl. Acad. Sci. U.S.A. **116**, 21274 (2019).

BIOMEDICINE

Mechanical tissue matching

Implantable electronic devices that can precisely detect metabolites and biomarkers of disease in real time are an exciting prospect. However, ensuring that such devices stably interact with human tissue without causing damage is difficult because the devices need to bend and adapt to the motions of the body. To overcome this impediment, Wang *et al*. developed an implantable electrochemical sensor composed of flexible helical bundles of carbon nanotubes that mimic the fibril structure of muscle. When injected into target

Published by AAAS



NATURAL HAZARDS

Irrigation endangerment

Earthquakes can trigger deadly and damaging landslides. Landslides during the 2018 earthquake at Palu, Indonesia, were responsible for many casualties. Bradley *et al*. identified irrigation from wet rice cultivation as an important factor that increased the landslide hazard. Irrigation changes the properties of soil and makes it more prone to a process by which it behaves like a liquid during ground shaking, making hillslopes more prone to landslides. Recognizing and avoiding irrigation practices that raise the water table where earthquake-triggered landslides may occur could help mitigate this hazard. —BG

Nat. Geosci. **12**, 935 (2019).

Landslide damage from the 2018 Palu, Indonesia, earthquake

tissues in cats, the bundles formed a stable interface with the tissue and could detect and transmit chemical concentrations in real time through Bluetooth. Matching the mechanical properties of tissue should thus allow implantable devices to be used long-term in humans. —GKA

Nat. Biomed. Eng. 10.1038/s41551-019-0462-8 (2019).

AGING

TRIIM trial to target thymus

Thanks to the demographic transition, aging human bodies are an increasing liability for health and social systems. The risk of many age-related diseases could be reduced by thymus regeneration to improve immunological protection. In the thymus regeneration, immunorestoration, and insulin mitigation (TRIIM) trial, Fahy *et al*. developed a method to promote thymus function and reverse immunosenescence aging markers in nine men aged 51 to 65 years. The men were treated with recombinant human growth hormone as well as two drugs (dehydroepiandrosterone and metformin) that are used to prevent hyperinsulinemia and the diabetogenic effects of growth hormone. Effects of treatment were measured on immune cells, cytokines, and epigenetic clocks, which monitor changes in DNA methylation and indicate biological age (as distinct from chronological age). Immunological measures improved in the treated men, and epigenetic age regressed by 1.5 years after 1 year of treatment. Larger studies are required to verify the findings. —LBR

Aging Cell **18**, e13028 (2019).

MAMMALIAN GENOMICS

All is not lost within the genome

The evolutionary loss of genes among species can occur if a gene is no longer needed. Paradoxically, many diseases in humans have been attributed to a loss of functional genes. Sharma and Hiller studied the genomes of a collection of 62 nonhuman mammals. They found that up to 22% of genes appear to have been lost in at least one species. Some gene losses may result in traits that associate with disease in humans—such as a lack of tooth enamel or impaired vision—but are not deleterious to the species concerned because of their specific lifestyles. Many of the missing genes affect olfaction and the immune system and do not affect the viability of human cells or mice when experimentally inactivated. —LMZ

NAR Genom. Bioinform. 10.1093/nargab/lqz012 (2019).

NANOPHOTONICS

Quantum light from a nanoantenna

Antennas are crucial components in wireless communication, allowing information to be transferred between senders and receivers. Recent developments have seen nanometer-scale optical antennas that allow for the efficient extraction of photons, plucking them out of a light beam and funneling them into the circuits of an optical chip or waveguide, and vice versa. Using a nanoantenna comprising an aluminium gallium arsenide (AlGaAs) nanocylinder, Marino *et al*. demonstrate the generation of photon pairs as the structure is excited by a pump laser. Antennas that can also be used as quantum light sources provide a useful route toward developing an integrated platform for quantum communication and sensing. —ISO

Optica **6**, 1416 (2019).

MASS EXTINCTIONS

Impact acidification

Surface-ocean acidification following the Chicxulub impact may have contributed to ecological collapse in the marine realm at the end of the Cretaceous. By measuring the boron isotopic composition of foraminifera, Henehan *et al*. show that surface-ocean pH fell rapidly—in a geological blink of the eye—after the bolide struck Earth and then recovered to preimpact values over the following hundred thousand years or so. Modeling these results suggests that marine primary productivity was reduced after impact and that lingering inefficiency in carbon export to the deep sea was responsible for the slower biogeochemical recovery. —HJS

Proc. Natl. Acad. Sci. U.S.A. **116**, 22500 (2019).

PHOTO: JEWEL SAMAD/GETTY IMAGES

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

NANOMATERIALS

Thinking small to store more

From mobile devices to the power grid, the needs for high-energy density or high-power density energy storage materials continue to grow. Materials that have at least one dimension on the nanometer scale offer opportunities for enhanced energy storage, although there are also challenges relating to, for example, stability and manufacturing. In this context, Pomerantseva *et al*. review fundamental processes of charge storage that apply specifically to nanostructured materials and briefly explore potential manufacturing processes. The authors also consider some of the skepticism, such as that found in the battery community, to the use of these materials. —MSL

Science, this issue p. 969

2D MAGNETISM

Stacking control

Recent searches for two-dimensional magnets have turned up a number of related van der Waals materials, such as chromium triiodide (CrI_3) and chromium tribromide ($CrBr_3$). Although their properties are expected to be similar, in bilayer form, the former is antiferromagnetic, whereas the latter appears to be a ferromagnet. Chen *et al*. used spin-polarized scanning tunneling microscopy to determine that the nature of the magnetic state in bilayer $CrBr_3$ depends on the type of stacking of its monolayers. An antiferromagnetic state formed when the two monolayers were oriented in the same direction, whereas the opposite orientation led to ferromagnetism. —JS

Science, this issue p. 983

PROTEIN DESIGN

Sense and respond

Many signaling pathways start with cellular proteins sensing and responding to small molecules. Despite advances in protein design, creating a protein-based sense-and-respond system remains challenging. Glasgow *et al*. designed binding sites at the interface of protein heterodimers (see the Perspective by Chica). By fusing each monomer to one half of a split reporter, they linked ligand-driven dimerization to the reporter output. The computational design strategy provides a generalizable approach to create synthetic sensing systems with different outputs. —VV

Science, this issue p. 1024; see also p. 952

ORGANIC CHEMISTRY

A tag team approach to forming C–N bonds

Many pharmaceutical compounds contain carbon-nitrogen (C–N) bonds in just one of two mirror-image orientations. Forging these bonds with electron-rich nitrogen reactants is challenging because the nitrogen groups can coordinate with, and thus interfere with, the catalyst. Li *et al*. report a cooperative approach to overcoming this obstacle (see the Perspective by Ovian and Jacobsen). They used a copper catalyst to activate the carbon reactant and then a hydrogen-bonding thiourea catalyst to set the product geometry with high selectivity. The reaction is compatible with a broad range of diazo ester and amine coupling partners. —JSY

Science, this issue p. 990; see also p. 948

INFECTION

Magnesium for *Salmonella* growth

In mammals, macrophages are a first line of defense against pathogenic bacteria like *Salmonella*. These immune cells possess a metal-ion transporter called SLC11A1 or NRAMP1, which is known to be involved in infection resistance. Cunrath and Bumann studied mice that only differ in one allele of the *SLC11A1* gene, which leaves mice either susceptible or resistant to infection. Proteomic analyses showed that bacteria isolated from mice with functional *SLC11A1* alleles experienced metal starvation, particularly of magnesium. The resulting impairment of bacterial growth seems to be the primary mode of action of SLC11A1 against invading pathogens. —CA

Science, this issue p. 995

TRAIT DEVELOPMENT

Where do horns come from?

One of the most pronounced examples of a sexually selected trait is the prothoracic horns of scarab beetles, which, in the most extreme cases, can be nearly half as long as the length of the beetle. It is fairly easy to understand how selection might have shaped these horns, but understanding how development shaped them from a hornless ancestor is a much more complex proposition. Hu *et al*. show that these horns are generated from wing homologs and argue that many other insect traits may have followed similar transcriptional paths (see the Perspective by Nijhout). —SNV

Science, this issue p. 1004; see also p. 946

NEUROSCIENCE

A brain circuit to control alcohol intake

Most people are exposed to alcohol at some point in their lives, but only a small fraction will develop a compulsive drinking disorder. Siciliano *et al*. first established a behavioral measure to assess how predisposition interacts with experience to produce compulsive drinking in a subset of mice (see the Perspective by Nixon and Mangieri). In search of the underlying neurobiological mechanism, they discovered that a discrete circuit between the medial prefrontal cortex and brainstem is central for the development of compulsive drinking. This circuit serves as both a biomarker for the development of compulsive drinking and a driver of its expression. It can bidirectionally control compulsive behavior by mitigating or mimicking punishment signals. —PRS

Science, this issue p. 1008; see also p. 947

GLOBAL HEALTH

AI in resource-poor health care systems

The increasing development of artificial intelligence (AI) applications for health care is mostly relevant to health care systems in high-income countries. In a Perspective, Hosny and Aerts discuss how AI applications can be developed to support health care in low- and middle-income countries. With increasing rates of noncommunicable diseases, such as cardiovascular disease and cancer, come complex challenges to health care systems that have limited resources, staffing, and expertise. The authors argue that AI applications on mobile devices for diagnostics and advice, those focused on clinical decision-making, and those that assess public health needs are likely to confer the greatest benefit to resource-poor health care systems, but numerous challenges must be overcome. —GKA

Science, this issue p. 955

INFECTIOUS DISEASE

Putting mice on a keto diet

Our immune responses to infections are influenced by extrinsic factors, including weather, social

Published by AAAS

interactions, and diet. Goldberg *et al*. found that feeding mice a high-fat, low-carbohydrate ketogenic diet confers protection in the context of lethal influenza infection. By characterizing the immune response in the lungs, the authors identified that a ketogenic diet promoted the expansion of γδ T cells in the lung. Experiments with mice lacking γδ T cells confirmed the functional importance of these cells in conferring protection. The γδ T cells appear to improve barrier function in the lungs by modifying differentiation and function of the airway epithelial cells. —AB

Sci. Immunol. **4**, eaav2026 (2019).

CALCIUM SIGNALING

Keeping a calcium sensor unSTIMulated

Upon endoplasmic reticulum calcium (Ca^{2+}) depletion, the Ca^{2+}-sensor STIM1 activates Ca^{2+} channels to replenish Ca^{2+} stores. Schober *et al*. investigated conformational changes in STIM1 during the early steps of activation using constitutively active STIM1 mutants associated with myopathy or cancer. Their analyses indicated that STIM1 could bind to multiple Ca^{2+} ions, a property that was disrupted by disease-associated mutations. These mutations caused STIM1 to adopt an active conformation similar to that caused by Ca^{2+} depletion in wild-type STIM1. —WW

Sci. Signal. **12**, eaax3194 (2019).

PSYCHOLOGY OF MUSIC

Cross-cultural analysis of song

It is unclear whether there are universal patterns to music across cultures. Mehr *et al*. examined ethnographic data and observed music in every society sampled (see the Perspective by Fitch and Popescu). For songs specifically, three dimensions characterize more than 25% of the performances studied: formality of the performance, arousal level, and religiosity. There is more variation in musical behavior within societies than between societies, and societies show similar levels of within-society variation in musical behavior. At the same time, one-third of societies significantly differ from average for any given dimension, and half of all societies differ from average on at least one dimension, indicating variability across cultures. —TSR

Science, this issue p. 970; see also p. 944

Published by AAAS

REVIEW SUMMARY

NANOMATERIALS

Energy storage: The future enabled by nanomaterials

Ekaterina Pomerantseva*, Francesco Bonaccorso*, Xinliang Feng*, Yi Cui*, Yury Gogotsi*

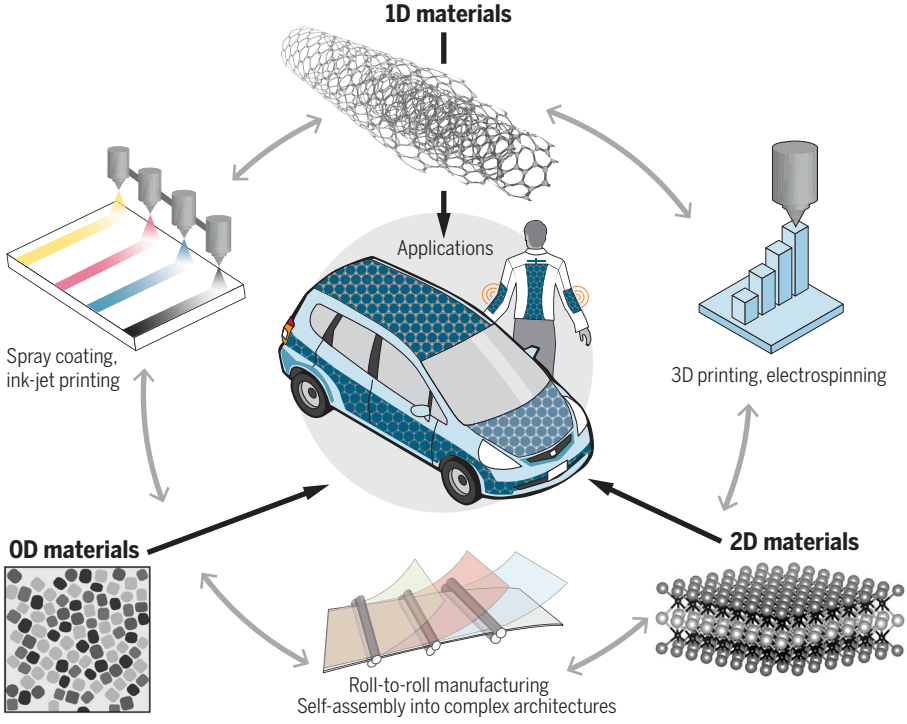
BACKGROUND: Nanomaterials offer greatly improved ionic transport and electronic conductivity compared with conventional battery and supercapacitor materials. They also enable the occupation of all intercalation sites available in the particle volume, leading to high specific capacities and fast ion diffusion. These features make nanomaterial-based electrodes able to tolerate high currents, offering a promising solution for high-energy and high-power energy storage. However, there are still many challenges associated with their use in energy storage technology and, with the exception of multiwall carbon-nanotube additives and carbon coatings on silicon particles in lithium-ion battery electrodes, the use of nanomaterials in commercial devices is very limited. After decades of development, a library of nanomaterials with versatile chemical compositions and shapes exists, ranging from oxides, chalcogenides, and carbides to carbon and elements forming alloys with lithium. This library includes various particle morphologies, such as zero-dimensional (0D) nanoparticles and quantum dots; 1D nanowires, nanotubes, and nanobelts; 2D nanoflakes and nanosheets; and 3D porous nanonetworks. Combined with lithium and beyond lithium ions, these chemically diverse nanoscale building blocks are available for creating energy storage solutions such as wearable and structural energy storage technology, which are not achievable with conventional materials.

ADVANCES: The success of nanomaterials in energy storage applications has manifold aspects. Nanostructuring is becoming key in controlling the electrochemical performance and exploiting various charge storage mechanisms, such as surface-based ion adsorption, pseudocapacitance, and diffusion-limited intercalation processes. The development of new high-performance materials, such as redox-active transition-metal carbides (MXenes) with conductivity exceeding that of carbons and other conventional electrode materials by at least an order of magnitude, open the door to the design of current collector–free and high-power next-generation energy storage devices. The combination of nanomaterials in hybrid architectures, such as carbon-silicon and carbon-sulfur, together with the development of versatile methods of nanostructuring, overcome challenges related to large volume change typical for alloying and conversion materials. These examples indicate that nanostructured materials and nanoarchitectured electrodes can provide solutions for designing and realizing high-energy, high-power, and long-lasting energy storage devices.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aan8285

OUTLOOK: The limitations of nanomaterials in energy storage devices are related to their high surface area—which causes parasitic reactions with the electrolyte, especially during the first cycle, known as the first cycle irreversibility—as well as their agglomeration. Therefore, future strategies aim to develop smart assembly of nanomaterials into architectures with controlled geometry. Moreover, combining nanomaterials with complementary functionalities, such as high electronic conductivity of graphene or MXenes with high operating voltage and high redox activity of oxides, is necessary. Building sophisticated electrode architectures requires innovative manufacturing approaches, such as printing, knitting, spray deposition, and so on. Already-developed techniques such as 3D printing, roll-to-roll manufacturing, self-assembly from solutions, atomic layer deposition, and other advanced techniques should be used to manufacture devices from nanomaterials that cannot be made by conventional slurry-based methods. Such manufacturing approaches can also enable long-sought flexible, stretchable, wearable, and structural energy storage and harvesting solutions for Internet of Things and other disruptive technologies. ■



Nanomaterials for energy storage applications. The high surface-to-volume ratio and short diffusion pathways typical of nanomaterials provide a solution for simultaneously achieving high energy and power density. Furthermore, the compatibility of nanomaterials with advanced manufacturing techniques—such as printing, spray coating, roll-to-roll assembly, and so on—allows for the design and realization of wearable, flexible, and foldable energy storage devices.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ep423@drexel.edu (E.P.); francesco.bonaccorso@iit.it (F.B.); xinliang.feng@tu-dresden.de (X.F.); yicui@stanford.edu (Y.C.); gogotsi@drexel.edu (Y.G.)
Cite this article as E. Pomerantseva *et al.*, *Science* **366**, eaan8285 (2019). DOI: 10.1126/science.aan8285



TOMORROW'S EARTH
Read more articles online at scim.ag/TomorrowsEarth

Downloaded from http://science.sciencemag.org/ on November 28, 2019

REVIEW

NANOMATERIALS

Energy storage: The future enabled by nanomaterials

Ekaterina Pomerantseva[1,2]*, Francesco Bonaccorso[3,4]*, Xinliang Feng[5,6]*, Yi Cui[7]*, Yury Gogotsi[1,2]*

Lithium-ion batteries, which power portable electronics, electric vehicles, and stationary storage, have been recognized with the 2019 Nobel Prize in chemistry. The development of nanomaterials and their related processing into electrodes and devices can improve the performance and/or development of the existing energy storage systems. We provide a perspective on recent progress in the application of nanomaterials in energy storage devices, such as supercapacitors and batteries. The versatility of nanomaterials can lead to power sources for portable, flexible, foldable, and distributable electronics; electric transportation; and grid-scale storage, as well as integration in living environments and biomedical systems. To overcome limitations of nanomaterials related to high reactivity and chemical instability caused by their high surface area, nanoparticles with different functionalities should be combined in smart architectures on nano- and microscales. The integration of nanomaterials into functional architectures and devices requires the development of advanced manufacturing approaches. We discuss successful strategies and outline a roadmap for the exploitation of nanomaterials for enabling future energy storage applications, such as powering distributed sensor networks and flexible and wearable electronics.

Energy usage is experiencing a large and fast shift toward electricity as the main power source. Reversible storage and release of electricity is an essential technology, driven by the needs of portable consumer electronics and medical devices, electric vehicles, and electric grids, as well as the emerging Internet of Things and wearable technologies. These applications and the need to store energy harvested by triboelectric and piezoelectric generators (e.g., from muscle movements), as well as solar panels, wind power generators, heat sources, and moving machinery, call for considerable improvement and diversification of energy storage technology. In this context, materials with nanometer-sized structural features and a large electrochemically active surface can change the paradigm for energy storage from within the electrode bulk to surface redox processes that occur orders of magnitude faster and allow a greatly improved power and cycle life (*1*–*3*). High electronic and ionic conductivities combined with intrinsic strength and flexibility of low-dimensional materials allow ultrathin, flexible, and structural energy storage solutions. The short diffusion path can enable the use of nonflammable solid electrolytes, leading to safer batteries, and large or multivalent ions for more affordable grid-scale applications. In addition to active energy-storing nanomaterials, passive components can benefit from the use of nanomaterials as well. For example, ultrathin hexagonal boron nitride (*h*-BN) and metal oxide separators and graphene or two-dimensional (2D) transition-metal carbide (MXene) current collectors can decrease the size and weight of the batteries (*4*, *5*). Today, we possess a large library of nanoparticles and nanostructured materials with a variety of compositions, electrochemical properties, and morphologies ranging from 0D nanoparticles and quantum dots (*6*, *7*) to 1D nanowires, nanotubes, and nanobelts (*8*, *9*), to 2D nanoflakes and nanosheets (*10*–*13*), and to 3D porous networks (*14*–*17*) (Fig. 1).

However, some of the challenges related to the reactivity of nanomaterials are due to their high specific surface area (SSA), which leads to parasitic, and often irreversible, reactions and strong interaction with electrolyte (*1*, *2*); in addition, cost and manufacturability of nanomaterials make the battery community skeptical about their promise for practical applications. Although the number of studies of various phenomena related to the performance of nanomaterials in energy storage is increasing year by year, only a few of them—such as graphene sheets, carbon nanotubes (CNTs), carbon black, and silicon nanoparticles—are currently used in commercial devices, primarily as additives (*18*).

High-capacity conversion (sulfur and fluorides) and alloying (Si and Sn) materials undergo considerable structure changes and large volume expansion and contraction (*19*, *20*), which can cause mechanical and chemomechanical instability across the length scales of individual nanoparticles, electrodes, and full electrochemical cells (*21*, *22*). Coatings with nanoscale thickness obtained via atomic or molecular layer deposition may be needed to suppress parasitic interfacial reactions, including the growth of metal dendrites, and/or form an artificial solid-electrolyte interphase (SEI) layer leading to the improved stability of electrochemical cells (*23*–*27*). Achieving future advancements in this research area will require broadening the compositional chemistry of interfacial layers and developing nanotechnology approaches that would allow for pinhole-free coating of 3D architectures with varying porosity. Advances in the development of autonomous microsystems and microdevices call for smaller power sources. As a result, many of the future energy storage devices need to be just several microns or even tens of nanometers thick. Therefore, thin film electrode and electrolyte layers need to be grown or printed not only on flat chips but also on surfaces of various shapes, such as packaging materials, or created as wearable tattoos, requiring manufacturing processes which differ drastically from the current battery manufacturing practice.

This review takes a holistic approach to energy storage, considering battery materials that exhibit bulk redox reactions and supercapacitor materials that store charge owing to the surface processes together, because nanostructuring often leads to erasing boundaries between these two energy storage solutions. We explain how the variety of 0D, 1D, 2D, and 3D nanoscale materials available today can be used as building blocks to create functional energy-storing architectures and what fundamental and engineering problems need to be resolved to enable the distributed energy storage required by the technologies of the next decade.

Nanomaterials platform

Figure 1 shows the variety of available 0D to 3D nanomaterials (nanoscale building blocks). What emerges is the large family of carbon nanomaterials (Fig. 1, top row). Carbon is invaluable for energy storage owing to its properties, such as low specific weight and high abundance, coupled with the high electronic conductivity of graphitic carbons. Moreover, because of $sp/sp^2/sp^3$ hybridization, multiple carbon structures and morphologies are available. However, nanostructured carbons usually provide limited, if any, redox capacity and only after functionalization (*28*, *29*). Therefore, they are usually used as a double-layer capacitor material, or as a conducting support backbone (*28*, *29*), rather than as active material for energy storage devices.

[1]A.J. Drexel Nanomaterials Institute, Drexel University, Philadelphia, PA 19104, USA. [2]Department of Materials Science and Engineering, Drexel University, Philadelphia, PA 19104, USA. [3]Graphene Labs, Istituto Italiano di Tecnologia, 16163 Genova, Italy. [4]BeDimensional Spa, 16163 Genova, Italy. [5]Center for Advancing Electronics Dresden (CFAED), Technische Universität Dresden, 01062 Dresden, Germany. [6]Faculty of Chemistry and Food Chemistry, Technische Universität Dresden, 01062 Dresden, Germany. [7]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: ep423@drexel.edu (E.P.); francesco.bonaccorso@iit.it (F.B.); xinliang.feng@tu-dresden.de (X.F.); yicui@stanford.edu (Y.C.); gogotsi@drexel.edu (Y.G.)

Downloaded from http://science.sciencemag.org/ on November 28, 2019

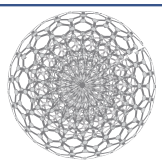
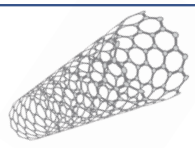
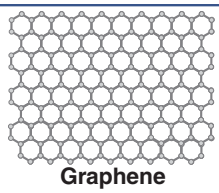
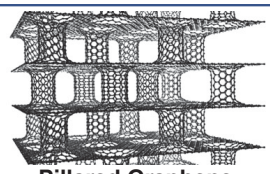
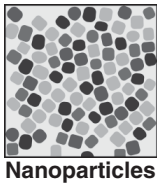
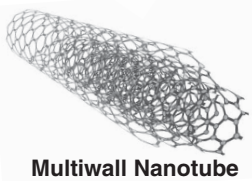
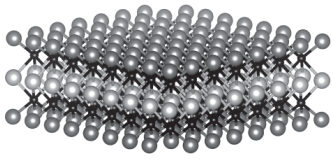
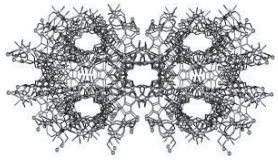
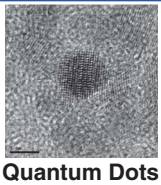
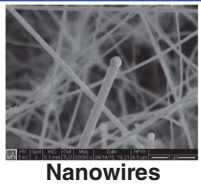
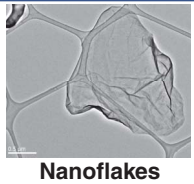
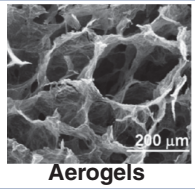
0D	1D	2D	3D
 Carbon Onion	 Single Wall Carbon Nanotube	 Graphene	 Pillared Graphene
 Nanoparticles	 Multiwall Nanotube	 Multielement 2D Compounds	 Metal-Organic Frameworks
 Quantum Dots	 Nanowires	 Nanoflakes	 Aerogels
Advantages			
Small in all dimensions Surfaces on all sites are accessible to electrolytes No bulk solid-state diffusion Can be integrated into multiple systems Can be used in stable inks for printing	Mechanical reliability Possibility to integrate with wearable devices Porous flexible freestanding films	Open 2D channels for ion transport; all surface is accessible enabling fast charge storage Compatible with flexible devices Small nanoflakes can be used in inks for printing	Can be used to create thick electrodes with large areal and volumetric storage properties
Limitations			
Agglomeration Do not densify and form only low density non-uniform structures Numerous points of contact lead to high resistance Poor chemical stability	Low packing density; cannot exhibit high volumetric performance Low yield and high cost of synthesis Diffusion pathways can be relatively long	Re-stacking Low out-of-plane electronic and ionic conductivity High cost of synthesis	Design Stability Manufacturing

Fig. 1. Overview of 0D, 1D, 2D, and 3D nanomaterials. An illustration of the chemical, structural, and morphological diversity of the available nanoscale building blocks that can be used to create complex functional architectures for next-generation energy storage devices with improved performance compared with the currently available ones. The advantages and challenges related to the application of each class of nanomaterials are summarized in the last two rows. [Metal organic frameworks, pillared graphene, quantum dots, and aerogels images reproduced with permission from (*140*–*143*), respectively; nanowires and nanoflakes images by Yayuan Liu/Stanford University and Meng-Qiang Zhao/Drexel University, respectively.]

The exception is graphite, which consists of an ordered stack of graphene layers and exhibits a specific capacity of 372 mA·hour g^{-1} for lithium ion storage in between the layers (*30*).

A more conventional approach to achieving high specific capacity is to exploit redox reactions in nanomaterials and thus utilize materials beyond carbons. High surface area of transition-metal compounds (oxides, sulfides, carbides, nitrides, etc.) and pure elements forming alloys with Li (Si, Ge, Sn, etc.) (*31*, *32*) allow those reactions to occur without solid-state diffusion limitations. More recently, highly crystalline conductive materials—such as metal organic frameworks (*33*–*35*), covalent organic frameworks (*36*), MXenes, and their composites, which form both 2D and 3D structures—have been used as electrodes for energy storage. They show promise to deliver high areal capacity owing to their high porosity, thus allowing the construction of thick electrodes (*37*). Organic nanomaterials, especially heteroatom-rich molecules and porous organic materials, not only can be directly used as electrodes for energy storage but can also be used as precursors to develop carbon-rich materials for energy storage (*38*).

In addition to chemical diversity, advances in nanomaterial synthesis enable control of morphological dimensionality, ranging from 0D to 3D nanostructures (Fig. 1), each of which has both advantages and limitations for energy storage applications. To benefit from their useful properties and alleviate shortcomings, redox-active 0D nanoparticles can be decorated on the surface of conductive 1D

Downloaded from http://science.sciencemag.org/ on November 28, 2019

and 2D materials with a double function—that is, preventing their aggregation and restacking—while at the same time contributing to charge storage. Examples of such structures are demonstrated by graphene-oxide (*6*, *39*, *40*) or MXene-oxide hybrids with enhanced energy storage capabilities enabled by the redox activity of oxide nanoparticles and the high electronic conductivity of graphene (*41*) or MXene (*42*). When produced as nanoscale particles, even typical battery materials can show pseudocapacitive behavior (pseudolinear charge-discharge), as was demonstrated for 6-nm nanoparticles of $LiCoO_2$ (*43*) and will be discussed in detail later. The self-assembly of 0D nanoparticles into denser systems is possible if the particles are monodisperse, as shown by the assembly of oxide nanoparticles on dichalcogenide sheets into ordered structures (*44*). Another limitation of 0D nanoparticle electrodes is a low electrical conductivity, owing to the contact resistance created by each contact point. An electrode containing multiple nanoparticle contact points can be considered as many resistors in series. If nanoparticles are conductive, as in the case of carbon onions, this problem is mitigated, but for most less conductive materials, this represents a severe limitation. Even in the case of conducting particles, Schottky junctions can form (*45*). In addition, nanoparticles of certain material classes, such as dichalcogenides, are not very stable (*46*) and require protective treatment during both synthesis and handling. Many semiconductor and metal nanoparticles form a passivating coating layer on the surface (*47*), creating a barrier for the charge transport, which either requires a further annealing step or encapsulation.

A key feature of 1D structures relies on the possibility of forming highly conducting continuous networks, by assembling conducting 1D nanomaterials (e.g., CNTs and metallic nanowires) over a large area and thickness of the electrode, leading to fast electron transport (*20*). CNTs are already widely used as conductive additives in battery electrodes (*20*). Additionally, many of the 1D nanostructures, especially CNTs, provide high mechanical strength (e.g., Young's modulus of 0.6 TPa) (*48*), ensuring mechanical integrity of the electrodes (*20*). Unlike the 0D particles, 1D nanomaterials do not require a binder to produce flexible structures, a feature that is particularly attractive for wearable applications. Moreover, although 1D nanomaterials form bundles (*49*), they still produce porous electrodes enabling electrolyte penetration (*20*, *35*). Porous free-standing membranes were formed using Al_2O_3 nanowires through direct transformation of bulk aluminum alloy into alkoxide nanowires, followed by heating in air (*8*). These membranes were used to produce ceramic separators for high-rate batteries, which are less prone to catching fire (*8*). The empty space between 1D structures allows the design of zero-expansion electrodes, leading to increased lifetime and decreased mechanical, thermal, and resistive losses. This property is particularly important for alloying anodes, such as Si, Ge, or Sn, in accommodating volume changes associated with interaction with up to 4.4 Li^+ ions per atom (*3*). The downside of 1D nanostructures is their low packing density, which makes the design and realization of 1D nanomaterial–based electrodes with high volumetric performance challenging (*3*). A strategy to overcome this limitation could be combining 1D nanostructures with 0D nanoparticles, thereby filling gaps between the nanotubes or nanowires and increasing the volumetric performance. The large-scale production of 1D nanostructures is another challenge. 1D nanomaterials such as single-wall CNTs or silicon nanowires are difficult to produce in large quantities at low cost. Another limitation of 1D nanomaterials is their intrinsic structure-morphology relationship, which affects the electrochemical performances. For example, for tunnel manganese oxide nanowires (*9*, *50*), the diffusion pathways are oriented along the length, not across the 1D nanowire, and can be relatively long, up to 100 μm.

For 2D nanomaterials, their primary advantage is related to their intrinsic structure, that is, the entire 2D surface can be accessible to electrolyte ions, enabling fast charge storage (*11*, *42*). Similar to 1D nanostructures, 2D sheets can form flexible electrodes without binders (*51*) or other additives, but they have the added value of achieving a much higher packing density and volumetric performance compared with the 1D counterparts. Although 2D sheets can be grown on some metal substrates, and progress is being made toward large-area single crystals (*52*), the large-scale bottom-up production of 2D materials is too expensive for the majority of energy storage applications, with the exception of small on-chip devices. By contrast, the direct liquid-phase exfoliation (LPE) of bulk layered crystals (*53*–*55*), wet chemical synthesis, and selective etching and dealloying (transformative synthesis) are much more relevant strategies for the scalable production and processing of 2D nanomaterials (*56*). The LPE process enables the formulation of inks of 2D materials in different solvents (*10*) for printing devices (*10*, *11*). 2D structures can provide high electronic conductivity in-plane, but there might be low electronic conductivity and slow transport of ions in the out-of-plane direction (*57*, *58*). Thus, the 2D morphology hinders the achievement of high (electro)chemical performance of thick planar electrodes with 2D sheets aligned along the surface of the current collector (*37*). In this respect, defect (pinhole) generation in 2D sheets would help, if it did not sacrifice the intrinsic electronic and electrochemical properties. Moreover, restacking of 2D structures limits electrolyte penetration and ionic transport (*57*). Smart architectures should have sufficient porosity and enlarged interlayer distance, as well as vertical alignment of 2D sheets enabling fast electrolyte penetration, while still enabling dense packing of nanoflakes. Hybridization of 2D sheets with 0D and 1D nanomaterials can either minimize or completely eliminate the restacking issue. Preventing restacking is also important to boost the performance of catalytically active 2D materials in order to make their surface accessible to reagents and increase the number of reaction sites. Layered materials and pillared clays offer opportunities for a tailored nanodesign, including tunable interlayer spacing and modification of the interlayer chemistry often accompanied by improved stability, leading to their diverse applications as multifunctional catalysts (*59*). Efficient strategies demonstrated in the catalysis area can also be explored for energy storage application and vice versa.

Many 3D nanomaterials, such as carbon (graphene) aerogels (*14*) and metal organic frameworks (MOFs), are a result of the assembly process of nanoparticles. Few 3D nanomaterials have intrinsic nanoscale features (e.g., thin walls of porous templated or carbide-derived carbons); most of them are built by combining 0D, 1D, and 2D nanomaterials of the same kind (carbons). To produce thick electrodes (e.g., 100 to 200 μm) with large areal and volumetric storage properties, it is necessary to develop 3D architectures optimized for both ion and electron transport. These architectures would minimize the amount of passive materials in cells, such as current collectors and separators that occupy additional volume and add dead weight. Examples of 3D electrodes with porous architectures that enable advances in energy storage have already been reported in literature (*60*–*62*). Building on these approaches, as well as developing new ones, is important for moving closer to nanomaterials-enabled energy storage.

Despite exciting diversity, none of the available nanomaterials are perfect, and none of them can solve all the problems of the current energy storage technologies. Carbon materials offer high electrical conductivity and chemical stability but a limited charge-storage capability. Transition-metal oxides and redox-active organic materials can often offer much larger charge storage, but most of them have low electrical conductivity. The latter issue can be overcome by combining the aforementioned active materials in a physical mixture or a hybrid structure with carbon or other conducting materials (*20*). 2D transition-metal carbides, nitrides, and carbonitrides, classified as MXenes, possess high electronic and high ionic conductivities (*42*). However, for this

Downloaded from http://science.sciencemag.org/ on November 28, 2019

class of materials, cost, stability, and an understanding of the electrochemical mechanisms are still open issues. This field is rapidly growing, with more than 30 highly conductive materials available, of which Ti_3C_2 has already demonstrated exceptional values of capacitance in aqueous electrolytes and excellent performance at rates up to 1000 V s^{-1} (*57*). Overall, the availability of nanoscale building blocks is continuously increasing. This provides a portfolio of materials with properties not achievable with the conventional materials used for batteries and supercapacitors.

Fundamental processes governing energy storage

The fundamental processes that control electrochemical performance can be somewhat different from bulk battery materials. For example, the transport of ions in nanoscale systems typically occurs in confinement between or at the surfaces of electrode materials (Fig. 2A), as in supercapacitors with porous carbon electrodes (*63*), instead of the transport in bulk electrolyte and solid electrodes in conventional batteries. Examples of extreme confinement could be narrow 2D slit pores between 2D sheets or narrow 1D channels in the structure of tunnel oxides or nanotubes. Although an understanding of such transport properties is largely missing thus far, there are data that show anomalous fast transport of water in hydrophobic channels of CNTs (*64*) or fast transport of protons by the Grotthuss mechanism between MXene nanosheets (*65*). The presence of confined water or electrolyte within nanochannels is known to facilitate the transport of ions (*66*), including large or multivalent ions, such as Mg^{2+} and Al^{3+} (*67*). Water or electrolyte confinement also allows the transport of complex ions—for example, $AlCl_4^-$ or organic ions—which is challenging and/or cannot be achieved in conventional battery electrodes (*68*), at least not for the intercalation mechanism. In the case of nanomaterials, we can explore fast storage mechanisms, such as intercalation pseudocapacitance, which is a process determined as non–diffusion-limited insertion of ions into the active electrode material (*69*). Confined structural water can be beneficial to improving charge-transfer resistance, especially in the case of aqueous energy storage systems (*70*, *71*). Moreover, it was shown that there was no detrimental effect on the performance of nonaqueous Li-ion cells over several hundreds of cycles when a small amount of water was added to the electrolyte (*72*). However, a considerable amount of water present in the electrode structure can be harmful to the device operation by causing parasitic reactions, which lead to irreversible chemical transformations of the device components (*73*). As of now, there is no clear understanding of how the kind of confinement or the amount of water in the material structure, the strength of bonds, crystallographic positions, and other parameters affect the electrochemical storage properties. A more in-depth study of these aspects is important to gain knowledge of the role of confined water in charge storage properties of nanomaterials.

It is also necessary to study the transport of electrons, because for some low-dimensional materials—such as CNTs, graphene, or Nb_2C—quantum capacitance (*74*) can become a limiting factor. The quantum capacitance is linked with the variation of the density of states of the materials (i.e., the effect of band filling and band emptying), which modifies the capacitance, imitating a second capacitor in series (*74*). More severely, many nanomaterials, such as oxides, are poor electronic conductors, raising multiple questions. How can we wire them? How can we inject electrons? Can electrons tunnel through thin mono-, double-, or triple-layers of ions of electrolyte between the sheets? What are the rate-limiting factors?

Hybridization of dissimilar nanomaterials—that is, the combination of oxide nanoparticles and carbons—maximizes heterointerfaces at which new phenomena can occur, as schematically illustrated in Fig. 2B. As a result of different work functions of carbon (e.g., graphene, CNTs, or carbon onions) and oxide nanoparticles (e.g., MnO_2 or Nb_2O_5), electrons may be injected from the carbon into the oxide, increasing the electrical conductivity of the latter. Consequently, formation of holes in graphene leads to an increase in the number of charge carriers and its quantum capacitance. Faster ionic transport and different ion-insertion potentials have been reported for hybrid 2D materials (*75*), such as graphene and MoS_2 (*76*) and graphene and Ti_3C_2 compared with the individual components (*77*). Therefore, the synergistic effects between different nanomaterials can be achieved by combining materials with different properties.

One challenge is the avoidance of unwanted chemical reactions associated with the high SSA of nanomaterials, which reaches to more than 2000 m^2 g^{-1} for porous carbon and graphene and is between 100 and 1000 m^2 g^{-1} for most other nanomaterials, such as CNTs, dichalcogenides, or MXenes (*42*). High SSA favors chemical interactions and the formation of SEI-like structures, often resulting in a very large first-cycle irreversible capacity or parasitic reactions during cycling, limiting the device's lifetime (Fig. 2C). Therefore, reactions between nanoparticles and electrolyte during cycling need to be studied in depth to fully understand and control them. An effective strategy to solve this problem is to assemble small nanostructures into secondary particles—that is, aggregates or clusters—of larger dimensions. In these structures, the electrolyte only contacts the outer surface of the secondary particles, thus reducing the effective contact surface area without losing the advantages of nanoscale dimensions of electrode material. This has been seen in

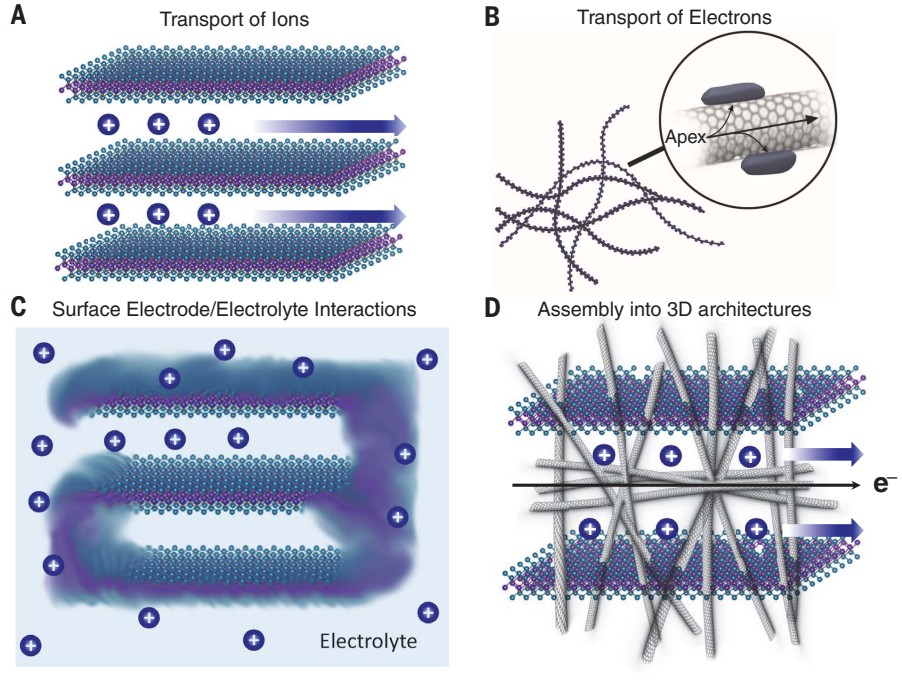


Fig. 2. Fundamental properties governing the performance of nanostructured materials for energy storage application. (**A**) Transport of ions. (**B**) Transport of electrons. (**C**) SEI formation and parasitic reactions between electrode and electrolyte. Blurry areas represent reaction products, such as SEI. (**D**) Connectivity and transport in 3D space.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

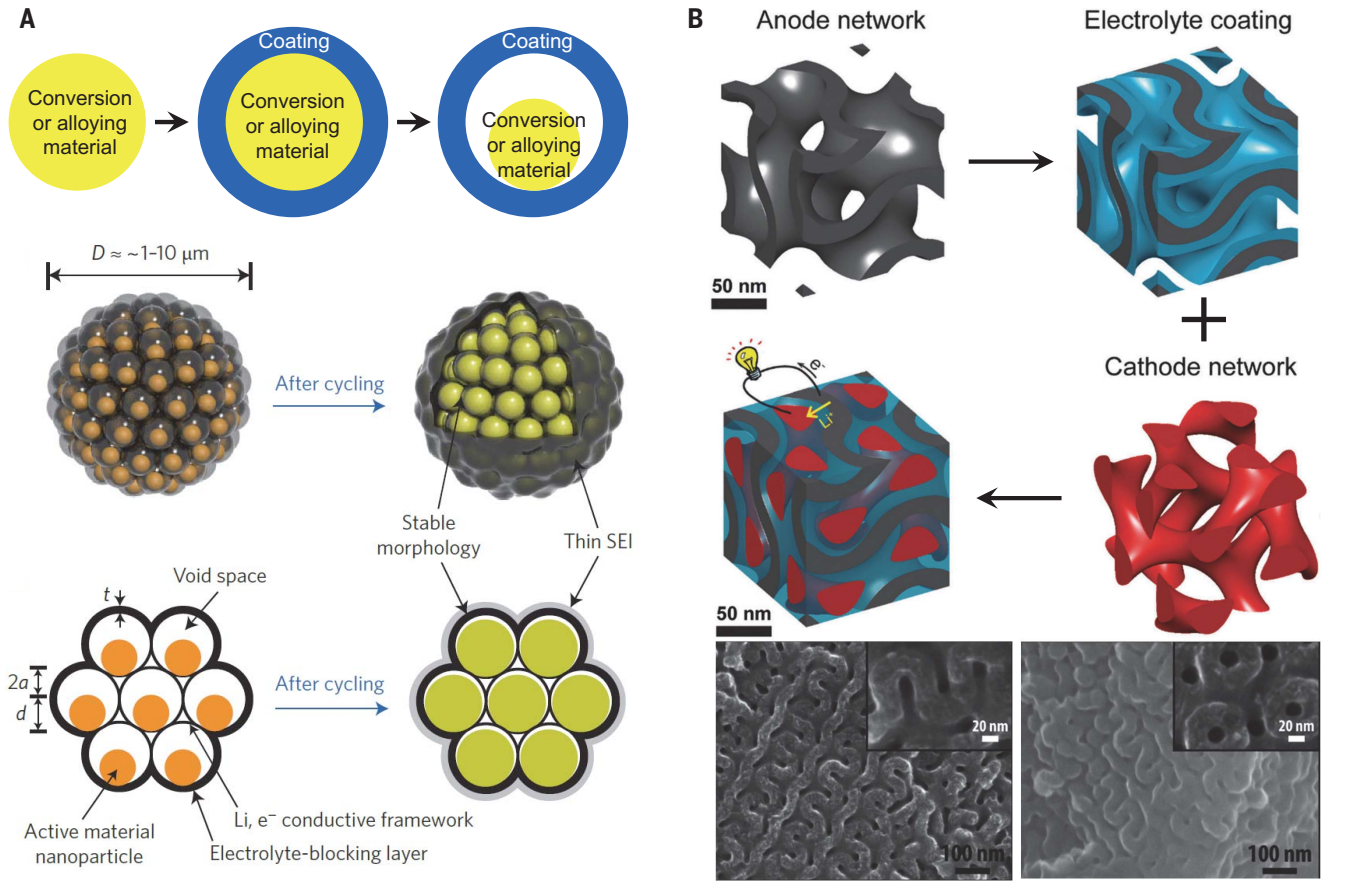


Fig. 3. Strategies developed to overcome performance limitations of nanomaterials in energy storage applications. (**A**) Nanoscale coatings on the surface of conversion and alloying electrode materials need to avoid mechanical instability caused by large-volume change and loss of the surface area as a result of agglomeration (*78*). *D*, diameter of pomegranate microparticle; *t*, thickness of the conducting framework; 2*a*, void dimension; *d*, diameter of the active material primary particles. (**B**) Nanostructured 3D electrode architecture can be realized through a scalable block-copolymer self-assembly process (*86*). [Images adapted with permission from (*78*, *86*)]

pomegranate-like Si nanoparticle–assembly anode materials (*78*) (Fig. 3A) and concentration-gradient $LiNiMnCoO_2$ cathode materials (*79*).

Finally, assembly into interconnected 3D electrode architectures with different length scales (from nanometers to microscale) can provide electronic and ionic conductivity in three dimensions. This requires pore engineering (*80*) and controllable assembly of different classes of nanomaterials (Fig. 2D). A goal is to generate a self-assembled 3D bicontinuous nanoarchitecture consisting of electrochemically active material with rapid ion and electron transport pathways (*81*). For building truly 3D architectures that integrate anode, cathode, and electrolyte together, which shortens the ionic diffusion length compared with conventional electrodes, the interdigitated types of electrode configurations have been proposed (*82*), although their experimental realization at a large scale is still challenging. A true 3D architecture was demonstrated using layer-by-layer self-assembly of interdigitated thin films on the surface of an open-cell aerogel substrate and other examples (*83*–*85*). An alternative strategy relies on the infiltration of a 3D scaffold that has a macroporous structure—for example, aerogel, graphene, or CNTs—with nanoparticles of active material. However, this approach has a number of limitations, such as limited volumetric performance. Fundamentally, it is necessary to learn how to build 3D architectures by self-assembly during electrode manufacturing. Relatively little has been done in this direction, but there are already encouraging demonstrations of truly integrated 3D electrodes (*86*) (Fig. 3B). Electrochemical or vapor-phase deposition of a separator and a counter electrode on a meso- or macroporous scaffold is promising as well.

Advances and phenomena enabled by nanomaterials in energy storage

Nanostructuring often enables the use of conventional materials that cannot be used in the microcrystalline state as either cathodes or anodes. Classical examples are alloying anodes—such as silicon, germanium, or tin—that experience large structure and volume changes during cycling (*31*, *32*). Bulk silicon, which has a theoretical capacity of up to 3579 mA·hour g^{-1}, considering $Li_{15}Si_4$ formation, cannot work as stand-alone anode in a Li-ion battery. The life cycle of silicon-based anodes is limited by the pulverization of the active material, which is determined by the volume swelling of silicon upon lithiation (up to 400 volume %) and subsequent shrinkage upon delithiation (*87*). However, reducing the particle size below ~150 nm (*3*) limits the electrode cracking upon the insertion of Li^+ ions, which mitigates the anode mechanical failure (Fig. 4A). There have been designs proposed to overcome the issues of large volume expansion and mechanical failure, including the use of nanowires (*1*, *8*), nanotubes (*88*), graphene flakes (*19*), hollow spheres, and core-shell and yolk-shell structures (*89*). To build a stable SEI for nanomaterials with large volume change, the concept of nanoscale double-walled hollowed structures was demonstrated. In this structure, the outer wall confines the expansion of the inner wall toward the hollow space inside and therefore generates a static outer surface for stable SEI formation (*31*).

Downloaded from http://science.sciencemag.org/ on November 28, 2019

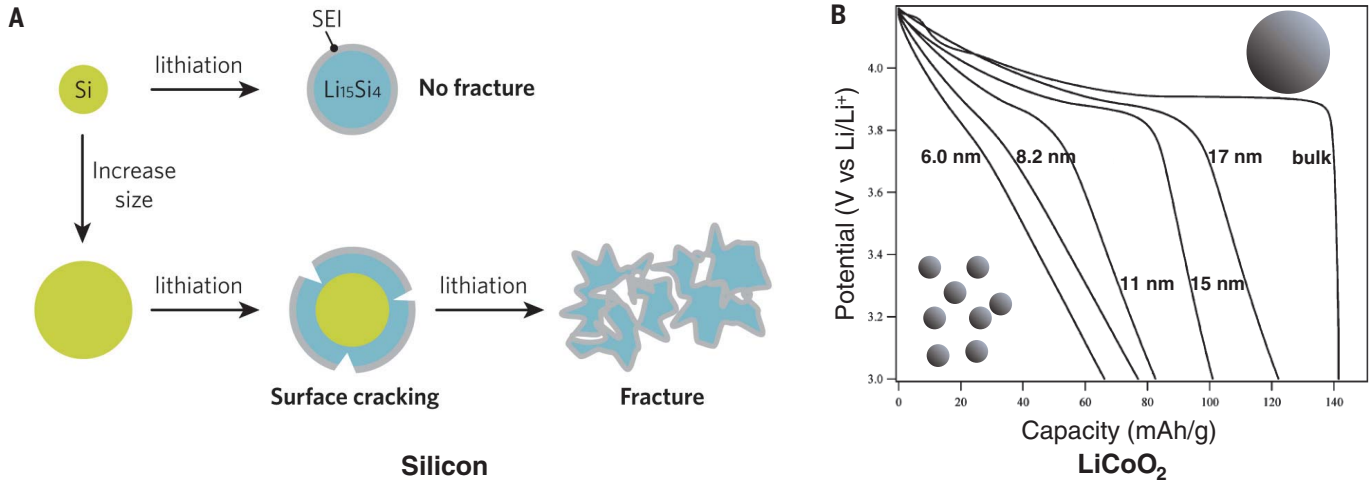


Fig. 4. Effect of nanostructuring on the electrochemical performance of energy storage materials in Li-ion batteries. (**A**) Schematic showing the effect of particle size on the lithiation of silicon anode material (*144*). (**B**) Effect of grain size on the galvanostatic discharge curve of $LiCoO_2$ cathode material (*43*). [Images adapted with permission from (*43*, *144*)]

A large family of conversion materials—such as oxides, sulfides, and fluorides—offer potential for storing a large amount of charge, but they have poor cyclability coupled with phase transformation and large volume change (*90*). Benefits of nanostructures have been fully demonstrated on these materials as well (*20*). For example, through a conversion reaction, lithiation and sodiation of FeF_2 electrodes generate a composite material consisting of ultrafine (1 to 4 nm) Fe nanoparticles, which are further fused into a continuous conductive network, and a fluoride phase. Metal nanoparticle networks in the resulting structures enable fast electron transport during further cycling (*91*, *92*). A similar effect is observed during lithiation of a $Ag_2VO_2PO_4$ cathode (*93*). High rate capability demonstrated by this material is attributed to the 15,000-fold increase in electronic conductivity associated with in situ formation of Ag nanoparticles during electrochemical reduction of $Ag_2VO_2PO_4$ (*93*). The sulfur cathode is another important example (*94*). Sulfur has a high theoretical capacity of 1673 mA·hour g^{-1} and offers a high theoretical specific energy of 2500 W·hour kg^{-1} for Li-S batteries. However, sulfur cathodes have three critical problems: (i) dissolution of lithium or sodium polysulfides into the organic electrolyte and their shuttling across the separator, (ii) lack of the electronic and ionic conductivity of sulfur and lithium sulfide resulting in poor utilization of active material, and (iii) large (80%) volume expansion of sulfur upon lithiation. The nanomaterials approach represents the most powerful solution to the aforementioned problems (*89*, *95*). Thin layers of 2D materials, such as MXene (*42*), or electrospun carbon nanofibers (*96*) at the separator on the cathode side can serve as barriers for polysulfide transfer across the separator. Encapsulation of sulfur in mesoporous carbon or MXene (*95*), S-TiO_2 yolk-shell structures (*89*), hollow sulfur spheres, and amphiphilic binding of sulfur and lithium sulfide species by oxide and nitrides offer solutions. In the case of metal-S batteries, nanomaterials with strong affinity to Li-polysulfides, such as carbides and nitrides of transition metals with high metallic conductivity (*97*), are preferred for building a scaffold for sulfur. The same materials with nanofiber or nanosheet morphology can be used for coating separators to prevent polysulfide shuttle. Another type of nanomaterial in the form of 0D or 2D particles or porous scaffolds can be used to prevent Li dendrite growth on the anode side (*98*, *99*). Such battery architecture highlights the importance of the use of nanomaterials in all the battery components but also requires a holistic approach toward selecting nanomaterials that can perform different functions within an energy storage device. This area of research is currently under active investigation, and specific material solutions are yet to be found for each individual energy storage system. Nevertheless, MXenes and graphene have already shown promise in all the three key aforementioned battery components.

The continued pursuit of high–energy density battery chemistries, such as Li-S, recently revived considerable interest in Li metal anodes. Li metal has the theoretical specific capacity of 3860 mA·hour g^{-1} and the lowest potential as an anode, which maximizes the specific energy. However, Li metal has a long list of problems, including extremely high chemical reactivity and large volume changes during Li metal plating and stripping, which create phenomenological problems of Li metal dendrites and mossy Li formation, instability of SEI, low coulombic efficiency, battery shorting, and fire hazard (*100*). The interfacial stability was recently improved with additives or coatings of nanodiamonds, *h*-BN, and other nanomaterials (*4*, *101*). The issue of large volume change can be addressed by designing host materials to house Li metal plating and striping, including hollow carbon nanospheres, graphene oxide, MXene, and polymer nanofiber scaffolds (*102*).

Many conventional cathode materials, such as $LiFePO_4$ or $LiCoO_2$, when downsized to the nanometer scale, can provide faster energy storage compared with the bulk counterparts (*43*). However, the energy storage mechanism changes, with the surface redox reaction becoming a dominant process. Large surface area creates a variety of sites for redox reactions, eliminates diffusion, and very often leads to changes in the electrochemical behavior, as has been shown, for example, for $LiCoO_2$ (*43*) (Fig. 4B). A decrease in the particle size leads to capacitor-like behavior, almost linear (supercapacitor-like) galvanostatic charge-discharge curves, and a decrease in capacity (*43*). Changes in electrochemical behavior induced by nanostructuring, similar to those observed for $LiCoO_2$, are likely to be exhibited by other intercalation cathodes, including high-capacity materials from the lithium nickel cobalt aluminum oxide (commonly $LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$ or NCA) and lithium nickel cobalt manganese oxide (often $LiNi_{0.6}Co_{0.2}Mn_{0.2}O_2$ or $LiNi_{0.8}Co_{0.1}Mn_{0.1}O_2$ or NCM) families. This behavior is not necessarily exhibited by all nanomaterials, but it is necessary to consider that capacity can be both increased or decreased, with the shrinking of crystal or particle size, depending on the type of material and charge-storage mechanism. Examples are the increase in capacity in the case of silicon (*3*) and the decrease in capacity in the case of $LiCoO_2$ (*43*) (Fig. 4B).

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Solid-state batteries, exhibiting substantially improved safety compared with the traditional ones, are considered to be one of the most promising storage technologies. In this context, there are a few considerations that make nanomaterials important for advancing this technology. First, in case of the solid-state batteries with planar geometry, nanostructuring promises to control 2D interfaces between battery components by means of the incorporation of specifically designed interface layers with nanoscale thickness and the ability to suppress parasitic reactions between electrode and electrolyte or metal dendrite growth (*23*–*27*). Additionally, nanomaterials can be used to create specific battery components. For example, nanoflakes of conductive 2D materials, such as graphene or MXene, can be assembled to form an exceptionally thin current collector layer (*103*). An exciting opportunity exists in the area of creating a highly conductive, uniform, and pinhole-free solid-state electrolyte layer with nanoscale thickness, which can potentially be achieved, for example, by using atomic layer deposition (*104*). Alternatively, Li-conducting nanofibers or nanowires could be embedded into the Li-conducting polymer matrix to produce a hybrid reinforced electrolyte layer with both high ionic conductivity and improved mechanical properties compared with the pristine one (*105*). Second, in case of the solid-state batteries with 3D architectures, the aforementioned nanomaterials-enabled advances are also important. However, in this case, nanomaterials can be used to construct 3D electrodes. Indeed, 3D elements in electrodes, such as pillars or cavities, often have micron-sized geometry to ensure mechanical stability. However, they can be hierarchically built out of nanoscale particles of various geometries to form pores for electrolyte penetration and to shorten diffusion distances to achieve fast transport of ions and electrons (*106*). Of course, building such architectures would require advanced manufacturing approaches, which we discuss below.

In addition to higher cost, compared with conventional materials, the remaining challenges related to the use of nanomaterials in energy storage devices include dealing with the sloping charge-discharge and the toxicity. For the former, the supercapacitor community knows how to handle this issue (*107*). Concerning the toxicity, it is important to ensure that nanomaterials have no or low toxicity and will not be harmful either during manufacturing or in the environment, once disposed. Nanomaterials, such as carbons, silicon, MXenes, or TiO_2, are nontoxic. Moreover, nanomaterials can be degraded in the environment faster than macroscopic materials with the same composition. For example, Ti_3C_2 degradation produces TiO_2 and CO_2, both nontoxic products, and the 2D morphology of MXene will lead to fast biodegradation when exposed to water and air. However, it is important to study the toxicity and environmental fate of new nanomaterials to ensure that they can safely be introduced into large-volume manufacturing. Minimization of the side reactions caused by a large surface area of many nanomaterials is another major challenge. Surface chemistry passivation, electrode materials design that minimizes exposed SSA (e.g., yolk-shell particles), preconditioning of electrodes, and use of electrolytes that produce stable SEI can be used separately or together to mitigate the negative effects of side reactions.

Nanomaterials with fast ion and electron transport

Low-dimensional materials can combine high electronic and ionic conductivities by using a mechanism that is usually referred to as pseudocapacitive or surface redox energy storage (*69*). It was realized more than 20 years ago by Conway *et al.* that ruthenium oxide (RuO_2), having a capacitor-like behavior, has redox energy storage (faradaic process) behind its large capacitance values (*108*). But the cost of the rare element ruthenium coupled with the fact that this material can only operate in very thin layers has limited its practical use (*109*). Additionally, the amount of structural water in ruthenium oxide is a key factor for its electrochemical performance (Fig. 5A), and its control during device operation represents a major challenge (*110*). Transition-metal atoms on the surface of MXenes can participate in redox reactions with fully electrochemically reversible redox wave in cyclic voltammetry curves overlaid on the large rectangular area corresponding to the double-layer capacitive charge storage mechanism (*57*) (Fig. 5B). The example of MXenes (*42*, *57*) has shown that both double-layer and redox capacitance can be used at very high current rates, with just ~20% electrochemical performance loss when going from 10 to 100,000 mV s^{-1} cycling (Fig. 5B). This rate would be impossible for conventional redox electrodes, which have low conductivity and a diffusion-limited charge-storage mechanism. MXenes have shown a charging time in the 1- to 10-ms range (*57*). At the same time, the capacitance of MXenes (up to 500 F g^{-1} and 1500 F cm^{-3} in thin films in acidic electrolyte, Fig. 5B) exceeds the capacitance of double-layer capacitor materials, such as carbons, which have 100 to 200 F g^{-1} or F cm^{-3}, while volumetrically rivaling that of ruthenium oxide thin films (1500 F cm^{-3}) (*110*). Thick electrodes can also work well, if restacking of the 2D sheets is prevented. In fact, vertical alignment of 2D sheets, achieved by exploiting their liquid crystalline behavior or through templating, would allow the development of MXene electrodes with tens of milligrams per square centimeter. In this context, thickness-independent (up to 200 μm) capacitance of vertically aligned MXene flakes has been demonstrated (*37*).

In many cases, however, it is necessary to combine different materials to achieve fast transport of both electrons and ions. A good example is the design and realization of hybrid structures, which have been reported for numerous oxides (Nb_2O_5, TiO_2, MoO_3, etc.) on a variety of carbon supports, such as nanotubes, graphene-based materials, and porous carbons (*13*). The carbon affects the electronic properties of both materials because it not only acts as a channel for electrons but also forms a heterojunction between the oxide and carbon surface. As a result, a higher capacity (~1000 mA·hour g^{-1}) has been achieved for a graphene–iron oxide electrode compared with both only oxide (~600 mA·hour g^{-1}), which cannot operate at high rates, and only carbon material (~400 mA·hour g^{-1}) (*111*) (Fig. 5C). When combined, these materials can operate at current densities exceeding several amperes per gram. Moreover, with the correct design of the electrode architecture, a very high rate performance can also be achieved, as demonstrated for Nb_2O_5 supported on graphene or MXene (*112*). Building such composite architectures can also allow the use of conversion electrode materials, such as FeF_3, $CuCl_2$, or S undergoing phase transformations (*21*).

Another way to enable fast transport of electrons and ions is through the creation of 2D heterostructures (*12*), which allow the combination of highly conducting and high–energy density 2D materials. Because at least one material in the hybrid structure should have good electronic conductivity, graphene has been the primary material of choice. This approach is rather universal, with a very large number of metallically conductive and redox-active materials available (*113*). It has been implemented in several different systems for applications ranging from pseudocapacitors to Li-ion and Li-S batteries (*113*). As of now, the governing assembly principles of integrating dissimilar nanomaterials into desired architectures are poorly understood. Moreover, it is not yet known how the transport of electrons and ions occurs between dissimilar 2D sheets and through the separating electrolyte or confined fluid. What is the optimal spacing between the sheets? Is the physical contact between the particles an essential requirement for electron transport or can the hopping or tunneling serve as the dominant transport mechanism? Machine learning should allow for optimization of those systems and for understanding the guiding principles for the selection of the optimum combinations of 2D materials to achieve the best electrochemical performance (*114*).

Downloaded from http://science.sciencemag.org/ on November 28, 2019

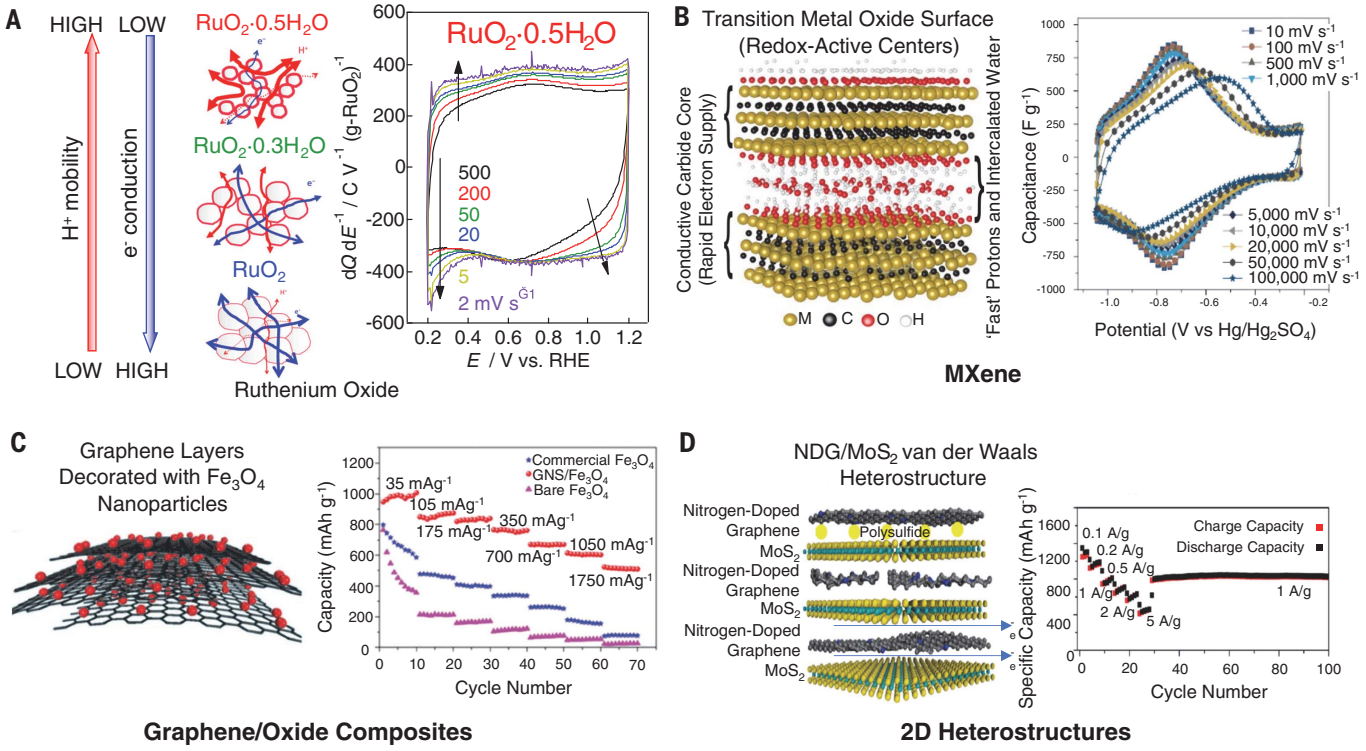


Fig. 5. Achieving fast transport of both electrons and ions in nanomaterials. (**A**) Schematic illustration of $RuO_2{\cdot}nH_2O$ demonstrating the ability to control electronic conductivity and proton transport by tuning the material hydration and cyclic voltammetry (CV) data obtained at scan rates from 2 to 500 mV s^{-1} (*145*). $dQ\ dE^{-1}$, capacitance; RHE, reversible hydrogen electrode. (**B**) Schematic illustration of a typical M_3C_2 MXene structure and cyclic voltammetry data collected at scan rates from 10 to 100,000 mV s^{-1} for a 90-nm-thick $Ti_3C_2T_x$ film. $Ti_3C_2T_x$ possesses excellent electronic conductivity owing to an electronically conductive transition metal carbide layer. Intercalated water molecules enable accessibility of protons to the redox-active TiO_2-like surface (*57*). (**C**) Schematic illustration of a composite material synthesized via decoration of graphene sheets with Fe_3O_4 nanoparticles and rate performance of the obtained material at current densities ranging from 35 to 1750 mA g^{-1} as compared with that of commercial Fe_3O_4 particles and nanoparticles of Fe_3O_4 synthesized using a similar approach but without the addition of graphene sheets (*146*). GNS, graphene nanosheets. (**D**) Schematic illustration of a 2D heterostructure consisting of the alternate MoS_2 and nitrogen-doped graphene (NDG) layers and its rate performance in a Li-S battery at current densities ranging from 0.1 to 5 A g^{-1} (*147*). [Images adapted with permission from (*57*, *145*–*147*)]

Manufacturing of the nanomaterial-enabled energy storage devices

For large-scale applications, we need to build batteries and supercapacitors in a conventional format, but exploiting nanomaterials will allow faster operation, higher power, and longer lifetime compared with the current technology. For example, replacing graphite with nanostructured silicon (*115*) can lead to a substantial increase in the energy density over conventional batteries. One of the key advantages of nanoscale materials is that they can be used to manufacture electrodes of any size, shape, or form factor. For example, we are used to seeing conventional batteries as separate units. But there is no reason why they cannot be combined with structural elements and distributed, when electrodes can conform to any shape and be made strong, robust, and flexible. As a result, instead of occupying space under the car body or in a trunk, they can become part of the automobile structure per se, for example, forming the car body or filling the empty space in doors and under the hood. All these properties can be achieved by using 1D and 2D materials with high mechanical properties and electronic conductivity. Beyond conventional energy storage devices for portable electronics and vehicles, there is increasing demand for flexible energy storage devices needed to power flexible electronics, including bendable, compressible, foldable, and stretchable devices. Wearable electronics (*116*) will require the incorporation of energy storage devices. This means that we need energy storage fibers, fabrics, and textiles and the ability to incorporate energy-storing materials into clothes. This involves the manufacture of nontoxic, strong, stretchable, and even washable conductive fibers, capable of both ionic and electronic transport. For the implantable devices, instead of traditionally used coin cells or patches, stretchable and biodegradable or biocompatible batteries could be incorporated inside the human body or battery-powered devices.

The design and realization of these devices require the incorporation of nanomaterials into functional architectures. Several strategies have been established by (i) using flexible substrates (*117*) and separators (*118*); (ii) designing new device patterns (*119*) and configurations—for example, fiber-like and spring-like devices (*120*); (iii) compositing electrode materials into flexible matrices (*32*, *121*), and so on. Flexible energy storage devices, including Li-ion battery (*122*), Na-ion battery (*7*), and Zn-air battery (*123*); flexible supercapacitors, including all-solid-state devices (*124*); and in-plane and fiber-like microsupercapacitors (*125*) have been reported. However, the packaged microdevice performance is usually inferior in terms of total volumetric or gravimetric energy density compared with conventional batteries of supercapacitors. Nanomaterials will use different manufacturing methods (Fig. 6). Spray coating and other high-throughput manufacturing techniques (e.g., doctor blade, dip coating, electrodeposition, layer-by-layer deposition, vacuum filtration, and ink-jet printing) (*10*) may make these devices sufficiently inexpensive for commodity applications. For example, electrochemically exfoliated graphene has been used to formulate

Downloaded from http://science.sciencemag.org/ on November 28, 2019

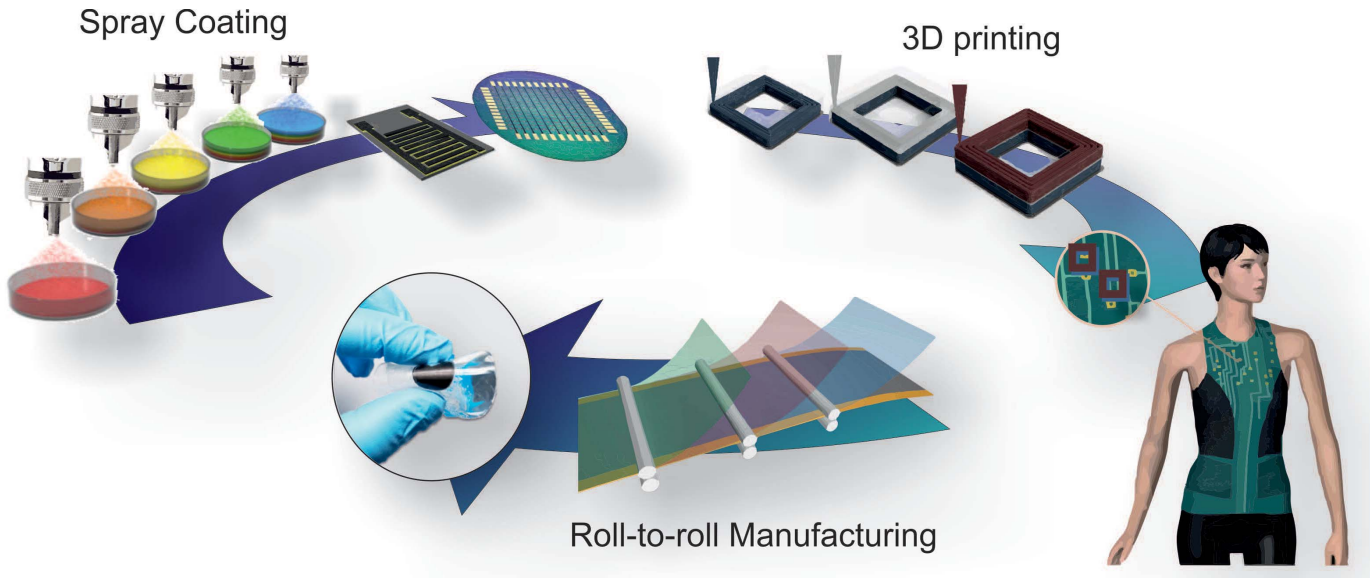


Fig. 6. Nanomaterials enable the production of next-generation energy storage systems by different manufacturing methods. [Supercapacitor array image by Husam N. Alshareef/King Abdullah University of Science and Technology (KAUST); figure wearing smart textiles image by Kristi Jost/Drexel University]

inks and manufacture in-plane micro-supercapacitors on flexible substrates and wafers by spray coating (*126*). Ink-jet printing has shown promise in the fabrication of flexible thin-film energy devices with large area and readily controllable thickness (*127*). Centrifugal casting can produce large-area sheets of 2D materials much faster than vacuum-assisted filtration. Roll-to-roll manufacturing can transform the assembly of battery-powered devices into a process similar to printing a newspaper. It is important to mention that conducting current collectors and insulating separators (in the case of sandwich-device architecture) need to be printed by the same method. Also, interdigitated and other non-sandwich energy storage device architectures become possible with printing technologies and polymer gel electrolytes, which both conduct ions and separate the electrodes.

The quality of printed films depends on the stability of active nanomaterial-based inks (*10*). In this context, functionalized nanoparticles are typically used for ink formulation because of their good dispersibility in solvents (both aqueous and organic). To increase the areal energy density of devices and their capacitance or capacity, it is necessary to build 3D devices with increased thickness and hierarchy of the electrodes. Therefore, it is necessary to develop 3D printing techniques and adjust them to specific conditions; for example, 0D particles and 2D sheets are easier to print compared with CNTs, but the latter can be more readily incorporated into fibers, providing mechanical strength and electrical conductivity required in this application. A 3D printed, interdigitated Li-ion microbattery was demonstrated using $Li_4Ti_5O_{12}$ (LTO) and $LiFePO_4$ (LFP) as the anode and cathode materials, respectively (*128*). This approach can produce distributed energy storage devices integrated with other electronic components.

The use of structural or printable energy storage (Fig. 6) adds additional requirements to the environmental and temperature stability of all components of the device. If a battery is located close to the hot part of an engine or incorporated into a part of the car body that will be subjected to sunshine over a prolonged period of time, the use of thermally stable electrolytes (e.g., ionic liquids) and other components (e.g., ceramic separators, such as ceramic nanofibers, boron nitride, or clay) may be required. In printable devices for Internet of Things and future miniaturized electronics, the use of nanomaterials should be considered not only as active components but also as binders, current collectors, sealants, and encapsulating elements. For example, the latter can be constructed using micrometer-thick polymer films with insulating 2D nanoparticles, decreasing water and oxygen permeability.

In-plane micro-supercapacitors made of carbide-derived carbons and 2D materials—including graphene, MXene, metal oxides, and conductive MOFs—are among the most popular flexible and integrated energy storage devices. Carbide-derived carbon films that possess a high SSA and narrow pore size can also be fabricated on a silicon wafer without cracking by two key steps: sputtering of metal carbides and chlorine-gas etching (*129*). 2D heterostructures can provide improved electron- and ion-transport pathways (*12*). However, most 2D materials–based in-plane micro-supercapacitors are fabricated by photolithography. 2D metal–organic coordination framework graphene- and MXene-based in-plane micro-supercapacitors with ac line-filtering performance were fabricated by in situ layer-by-layer growth of active material on prefabricated current collectors (*34*). Such an approach, coupled with the exploitation of 2D materials, offers a pathway for the scalable fabrication of in-plane micro-supercapacitors.

Smart energy storage devices, which can deliver extra functions under external stimuli beyond energy storage, enable a wide range of applications. In particular, electrochromic (*130*), photoresponsive (*131*), self-healing (*132*), thermally responsive supercapacitors and batteries have been demonstrated. However, the fade of the performance under stimuli still hinders their practical applications. Another pathway to achieve stimuli is manipulating electrolyte—for example, by using thermally responsive polymer gels—to control the ion transport between the electrodes, which can eventually cause the on and off switching of the device. Although several prototypes have already been demonstrated, considerable challenges—for example, balance of high performance and extra functions and the integration of such smart devices into fully functioning systems—still need to be resolved (*133*).

Parasitic reactions between electrode nanomaterials and electrolytes (*3*) can cause the decomposition of electrolyte and metal consumption for metal-ion batteries and consequently result in poor energy storage performance, including low Coulombic efficiency, cycle life, and energy density, compromised safety, and so on. Many efforts—such as the development of coated electrode materials (*134*), electrolytes and additives (*135*), membranes (*136*), and

Downloaded from http://science.sciencemag.org/ on November 28, 2019

metal-electrolyte interfaces (*137*)—have been made to suppress parasitic reactions. For example, fluoroethylene carbonate additive has been used to improve the cycle life of Li-ion batteries with Si nanoparticles (60 nm) by suppression of parasitic reactions, avoiding the formation of metastable c-$Li_{15}Si_4$ phase (*138*). The use of nanomaterials and new materials, in general, may require the development of new electrolytes compatible with those materials, especially in confinement (*139*).

Conclusions

Despite certain skepticism within the battery community related to the use of nanomaterials in commercial devices, several examples in which nanostructuring led to breakthroughs in performance, such as in the case of silicon-carbon composite anodes, indicate that nanostructured materials can provide solutions to create high-energy, high-power, and long-lasting energy storage devices. Research carried out over the past decade has shown that a device's lifetime increases as a result of nanostructuring. Indeed, overcoming the strain induced by cyclic expansion and contraction of macroscale particles can compensate for lifetime limitations resulting from electrolyte breakdown on a large surface of nanomaterials. Design of encapsulated nanoparticle assemblies can further minimize the contact area with electrolyte, decreasing the irreversible processes of the first charge-discharge cycle. Large-scale implementation of Si nanoparticles in Li-ion battery anodes by Sila Nanotechnologies and other companies is a convincing demonstration of the scalability of nanomaterials for large-volume battery production. The use of hundreds of tons of multiwall CNTs as conducting and reinforcing additives in battery electrodes is an excellent example of nanoscale additive use. There are other nanomaterials—such as single-wall CNTs, graphene, and so on—used in small-volume or small-size batteries and supercapacitors. Decreased prices and increased confidence in safety (health, environmental, and operational) will open doors for a wider implementation of nanomaterials in energy storage technology. To reach full potential, nanomaterials need to be combined in sophisticated architectures that enable multiple functionalities related to the transport of electrons and ions as well as interactions between various device components or suppression of such interactions. However, enabling complex architectures requires the use of advanced processing and manufacturing techniques compatible with nanomaterials, such as 3D printing, knitting, spray and/or spin coating, roll-to-roll assembly, and others.

REFERENCES AND NOTES

1. A. S. Aricò, P. Bruce, B. Scrosati, J. M. Tarascon, W. van Schalkwijk, Nanostructured materials for advanced energy conversion and storage devices. *Nat. Mater.* **4**, 366–377 (2005). doi: 10.1038/nmat1368; pmid: 15867920
2. P. G. Bruce, B. Scrosati, J. M. Tarascon, Nanomaterials for rechargeable lithium batteries. *Angew. Chem. Int. Ed.* **47**, 2930–2946 (2008). doi: 10.1002/anie.200702505; pmid: 18338357
3. Y. Sun, N. Liu, Y. Cui, Promises and challenges of nanomaterials for lithium-based rechargeable batteries. *Nat. Energy* **1**, 16071 (2016). doi: 10.1038/nenergy.2016.71
4. W. Luo *et al.*, A thermally conductive separator for stable Li metal anodes. *Nano Lett.* **15**, 6149–6154 (2015). doi: 10.1021/acs.nanolett.5b02432; pmid: 26237519
5. C. H. Wang *et al.*, Titanium carbide (MXene) as a current collector for lithium-ion batteries. *ACS Omega* **3**, 12489–12494 (2018). doi: 10.1021/acsomega.8b02032; pmid: 31457980
6. Z. S. Wu *et al.*, Graphene anchored with Co_3O_4 nanoparticles as anode of lithium ion batteries with enhanced reversible capacity and cyclic performance. *ACS Nano* **4**, 3187–3194 (2010). doi: 10.1021/nn100740x; pmid: 20455594
7. H. J. Song, D. S. Kim, J. C. Kim, S. H. Hong, D. W. Kim, An approach to flexible Na-ion batteries with exceptional rate capability and long lifespan using $Na_2FeP_2O_7$ nanoparticles on porous carbon cloth. *J. Mater. Chem. A* **5**, 5502–5510 (2017). doi: 10.1039/C7TA00727B
8. D. Lei, J. Benson, A. Magasinski, G. Berdichevsky, G. Yushin, Transformation of bulk alloys to oxide nanowires. *Science* **355**, 267–271 (2017). doi: 10.1126/science.aal2239; pmid: 28104885
9. F. Xu *et al.*, Visualization of lithium-ion transport and phase evolution within and between manganese oxide nanorods. *Nat. Commun.* **8**, 15400 (2017). doi: 10.1038/ncomms15400; pmid: 28537250
10. F. Bonaccorso, A. Bartolotta, J. N. Coleman, C. Backes, 2D-crystal-based functional inks. *Adv. Mater.* **28**, 6136–6166 (2016). doi: 10.1002/adma.201506410; pmid: 27273554
11. F. Bonaccorso *et al.*, Graphene, related two-dimensional crystals, and hybrid systems for energy conversion and storage. *Science* **347**, 1246501 (2015). doi: 10.1126/science.1246501; pmid: 25554791
12. E. Pomerantseva, Y. Gogotsi, Two-dimensional heterostructures for energy storage. *Nat. Energy* **2**, 17089 (2017). doi: 10.1038/nenergy.2017.89
13. Q. Qu, S. Yang, X. Feng, 2D sandwich-like sheets of iron oxide grown on graphene as high energy anode material for supercapacitors. *Adv. Mater.* **23**, 5574–5580 (2011). doi: 10.1002/adma.201103042; pmid: 22052661
14. Z. S. Wu *et al.*, Three-dimensional graphene-based macro- and mesoporous frameworks for high-performance electrochemical capacitive energy storage. *J. Am. Chem. Soc.* **134**, 19532–19535 (2012). doi: 10.1021/ja308676h; pmid: 23148416
15. Z. S. Wu *et al.*, Three-dimensional nitrogen and boron co-doped graphene for high-performance all-solid-state supercapacitors. *Adv. Mater.* **24**, 5130–5135 (2012). doi: 10.1002/adma.201201948; pmid: 22807002
16. D. Wu, F. Zhang, H. Liang, X. Feng, Nanocomposites and macroscopic materials: Assembly of chemically modified graphene sheets. *Chem. Soc. Rev.* **41**, 6160–6177 (2012). doi: 10.1039/c2cs35179j; pmid: 22875044
17. S. Han, D. Wu, S. Li, F. Zhang, X. Feng, Porous graphene materials for advanced electrochemical energy storage and conversion devices. *Adv. Mater.* **26**, 849–864 (2014). doi: 10.1002/adma.201303115; pmid: 24347321
18. K. Turcheniuk, D. Bondarev, V. Singhal, G. Yushin, Ten years left to redesign lithium-ion batteries. *Nature* **559**, 467–470 (2018). doi: 10.1038/d41586-018-05752-3; pmid: 30046087
19. E. Greco *et al.*, Few-layer graphene improves silicon performance in Li-ion battery anodes. *J. Mater. Chem. A* **5**, 19306–19315 (2017). doi: 10.1039/C7TA05395A
20. H. Y. Sun *et al.*, Carbon nanotubes-bridged molybdenum trioxide nanosheets as high performance anode for lithium ion batteries. *2D Materials* **5**, 015024 (2018). doi: 10.1088/2053-1583/aa9963
21. S.-H. Yu, X. Feng, N. Zhang, J. Seok, H. D. Abruña, Understanding conversion-type electrodes for lithium rechargeable batteries. *Acc. Chem. Res.* **51**, 273–281 (2018). doi: 10.1021/acs.accounts.7b00487; pmid: 29373023
22. R. Koerver *et al.*, Chemo-mechanical expansion of lithium electrode materials—on the route to mechanically optimized all-solid-state batteries. *Energy Environ. Sci.* **11**, 2142–2158 (2018). doi: 10.1039/C8EE00907D
23. J. W. Jung *et al.*, Rational design of protective In_2O_3 layer-coated carbon nanopaper membrane: Toward stable cathode for long-cycle Li-O_2 batteries. *Nano Energy* **46**, 193–202 (2018). doi: 10.1016/j.nanoen.2018.01.045
24. Y. S. Jung *et al.*, Unexpected improved performance of ALD coated $LiCoO_2$/graphite Li-ion batteries. *Adv. Energy Mater.* **3**, 213–219 (2013). doi: 10.1002/aenm.201200370
25. S. Neudeck *et al.*, Effect of low-temperature Al_2O_3 ALD coating on Ni-rich layered oxide composite cathode on the long-term cycling performance of lithium-ion batteries. *Sci. Rep.* **9**, 5328 (2019). doi: 10.1038/s41598-019-41767-0; pmid: 30926918
26. A. P. Wang, S. Kadam, H. Li, S. Q. Shi, Y. Qi, Review on modeling of the anode solid electrolyte interphase (SEI) for lithium-ion batteries. *npj Comput. Mater.* **4**, 15 (2018). doi: 10.1038/s41524-018-0064-0
27. L. Fan, H. L. L. Zhuang, L. N. Gao, Y. Y. Lu, L. A. Archer, Regulating Li deposition at artificial solid electrolyte interphases. *J. Mater. Chem. A Mater. Energy Sustain.* **5**, 3483–3492 (2017). doi: 10.1039/C6TA10204B
28. P. Simon, Y. Gogotsi, Materials for electrochemical capacitors. *Nat. Mater.* **7**, 845–854 (2008). doi: 10.1038/nmat2297; pmid: 18956000
29. G. Wang, L. Zhang, J. Zhang, A review of electrode materials for electrochemical supercapacitors. *Chem. Soc. Rev.* **41**, 797–828 (2012). doi: 10.1039/C1CS15060J; pmid: 21779609
30. B. Scrosati, J. Garche, Lithium batteries: Status, prospects and future. *J. Power Sources* **195**, 2419–2430 (2010). doi: 10.1016/j.jpowsour.2009.11.048
31. H. Wu *et al.*, Stable cycling of double-walled silicon nanotube battery anodes through solid-electrolyte interphase control. *Nat. Nanotechnol.* **7**, 310–315 (2012). doi: 10.1038/nnano.2012.35; pmid: 22447161
32. R. Mo, D. Rooney, K. Sun, H. Y. Yang, 3D nitrogen-doped graphene foam with encapsulated germanium/nitrogen-doped graphene yolk-shell nanoarchitecture for high-performance flexible Li-ion battery. *Nat. Commun.* **8**, 13949 (2017). doi: 10.1038/ncomms13949; pmid: 28051065
33. D. Sheberla *et al.*, Conductive MOF electrodes for stable supercapacitors with high areal capacitance. *Nat. Mater.* **16**, 220–224 (2017). doi: 10.1038/nmat4766; pmid: 27723738
34. C. Yang *et al.*, Coordination polymer framework based on-chip micro-supercapacitors with AC line-filtering performance. *Angew. Chem. Int. Ed.* **56**, 3920–3924 (2017). doi: 10.1002/anie.201700679; pmid: 28267257
35. Y. Mao *et al.*, Foldable interpenetrated metal-organic frameworks/carbon nanotubes thin film for lithium-sulfur batteries. *Nat. Commun.* **8**, 14628 (2017). doi: 10.1038/ncomms14628; pmid: 28262801
36. C. R. DeBlase, K. E. Silberstein, T. T. Truong, H. D. Abruña, W. R. Dichtel, β-Ketoenamine-linked covalent organic frameworks capable of pseudocapacitive energy storage. *J. Am. Chem. Soc.* **135**, 16821–16824 (2013). doi: 10.1021/ja409421d; pmid: 24147596
37. Y. Xia *et al.*, Thickness-independent capacitance of vertically aligned liquid-crystalline MXenes. *Nature* **557**, 409–412 (2018). doi: 10.1038/s41586-018-0109-z; pmid: 29769673
38. L. Borchardt *et al.*, Toward a molecular design of porous carbon materials. *Mater. Today* **20**, 592–610 (2017). doi: 10.1016/j.mattod.2017.06.002
39. Z. J. Fan *et al.*, Asymmetric supercapacitors based on graphene/MnO_2 and activated carbon nanofiber electrodes with high power and energy density. *Adv. Funct. Mater.* **21**, 2366–2375 (2011). doi: 10.1002/adfm.201100058
40. D. Wang *et al.*, Self-assembled TiO_2-graphene hybrid nanostructures for enhanced Li-ion insertion. *ACS Nano* **3**, 907–914 (2009). doi: 10.1021/nn900150y; pmid: 19323486
41. A. K. Geim, K. S. Novoselov, The rise of graphene. *Nat. Mater.* **6**, 183–191 (2007). doi: 10.1038/nmat1849 pmid: 17330084
42. B. Anasori, M. R. Lukatskaya, Y. Gogotsi, 2D metal carbides and nitrides (MXenes) for energy storage. *Nat. Rev. Mater.* **2**, 16098 (2017). doi: 10.1038/natrevmats.2016.98
43. M. Okubo *et al.*, Nanosize effect on high-rate Li-ion intercalation in $LiCoO_2$ electrode. *J. Am. Chem. Soc.* **129**, 7444–7452 (2007). doi: 10.1021/ja0681927; pmid: 17511453
44. E. V. Shevchenko, D. V. Talapin, N. A. Kotov, S. O'Brien, C. B. Murray, Structural diversity in binary nanoparticle superlattices. *Nature* **439**, 55–59 (2006). doi: 10.1038/nature04414; pmid: 16397494
45. Y. Song *et al.*, Role of interfacial oxide in high-efficiency graphene-silicon Schottky barrier solar cells. *Nano Lett.* **15**, 2104–2110 (2015). doi: 10.1021/nl505011f; pmid: 25685934

Downloaded from http://science.sciencemag.org/ on November 28, 2019

46. J. Gao *et al.*, Aging of transition metal dichalcogenide monolayers. *ACS Nano* **10**, 2628–2635 (2016). doi: 10.1021/acsnano.5b07677; pmid: 26808328
47. C. R. Ryder, J. D. Wood, S. A. Wells, M. C. Hersam, Chemically tailoring semiconducting two-dimensional transition metal dichalcogenides and black phosphorus. *ACS Nano* **10**, 3900–3917 (2016). doi: 10.1021/acsnano.6b01091; pmid: 27018800
48. G. H. Gao, T. Cagin, W. A. Goddard III, Energetics, structure, mechanical and vibrational properties of single-walled carbon nanotubes. *Nanotechnology* **9**, 184–191 (1998). doi: 10.1088/0957-4484/9/3/007
49. F. Bonaccorso *et al.*, Density gradient ultracentrifugation of nanotubes: Interplay of bundling and surfactants encapsulation. *J. Phys. Chem. C* **114**, 17267–17285 (2010). doi: 10.1021/jp1030174
50. Y. Yuan *et al.*, The influence of large cations on the electrochemical properties of tunnel-structured metal oxides. *Nat. Commun.* **7**, 13374 (2016). doi: 10.1038/ncomms13374; pmid: 27869120
51. J. Hassoun *et al.*, An advanced lithium-ion battery based on a graphene anode and a lithium iron phosphate cathode. *Nano Lett.* **14**, 4901–4906 (2014). doi: 10.1021/nl502429m; pmid: 25026051
52. Y. Hao *et al.*, The role of surface oxygen in the growth of large single-crystal graphene on copper. *Science* **342**, 720–723 (2013). doi: 10.1126/science.1243879; pmid: 24158906
53. J. S. Bunch, Y. Yaish, M. Brink, K. Bolotin, P. L. McEuen, Coulomb oscillations and Hall effect in quasi-2D graphite quantum dots. *Nano Lett.* **5**, 287–290 (2005). doi: 10.1021/nl048111+; pmid: 15794612
54. Y. Hernandez *et al.*, High-yield production of graphene by liquid-phase exfoliation of graphite. *Nat. Nanotechnol.* **3**, 563–568 (2008). doi: 10.1038/nnano.2008.215; pmid: 18772919
55. J. N. Coleman *et al.*, Two-dimensional nanosheets produced by liquid exfoliation of layered materials. *Science* **331**, 568–571 (2011). doi: 10.1126/science.1194975; pmid: 21292974
56. A. E. Del Rio Castillo *et al.*, High-yield production of 2D crystals by wet-jet milling. *Mater. Horiz.* **5**, 890–904 (2018). doi: 10.1039/C8MH00487K
57. M. R. Lukatskaya *et al.*, Ultra-high-rate pseudocapacitive energy storage in two-dimensional transition metal carbides. *Nat. Energy* **2**, 17105 (2017). doi: 10.1038/nenergy.2017.105
58. J. Qiao, X. Kong, Z.-X. Hu, F. Yang, W. Ji, High-mobility transport anisotropy and linear dichroism in few-layer black phosphorus. *Nat. Commun.* **5**, 4475 (2014). doi: 10.1038/ncomms5475; pmid: 25042376
59. G. Centi, S. Perathoner, Catalysis by layered materials: A review. *Microporous Mesoporous Mater.* **107**, 3–15 (2008). doi: 10.1016/j.micromeso.2007.03.011
60. A. Vu, Y. Q. Qian, A. Stein, Porous electrode materials for lithium-ion batteries—how to prepare them and what makes them special. *Adv. Energy Mater.* **2**, 1056–1085 (2012). doi: 10.1002/aenm.201200320
61. L. M. Zhou *et al.*, Recent developments on and prospects for electrode materials with hierarchical structures for lithium-ion batteries. *Adv. Energy Mater.* **8**, 1701415 (2018). doi: 10.1002/aenm.201701415
62. E. Uchaker, G. Z. Cao, Mesocrystals as electrode materials for lithium-ion batteries. *Nano Today* **9**, 499–524 (2014). doi: 10.1016/j.nantod.2014.06.004
63. D.-W. Wang, F. Li, M. Liu, G. Q. Lu, H.-M. Cheng, 3D aperiodic hierarchical porous graphitic carbon material for high-rate electrochemical capacitive energy storage. *Angew. Chem. Int. Ed.* **47**, 373–376 (2008). doi: 10.1002/anie.200702721; pmid: 18022983
64. M. Melillo, F. Zhu, M. A. Snyder, J. Mittal, Water transport through nanotubes with varying interaction strength between tube wall and water. *J. Phys. Chem. Lett.* **2**, 2978–2983 (2011). doi: 10.1021/jz2012319; pmid: 25606067
65. T. Miyake, M. Rolandi, Grotthuss mechanisms: From proton transport in proton wires to bioprotonic devices. *J. Phys. Condens. Matter* **28**, 023001 (2016). doi: 10.1088/0953-8984/28/2/023001; pmid: 26657711
66. N. Li, M. D. Guiver, Ion transport by nanochannels in ion-containing aromatic copolymers. *Macromolecules* **47**, 2175–2198 (2014). doi: 10.1021/ma402254h
67. M.-Q. Zhao *et al.*, Magnesium-ion storage capability of MXenes. *ACS Appl. Energy Mater.* **2**, 1572–1578 (2019). doi: 10.1021/acsaem.8b02253
68. A. VahidMohammadi, A. Hadjikhani, S. Shahbazmohamadi, M. Beidaghi, Two-dimensional vanadium carbide (MXene) as a high-capacity cathode material for rechargeable aluminum batteries. *ACS Nano* **11**, 11135–11144 (2017). doi: 10.1021/acsnano.7b05350; pmid: 29039915
69. P. Simon, Y. Gogotsi, B. Dunn, Materials science. Where do batteries end and supercapacitors begin? *Science* **343**, 1210–1211 (2014). doi: 10.1126/science.1249625; pmid: 24626920
70. D. Kundu, B. D. Adams, V. Duffort, S. H. Vajargah, L. F. Nazar, A high-capacity and long-life aqueous rechargeable zinc battery using a metal oxide intercalation cathode. *Nat. Energy* **1**, 16119 (2016). doi: 10.1038/nenergy.2016.119
71. V. Augustyn, Y. Gogotsi, 2D materials with nanoconfined fluids for electrochemical energy storage. *Joule* **1**, 443–452 (2017). doi: 10.1016/j.joule.2017.09.008
72. J. C. Burns *et al.*, The impact of intentionally added water to the electrolyte of Li-ion cells. *J. Electrochem. Soc.* **160**, A2281–A2287 (2013). doi: 10.1149/2.101311jes
73. L. W. Wangoh *et al.*, Correlating lithium hydroxyl accumulation with capacity retention in V_2O_5 aerogel cathodes. *ACS Appl. Mater. Interfaces* **8**, 11532–11538 (2016). doi: 10.1021/acsami.6b02759; pmid: 27104947
74. S. Ilani, L. A. K. Donev, M. Kindermann, P. L. McEuen, Measurement of the quantum capacitance of interacting electrons in carbon nanotubes. *Nat. Phys.* **2**, 687–691 (2006). doi: 10.1038/nphys412
75. I. Demiroglu, F. M. Peeters, O. Gülseren, D. Çakır, C. Sevik, Alkali metal intercalation in MXene/graphene heterostructures: A new platform for ion battery applications. *J. Phys. Chem. Lett.* **10**, 727–734 (2019). doi: 10.1021/acs.jpclett.8b03056; pmid: 30694678
76. D. K. Bediako *et al.*, Heterointerface effects in the electrointercalation of van der Waals heterostructures. *Nature* **558**, 425–429 (2018). doi: 10.1038/s41586-018-0205-0; pmid: 29925970
77. J. Yan *et al.*, Flexible MXene/graphene films for ultrafast supercapacitors with outstanding volumetric capacitance. *Adv. Funct. Mater.* **27**, 1701264 (2017). doi: 10.1002/adfm.201701264
78. N. Liu *et al.*, A pomegranate-inspired nanoscale design for large-volume-change lithium battery anodes. *Nat. Nanotechnol.* **9**, 187–192 (2014). doi: 10.1038/nnano.2014.6; pmid: 24531496
79. Y. K. Sun *et al.*, High-energy cathode material for long-life and safe lithium batteries. *Nat. Mater.* **8**, 320–324 (2009). doi: 10.1038/nmat2418; pmid: 19305398
80. Y. Xu *et al.*, Solution processable holey graphene oxide and its derived macrostructures for high-performance supercapacitors. *Nano Lett.* **15**, 4605–4610 (2015). doi: 10.1021/acs.nanolett.5b01212; pmid: 26056845
81. H. Zhang, X. Yu, P. V. Braun, Three-dimensional bicontinuous ultrafast-charge and -discharge bulk battery electrodes. *Nat. Nanotechnol.* **6**, 277–281 (2011). doi: 10.1038/nnano.2011.38; pmid: 21423184
82. J. W. Long, B. Dunn, D. R. Rolison, H. S. White, Three-dimensional battery architectures. *Chem. Rev.* **104**, 4463–4492 (2004). doi: 10.1021/cr020740l; pmid: 15669159
83. N. S. Ergang *et al.*, Photonic crystal structures as a basis for a three-dimensionally interpenetrating electrochemical-cell system. *Adv. Mater.* **18**, 1750–1753 (2006). doi: 10.1002/adma.200600295
84. H. S. Min *et al.*, Fabrication and properties of a carbon/polypyrrole three-dimensional microbattery. *J. Power Sources* **178**, 795–800 (2008). doi: 10.1016/j.jpowsour.2007.10.003
85. K. T. Nam *et al.*, Stamped microbattery electrodes based on self-assembled M13 viruses. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17227–17231 (2008). doi: 10.1073/pnas.0711620105; pmid: 18753629
86. J. G. Werner, G. G. Rodriguez-Calero, H. D. Abruna, U. Wiesner, Block copolymer derived 3-D interpenetrating multifunctional gyroidal nanohybrids for electrical energy storage. *Energy Environ. Sci.* **11**, 1261–1270 (2018). doi: 10.1039/C7EE03571C
87. M. N. Obrovac, V. L. Chevrier, Alloy negative electrodes for Li-ion batteries. *Chem. Rev.* **114**, 11444–11502 (2014). doi: 10.1021/cr500207g; pmid: 25399614
88. F. Luo *et al.*, Review—Nano-silicon/carbon composite anode materials towards practical application for next generation Li-ion batteries. *J. Electrochem. Soc.* **162**, A2509–A2528 (2015). doi: 10.1149/2.0131514jes
89. Z. Wei Seh *et al.*, Sulphur-TiO_2 yolk-shell nanoarchitecture with internal void space for long-cycle lithium-sulphur batteries. *Nat. Commun.* **4**, 1331 (2013). doi: 10.1038/ncomms2327; pmid: 23299881
90. N. Nitta, F. Wu, J. T. Lee, G. Yushin, Li-ion battery materials: Present and future. *Mater. Today* **18**, 252–264 (2015). doi: 10.1016/j.mattod.2014.10.040
91. K. He *et al.*, Sodiation via heterogeneous disproportionation in FeF_2 electrodes for sodium-ion batteries. *ACS Nano* **8**, 7251–7259 (2014). doi: 10.1021/nn502284y; pmid: 24911154
92. F. Wang *et al.*, Tracking lithium transport and electrochemical reactions in nanoparticles. *Nat. Commun.* **3**, 1201 (2012). doi: 10.1038/ncomms2185; pmid: 23149745
93. E. S. Takeuchi *et al.*, Electrochemical reduction of silver vanadium phosphorous oxide, $Ag_2VO_2PO_4$: The formation of electrically conductive metallic silver nanoparticles. *Chem. Mater.* **21**, 4934–4939 (2009). doi: 10.1021/cm902102k; pmid: 20161435
94. Z. W. Seh, Y. Sun, Q. Zhang, Y. Cui, Designing high-energy lithium-sulfur batteries. *Chem. Soc. Rev.* **45**, 5605–5634 (2016). doi: 10.1039/C5CS00410A; pmid: 27460222
95. X. Ji, K. T. Lee, L. F. Nazar, A highly ordered nanostructured carbon-sulphur cathode for lithium-sulphur batteries. *Nat. Mater.* **8**, 500–506 (2009). doi: 10.1038/nmat2460; pmid: 19448613
96. A. S. Levitt *et al.*, Electrospun MXene/carbon nanofibers as supercapacitor electrodes. *J. Mater. Chem. A Mater. Energy Sustain.* **7**, 269–277 (2019). doi: 10.1039/C8TA09810G
97. X. Liang, Y. Rangom, C. Y. Kwok, Q. Pang, L. F. Nazar, Interwoven MXene nanosheet/carbon-nanotube composites as Li-S cathode hosts. *Adv. Mater.* **29**, 1603040 (2017). doi: 10.1002/adma.201603040; pmid: 27859697
98. Y. M. Sun *et al.*, Graphite-encapsulated Li-metal hybrid anodes for high-capacity Li batteries. *Chem* **1**, 287–297 (2016). doi: 10.1016/j.chempr.2016.07.009
99. W. Liu *et al.*, Core-shell nanoparticle coating as an interfacial layer for dendrite-free lithium metal anodes. *ACS Cent. Sci.* **3**, 135–140 (2017). doi: 10.1021/acscentsci.6b00389; pmid: 28280780
100. D. Lin, Y. Liu, Y. Cui, Reviving the lithium metal anode for high-energy batteries. *Nat. Nanotechnol.* **12**, 194–206 (2017). doi: 10.1038/nnano.2017.16; pmid: 28265117
101. X.-B. Cheng *et al.*, Nanodiamonds suppress the growth of lithium dendrites. *Nat. Commun.* **8**, 336 (2017). doi: 10.1038/s41467-017-00519-2; pmid: 28839134
102. K. Yan *et al.*, Selective deposition and stable encapsulation of lithium through heterogeneous seeded growth. *Nat. Energy* **1**, 16010 (2016). doi: 10.1038/nenergy.2016.10
103. Y. Shi *et al.*, Graphene-based integrated electrodes for flexible lithium ion batteries. *2D Materials* **2**, 024004 (2015). doi: 10.1088/2053-1583/2/2/024004
104. A. J. Pearse *et al.*, Nanoscale solid state batteries enabled by thermal atomic layer deposition of a lithium polyphosphazene solid state electrolyte. *Chem. Mater.* **29**, 3740–3753 (2017). doi: 10.1021/acs.chemmater.7b00805
105. K. K. Fu *et al.*, Flexible, solid-state, ion-conducting membrane with 3D garnet nanofiber networks for lithium batteries. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7094–7099 (2016). doi: 10.1073/pnas.1600422113; pmid: 27307440
106. K. Gerasopoulos *et al.*, Hierarchical three-dimensional microbattery electrodes combining bottom-up self-assembly and top-down micromachining. *ACS Nano* **6**, 6422–6432 (2012). doi: 10.1021/nn301981p; pmid: 22670634
107. M. R. Lukatskaya, B. Dunn, Y. Gogotsi, Multidimensional materials and device architectures for future hybrid energy storage. *Nat. Commun.* **7**, 12647 (2016). doi: 10.1038/ncomms12647; pmid: 27600869
108. B. E. Conway, Transition from supercapacitor to battery behavior in electrochemical energy storage. *J. Electrochem. Soc.* **138**, 1539–1548 (1991). doi: 10.1149/1.2085829
109. G. Calogero, A. Bartolotta, G. Di Marco, A. Di Carlo, F. Bonaccorso, Vegetable-based dye-sensitized solar cells. *Chem. Soc. Rev.* **44**, 3244–3294 (2015). doi: 10.1039/C4CS00309H; pmid: 25855097
110. W. Sugimoto, H. Iwata, Y. Yasunaga, Y. Murakami, Y. Takasu, Preparation of ruthenic acid nanosheets and utilization of its interlayer surface for electrochemical energy storage. *Angew. Chem. Int. Ed.* **42**, 4092–4096 (2003). doi: 10.1002/anie.200351691; pmid: 12973779
111. S. Yang *et al.*, Porous iron oxide ribbons grown on graphene for high-performance lithium storage. *Sci. Rep.* **2**, 427 (2012). doi: 10.1038/srep00427; pmid: 22645643
112. C. Zhang *et al.*, Layered orthorhombic Nb_2O_5@$Nb_4C_3T_x$ and TiO_2@$Ti_3C_2T_x$ hierarchical composites for high performance Li-ion batteries. *Adv. Funct. Mater.* **26**, 4143–4151 (2016). doi: 10.1002/adfm.201600682

Downloaded from http://science.sciencemag.org/ on November 28, 2019

113. Y. Zhu *et al*., Structural engineering of 2D nanomaterials for energy storage and catalysis. *Adv. Mater.* **30**, e1706347 (2018). doi: 10.1002/adma.201706347; pmid: 29430788
114. N. C. Frey *et al*., Prediction of synthesis of 2D metal carbides and nitrides (MXenes) and their precursors with positive and unlabeled machine learning. *ACS Nano* **13**, 3031–3041 (2019). doi: 10.1021/acsnano.8b08014; pmid: 30830760
115. I. Kovalenko *et al*., A major constituent of brown algae for use in high-capacity Li-ion batteries. *Science* **334**, 75–79 (2011). doi: 10.1126/science.1209150; pmid: 21903777
116. A. Nathan *et al*., Flexible electronics: The next ubiquitous platform. *Proc. IEEE* **100**, 1486–1517 (2012). doi: 10.1109/JPROC.2012.2190168
117. W. Gao *et al*., Direct laser writing of micro-supercapacitors on hydrated graphite oxide films. *Nat. Nanotechnol.* **6**, 496–500 (2011). doi: 10.1038/nnano.2011.110; pmid: 21804554
118. L. Y. Yuan *et al*., Polypyrrole-coated paper for flexible solid-state energy storage. *Energy Environ. Sci.* **6**, 470–476 (2013). doi: 10.1039/c2ee23977a
119. M. Park *et al*., Postpatterned electrodes for flexible node-type lithium-ion batteries. *Adv. Mater.* **29**, 1605773 (2017). doi: 10.1002/adma.201605773; pmid: 28102602
120. Y. Zhang *et al*., Flexible and stretchable lithium-ion batteries and supercapacitors based on electrically conducting carbon nanotube fiber springs. *Angew. Chem. Int. Ed.* **53**, 14564–14568 (2014). doi: 10.1002/anie.201409366; pmid: 25358468
121. H. Sun *et al*., Three-dimensional holey-graphene/niobia composite architectures for ultrahigh-rate energy storage. *Science* **356**, 599–604 (2017). doi: 10.1126/science.aam5852; pmid: 28495745
122. X. Pu *et al*., A self-charging power unit by integration of a textile triboelectric nanogenerator and a flexible lithium-ion battery for wearable electronics. *Adv. Mater.* **27**, 2472–2478 (2015). doi: 10.1002/adma.201500311; pmid: 25736078
123. P. Tan *et al*., Flexible Zn- and Li-air batteries: Recent advances, challenges, and future perspectives. *Energy Environ. Sci.* **10**, 2056–2080 (2017). doi: 10.1039/C7EE01913K
124. Z. S. Wu, Z. Liu, K. Parvez, X. Feng, K. Müllen, Ultrathin printable graphene supercapacitors with AC line-filtering performance. *Adv. Mater.* **27**, 3669–3675 (2015). doi: 10.1002/adma.201501208; pmid: 25973974
125. N. A. Kyeremateng, T. Brousse, D. Pech, Microsupercapacitors as miniaturized energy-storage components for on-chip electronics. *Nat. Nanotechnol.* **12**, 7–15 (2017). doi: 10.1038/nnano.2016.196; pmid: 27819693
126. Z. Liu *et al*., Ultraflexible in-plane micro-supercapacitors by direct printing of solution-processable electrochemically exfoliated graphene. *Adv. Mater.* **28**, 2217–2222 (2016). doi: 10.1002/adma.201505304; pmid: 26784382
127. J. Li *et al*., Scalable fabrication and integration of graphene microsupercapacitors through full inkjet printing. *ACS Nano* **11**, 8249–8256 (2017). doi: 10.1021/acsnano.7b03354; pmid: 28682595
128. K. Sun *et al*., 3D printing of interdigitated Li-ion microbattery architectures. *Adv. Mater.* **25**, 4539–4543 (2013). doi: 10.1002/adma.201301036; pmid: 23776158
129. P. Huang *et al*., On-chip and freestanding elastic carbon films for micro-supercapacitors. *Science* **351**, 691–695 (2016). doi: 10.1126/science.aad3345; pmid: 26912855
130. P. Zhang *et al*., Stimulus-responsive micro-supercapacitors with ultrahigh energy density and reversible electrochromic window. *Adv. Mater.* **29**, 1604491 (2017). doi: 10.1002/adma.201604491; pmid: 27922733
131. Z. Liu *et al*., Photoswitchable micro-supercapacitor based on a diarylethene-graphene composite film. *J. Am. Chem. Soc.* **139**, 9443–9446 (2017). doi: 10.1021/jacs.7b04491; pmid: 28650642
132. Y. Huang *et al*., A self-healable and highly stretchable supercapacitor based on a dual crosslinked polyelectrolyte. *Nat. Commun.* **6**, 10310 (2015). doi: 10.1038/ncomms10310; pmid: 26691661
133. P. Zhang, F. Wang, M. Yu, X. Zhuang, X. Feng, Two-dimensional materials for miniaturized energy storage devices: From individual devices to smart integrated systems. *Chem. Soc. Rev.* **47**, 7426–7451 (2018). doi: 10.1039/C8CS00561C; pmid: 30206606
134. J. Y. Luo, W. J. Cui, P. He, Y. Y. Xia, Raising the cycling stability of aqueous lithium-ion batteries by eliminating oxygen in the electrolyte. *Nat. Chem.* **2**, 760–765 (2010). doi: 10.1038/nchem.763; pmid: 20729897
135. M. A. Deyab, 1-Allyl-3-methylimidazolium bis (trifluoromethylsulfonyl)imide as an effective organic additive in aluminum-air battery. *Electrochim. Acta* **244**, 178–183 (2017). doi: 10.1016/j.electacta.2017.05.116
136. S. Wei *et al*., Highly stable sodium batteries enabled by functional ionic polymer membranes. *Adv. Mater.* **29**, 1605512 (2017). doi: 10.1002/adma.201605512; pmid: 28112842
137. M. D. Tikekar, S. Choudhury, Z. Y. Tu, L. A. Archer, Design principles for electrolytes and interfaces for stable lithium-metal batteries. *Nat. Energy* **1**, 16114 (2016). doi: 10.1038/nenergy.2016.114
138. H. Gao *et al*., Parasitic reactions in nanosized silicon anodes for lithium-ion batteries. *Nano Lett.* **17**, 1512–1519 (2017). doi: 10.1021/acs.nanolett.6b04551; pmid: 28177638
139. X. Wang *et al*., Influences from solvents on charge storage in titanium carbide MXenes. *Nat. Energy* **4**, 241–248 (2019). doi: 10.1021/acs.nanolett.6b04551; pmid: 28177638
140. A. J. Calahorro *et al*., Luminescence and magnetic properties of three metal–organic frameworks based on the 5-(1*H*-tetrazol-5-yl)isophthalic acid ligand. *CrystEngComm* **15**, 7636–7639 (2013). doi: 10.1039/c3ce40869h
141. G. K. Dimitrakakis, E. Tylianakis, G. E. Froudakis, Pillared graphene: A new 3-D network nanostructure for enhanced hydrogen storage. *Nano Lett.* **8**, 3166–3170 (2008). doi: 10.1021/nl801417w; pmid: 18800853
142. L. E. Shea-Rohwer, J. E. Martin, X. Cai, D. F. Kelley, Red-emitting quantum dots for solid-state lighting. *ECS J. Solid State Sci. Technol.* **2**, R3112–R3118 (2013). doi: 10.1149/2.015302jss
143. H. Hu, Z. Zhao, Y. Gogotsi, J. Qiu, Compressible carbon nanotube–graphene hybrid aerogels with superhydrophobicity and superoleophilicity for oil sorption. *Environ. Sci. Technol. Lett.* **1**, 214–220 (2014). doi: 10.1021/ez500021w
144. X. H. Liu *et al*., Size-dependent fracture of silicon nanoparticles during lithiation. *ACS Nano* **6**, 1522–1531 (2012). doi: 10.1021/nn204476h; pmid: 22217200
145. W. Sugimoto, H. Iwata, K. Yokoshima, Y. Murakami, Y. Takasu, Proton and electron conductivity in hydrous ruthenium oxides evaluated by electrochemical impedance spectroscopy: The origin of large capacitance. *J. Phys. Chem. B* **109**, 7330–7338 (2005). doi: 10.1021/jp044252o; pmid: 16851839
146. G. Zhou *et al*., Graphene-wrapped Fe_3O_4 anode material with improved reversible capacity and cyclic stability for lithium ion batteries. *Chem. Mater.* **22**, 5306–5313 (2010). doi: 10.1021/cm101532x
147. C. Zhao *et al*., Self-assembly-induced alternately stacked single-layer MoS_2 and N-doped graphene: A novel van der Waals heterostructure for lithium-ion batteries. *ACS Appl. Mater. Interfaces* **8**, 2372–2379 (2016). doi: 10.1021/acsami.5b11492; pmid: 26745784

ACKNOWLEDGMENTS

Funding: E.P. acknowledges support from National Science Foundation award no. DMR-1752623. F.B. acknowledges funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 785219–GrapheneCore2 and thanks A. E. Del Rio Castillo and H. Sun for useful discussions. X.F. thanks EU Graphene Flagship, ERC T2DCP, Coordination Networks: Building Blocks for Functional Systems (SPP 1928, COORNET), and the Center of Advancing Electronics Dresden (cfaed). Y.G.'s research on energy storage was supported through the Fluid Interface Reactions, Structures, and Transport (FIRST) Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, and Office of Basic Energy Sciences.
Competing interests: None declared.

10.1126/science.aan8285

Downloaded from http://science.sciencemag.org/ on November 28, 2019

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY OF MUSIC

Universality and diversity in human song

Samuel A. Mehr*, Manvir Singh*, Dean Knox, Daniel M. Ketter, Daniel Pickens-Jones, S. Atwood, Christopher Lucas, Nori Jacoby, Alena A. Egner, Erin J. Hopkins, Rhea M. Howard, Joshua K. Hartshorne, Mariela V. Jennings, Jan Simson, Constance M. Bainbridge, Steven Pinker, Timothy J. O'Donnell, Max M. Krasnow, Luke Glowacki*

INTRODUCTION: Music is often assumed to be a human universal, emerging from an evolutionary adaptation specific to music and/or a by-product of adaptations for affect, language, motor control, and auditory perception. But universality has never actually been systematically demonstrated, and it is challenged by the vast diversity of music across cultures. Hypotheses of the evolutionary function of music are also untestable without comprehensive and representative data on its forms and behavioral contexts across societies.

RATIONALE: We conducted a natural history of song: a systematic analysis of the features of vocal music found worldwide. It consists of a corpus of ethnographic text on musical behavior from a representative sample of mostly small-scale societies, and a discography of audio recordings of the music itself. We then applied tools of computational social science, which minimize the influence of sampling error and other biases, to answer six questions. Does music appear universally? What kinds of behavior are associated with song, and how do they vary among societies? Are the musical features of a song indicative of its behavioral context (e.g., infant care)? Do the melodic and rhythmic patterns of songs vary systematically, like those patterns found in language? And how prevalent is tonality across musical idioms?

RESULTS: Analysis of the ethnography corpus shows that music appears in every society observed; that variation in song events is well characterized by three dimensions (formality, arousal, religiosity); that musical behavior varies more within societies than across them on these dimensions; and that music is regularly associated with behavioral contexts such as infant care, healing, dance, and love. Analysis of the discography corpus shows that identifiable acoustic features of songs (accent, tempo, pitch range, etc.) predict their primary behavioral context (love, healing, etc.); that musical forms vary along two dimensions (melodic and rhythmic complexity); that melodic and rhythmic bigrams fall into power-law distributions; and that tonality is widespread, perhaps universal.

CONCLUSION: Music is in fact universal: It exists in every society (both with and without words), varies more within than between societies, regularly supports certain types of behavior, and has acoustic features that are systematically related to the goals and responses of singers and listeners. But music is not a fixed biological response with a single prototypical adaptive function: It is produced worldwide in diverse behavioral contexts that vary in formality, arousal, and religiosity. Music does appear to be tied to specific perceptual, cognitive, and affective faculties, including language (all societies put words to their songs), motor control (people in all societies dance), auditory analysis (all musical systems have signatures of tonality), and aesthetics (their melodies and rhythms are balanced between monotony and chaos). These analyses show how applying the tools of computational social science to rich bodies of humanistic data can reveal both universal features and patterns of variability in culture, addressing long-standing debates about each. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aax0868

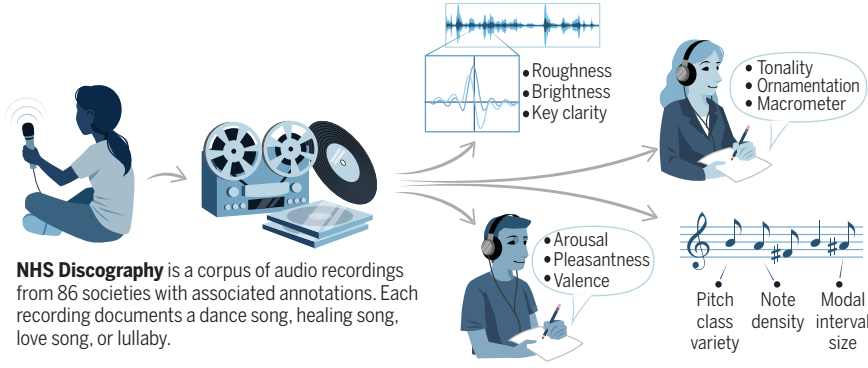
The list of author affiliations is available in the full article online.
*Corresponding author. Email: sam@wjh.harvard.edu (S.A.M.); manvirsingh@fas.harvard.edu (M.S.); glowacki@psu.edu (L.G.)
Cite this article as S. A. Mehr *et al.*, *Science* **366**, eaax0868 (2019). DOI: 10.1126/science.aax0868

Studying world music systematically. We used primary ethnographic text and field recordings of song performances to build two richly annotated cross-cultural datasets: NHS Ethnography and NHS Discography. The original material in each dataset was annotated by humans (both amateur and expert) and by automated algorithms.





NHS Ethnography is a corpus of ethnographic text from 60 societies with associated annotations. Each text excerpt describes a song performance, summarizes the use of song in a society, or both.



NHS Discography is a corpus of audio recordings from 86 societies with associated annotations. Each recording documents a dance song, healing song, love song, or lullaby.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

RESEARCH ARTICLE

PSYCHOLOGY OF MUSIC

Universality and diversity in human song

Samuel A. Mehr[1,2,3]*, Manvir Singh[4]*, Dean Knox[5], Daniel M. Ketter[6,7], Daniel Pickens-Jones[8], S. Atwood[2], Christopher Lucas[9], Nori Jacoby[10], Alena A. Egner[2], Erin J. Hopkins[2], Rhea M. Howard[2], Joshua K. Hartshorne[11], Mariela V. Jennings[11], Jan Simson[2,12], Constance M. Bainbridge[2], Steven Pinker[2], Timothy J. O'Donnell[13], Max M. Krasnow[2], Luke Glowacki[14]*

What is universal about music, and what varies? We built a corpus of ethnographic text on musical behavior from a representative sample of the world's societies, as well as a discography of audio recordings. The ethnographic corpus reveals that music (including songs with words) appears in every society observed; that music varies along three dimensions (formality, arousal, religiosity), more within societies than across them; and that music is associated with certain behavioral contexts such as infant care, healing, dance, and love. The discography—analyzed through machine summaries, amateur and expert listener ratings, and manual transcriptions—reveals that acoustic features of songs predict their primary behavioral context; that tonality is widespread, perhaps universal; that music varies in rhythmic and melodic complexity; and that elements of melodies and rhythms found worldwide follow power laws.

At least since Henry Wadsworth Longfellow declared in 1835 that "music is the universal language of mankind" (*1*), the conventional wisdom among many authors, scholars, and scientists is that music is a human universal, with profound similarities across societies (*2*). On this understanding, musicality is embedded in the biology of *Homo sapiens* (*3*), whether as one or more evolutionary adaptations for music (*4*, *5*), the by-products of adaptations for auditory perception, motor control, language, and affect (*6*–*9*), or some amalgam of these.

Music certainly is widespread (*10*–*12*), ancient (*13*), and appealing to almost everyone (*14*). Yet claims that it is universal or has universal features are commonly made without citation [e.g., (*15*–*17*)], and those with the greatest expertise on the topic are skeptical. With a few exceptions (*18*), most music scholars suggest that few if any universals exist in music (*19*–*23*). They point to variability in the interpretations of a given piece of music (*24*–*26*), the importance of natural and social environments in shaping music (*27*–*29*), the diverse forms of music that can share similar behavioral functions (*30*), and the methodological difficulty of comparing the music of different societies (*12*, *31*, *32*). Given these criticisms, along with a history of some scholars using comparative work to advance erroneous claims of cultural or racial superiority (*33*), the common view among music scholars today (*34*, *35*) is summarized by the ethnomusicologist George List: "The only universal aspect of music seems to be that most people make it. … I could provide pages of examples of the non-universality of music. This is hardly worth the trouble" (*36*).

Are there, in fact, meaningful universals in music? No one doubts that music varies across cultures, but diversity in behavior can shroud regularities emerging from common underlying psychological mechanisms. Beginning with Chomsky's hypothesis that the world's languages conform to an abstract Universal Grammar (*37*, *38*), many anthropologists, psychologists, and cognitive scientists have shown that behavioral patterns once considered arbitrary cultural products may exhibit deeper, abstract similarities across societies emerging from universal features of human nature. These include religion (*39*–*41*), mate preferences (*42*), kinship systems (*43*), social relationships (*44*, *45*), morality (*46*, *47*), violence and warfare (*48*–*50*), and political and economic beliefs (*51*, *52*).

Music may be another example, although it is perennially difficult to study. A recent analysis of the *Garland Encyclopedia of World Music* revealed that certain features—such as the use of words, chest voice, and an isochronous beat—appear in a majority of songs recorded within each of nine world regions (*53*). But the corpus was sampled opportunistically, which made generalizations to all of humanity impossible; the musical features were ambiguous, leading to poor interrater reliability; and the analysis studied only the forms of the societies' music, not the behavioral contexts in which it is performed, leaving open key questions about functions of music and their connection to its forms.

Music perception experiments have begun to address some of these issues. In one, internet users reliably discriminated dance songs, healing songs, and lullabies sampled from 86 mostly small-scale societies (*54*); in another, listeners from the Mafa of Cameroon rated "happy," "sad," and "fearful" examples of Western music somewhat similarly to Canadian listeners, despite having had limited exposure to Western music (*55*); in a third, Americans and Kreung listeners from a rural Cambodian village were asked to create music that sounded "angry," "happy," "peaceful," "sad," or "scared" and generated similar melodies to one another within these categories (*56*). These studies suggest that the form of music is systematically related to its affective and behavioral effects in similar ways across cultures. But they can only provide provisional clues about which aspects of music, if any, are universal, because the societies, genres, contexts, and judges are highly limited, and because they too contain little information about music's behavioral contexts across cultures.

A proper evaluation of claims of universality and variation requires a natural history of music: a systematic analysis of the features of musical behavior and musical forms across cultures, using scientific standards of objectivity, representativeness, quantification of variability, and controls for data integrity. We take up this challenge here. We focus on vocal music (hereafter, song) rather than instrumental music [see (*57*)] because it does not depend on technology, has well-defined physical correlates [i.e., pitched vocalizations (*19*)], and has been the primary focus of biological explanations for music (*4*, *5*).

Leveraging more than a century of research from anthropology and ethnomusicology, we built two corpora, which collectively we call the Natural History of Song (NHS). The NHS Ethnography is a corpus of descriptions of song performances, including their context, lyrics, people present, and other details, systematically assembled from the ethnographic record to representatively sample diversity across societies. The NHS Discography is a corpus of field recordings of performances of four kinds of song—dance, healing, love, and lullaby—from an approximately representative sample of human societies, mostly small-scale.

[1]Data Science Initiative, Harvard University, Cambridge, MA 02138, USA. [2]Department of Psychology, Harvard University, Cambridge, MA 02138, USA. [3]School of Psychology, Victoria University of Wellington, Wellington, New Zealand. [4]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [5]Department of Politics, Princeton University, Princeton, NJ 08544, USA. [6]Eastman School of Music, University of Rochester, Rochester, NY 14604, USA. [7]Department of Music, Missouri State University, Springfield, MO 65897, USA. [8]Unaffiliated scholar, Portland, OR 97212, USA. [9]Department of Political Science, Washington University, St. Louis, MO 63130, USA. [10]Computational Auditory Perception Group, Max Planck Institute for Empirical Aesthetics, 60322 Frankfurt am Main, Germany. [11]Department of Psychology, Boston College, Chestnut Hill, MA 02467, USA. [12]Department of Psychology, University of Konstanz, 78464 Konstanz, Germany. [13]Department of Linguistics, McGill University, Montreal, QC H3A 1A7, Canada. [14]Department of Anthropology, Pennsylvania State University, State College, PA 16802, USA.
*Corresponding author. Email: sam@wjh.harvard.edu (S.A.M.); manvirsingh@fas.harvard.edu (M.S.); glowacki@psu.edu (L.G.)

Downloaded from http://science.sciencemag.org/ on November 28, 2019

We used the corpora to test five sets of hypotheses about universality and variability in musical behavior and musical forms:

1) We tested whether music is universal by examining the ethnographies of 315 societies, and then a geographically stratified pseudorandom sample of them.

2) We assessed how the behaviors associated with song differ among societies. We reduced the high-dimensional NHS Ethnography annotations to a small number of dimensions of variation while addressing challenges in the analysis of ethnographic data, such as selective nonreporting. This allowed us to assess how the variation in musical behavior across societies compares with the variation within a single society.

3) We tested which behaviors are universally or commonly associated with song. We cataloged 20 common but untested hypotheses about these associations, such as religious activity, dance, and infant care (*4*, *5*, *40*, *54*, *58*–*60*), and tested them after adjusting for sampling error and ethnographer bias, problems that have bedeviled prior tests.

4) We analyzed the musical features of songs themselves, as documented in the NHS Discography. We derived four representations of each song, including blind human ratings and machine summaries. We then applied machine classifiers to these representations to test whether the musical features of a song predict its association with particular behavioral contexts.

5) In exploratory analyses, we assessed the prevalence of tonality in the world's songs, found that variation in their annotations falls along a small number of dimensions, and plotted the statistical distributions of melodic and rhythmic patterns in them.

All data and materials are publicly available at http://osf.io/jmv3q. We also encourage readers to view and listen to the corpora interactively via the plots available at http://themusiclab.org/nhsplots.

Music appears in all measured human societies

Is music universal? We first addressed this question by examining the eHRAF World Cultures database (*61*, *62*), developed and maintained by the Human Relations Area Files organization. It includes high-quality ethnographic documents from 315 societies, subject-indexed by paragraph. We searched for text that was tagged as including music (instrumental or vocal) or that contained at least one keyword identifying vocal music (e.g., "singers").

Music was widespread: The eHRAF ethnographies describe music in 309 of the 315 societies. Moreover, the remaining six (the Turkmen, Dominican, Hazara, Pamir, Tajik, and Ghorbat peoples) do in fact have music, according to primary ethnographic documents available outside the database (*63*–*68*). Thus, music is present in 100% of a large sample of societies, consistent with the claims of writers and scholars since Longfellow (*1*, *4*, *5*, *10*, *12*, *53*, *54*, *58*–*60*, *69*–*73*). Given these data, and assuming that the sample of human societies is representative, the Bayesian 95% posterior credible interval for the population proportion of human societies that have music, with a uniform prior, is [0.994, 1].

To examine what about music is universal and how music varies worldwide, we built the NHS Ethnography (Fig. 1 and Text S1.1), a corpus of 4709 descriptions of song performances drawn from the Probability Sample File (*74*–*76*). This is a ~45-million-word subset of the 315-society database, comprising 60 traditionally living societies that were drawn pseudorandomly from each of Murdock's 60 cultural clusters (*62*), covering 30 distinct geographical

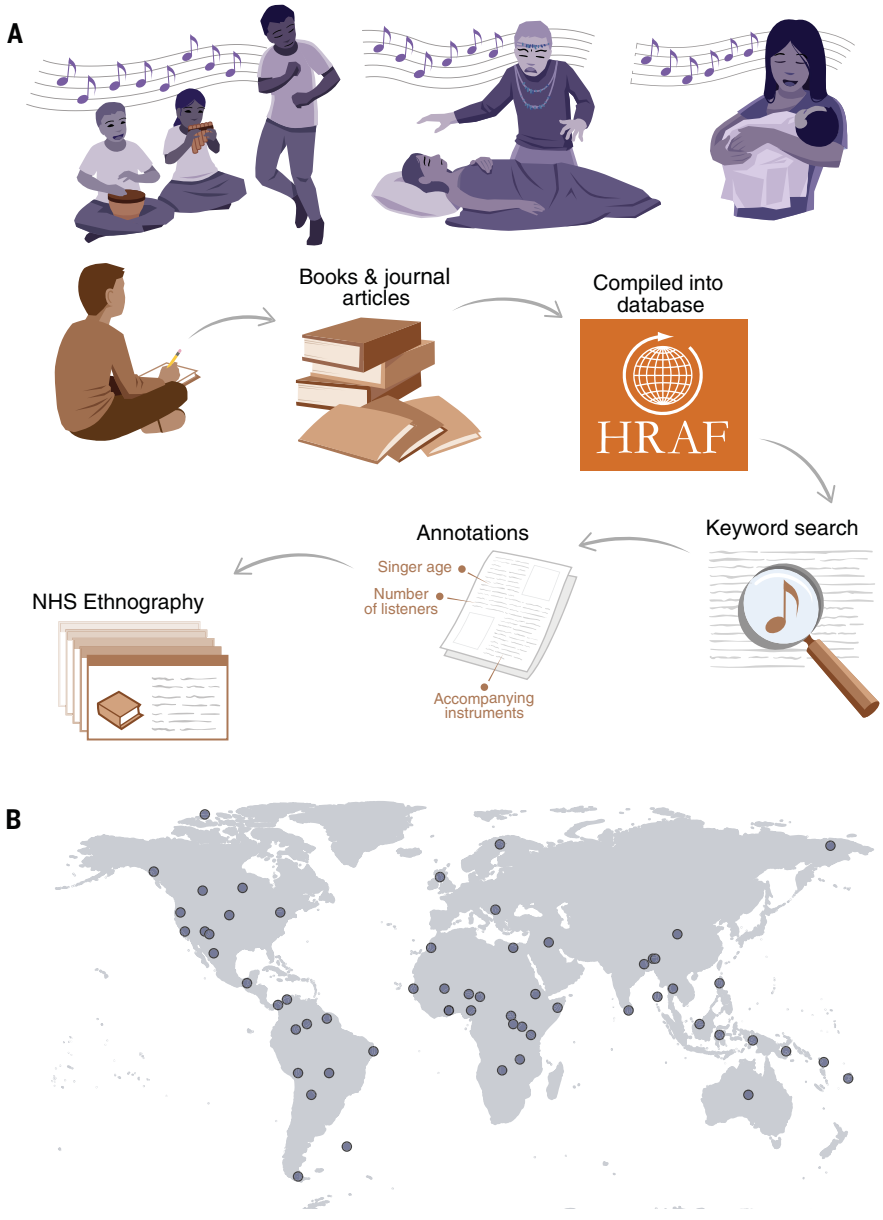


Fig. 1. Design of the NHS Ethnography. The illustration depicts the sequence from acts of singing to the ethnography corpus. (**A**) People produce songs in conjunction with other behavior, which scholars observe and describe in text. These ethnographies are published in books, reports, and journal articles and then compiled, translated, cataloged, and digitized by the Human Relations Area Files organization. (**B**) We conduct searches of the online eHRAF corpus for all descriptions of songs in the 60 societies of the Probability Sample File and annotate them with a variety of behavioral features. The raw text, annotations, and metadata together form the NHS Ethnography. Codebooks listing all available data are in tables S1 to S6; a listing of societies and locations from which texts were gathered is in table S12.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

regions and selected to be historically mostly independent of one another. Because the corpus representatively samples from the world's societies, it has been used to test cross-cultural regularities in many domains (*46*, *77*–*83*), and these regularities may be generalized (with appropriate caution) to all societies.

The NHS Ethnography, it turns out, includes examples of songs in all 60 societies. Moreover, each society has songs with words, as opposed to just humming or nonsense syllables (which are reported in 22 societies). Because the societies were sampled independently of whether their people were known to produce music, in contrast to prior cross-cultural studies (*10*, *53*, *54*), the presence of music in each one—as recognized by the anthropologists who embedded themselves in the society and wrote their authoritative ethnographies—constitutes the clearest evidence supporting the claim that song is a human universal. Readers interested in the nature of the ethnographers' reports, which bear on what constitutes "music" in each society [see (*27*)], are encouraged to consult the interactive NHS Ethnography Explorer at http://themusiclab.org/nhsplots.

Musical behavior worldwide varies along three dimensions

How do we reconcile the discovery that song is universal with the research from ethnomusicology showing radical variability? We propose that the music of a society is not a fixed inventory of cultural behaviors, but rather the product of underlying psychological faculties that make certain kinds of sound feel appropriate to certain social and emotional circumstances. These include entraining the body to acoustic and motoric rhythms, analyzing harmonically complex sounds, segregating and grouping sounds into perceptual streams (*6*, *7*), parsing the prosody of speech, responding to emotional calls, and detecting ecologically salient sounds (*8*, *9*). These faculties may interact with others that specifically evolved for music (*4*, *5*). Musical idioms differ with respect to which acoustic features they use and which emotions they engage, but they all draw from a common suite of psychological responses to sound.

If so, what should be universal about music is not specific melodies or rhythms but clusters of correlated behaviors, such as slow soothing lullabies sung by a mother to a child or lively rhythmic songs sung in public by a group of dancers. We thus asked how musical behavior varies worldwide and how the variation within societies compares to the variation between them.

Reducing the dimensionality of variation in musical behavior

To determine whether the wide variation in the annotations of the behavioral context of songs in the database (Text S1.1) falls along a smaller number of dimensions capturing the principal ways that musical behavior varies worldwide, we used an extension of Bayesian principal components analysis (*84*), which, in addition to reducing dimensionality, handles missing data in a principled way and provides a credible interval for each observation's coordinates in the resulting space. Each observation is a "song event," namely, a description in the NHS Ethnography of a song performance, a characterization of how a society uses songs, or both.

We found that three latent dimensions is the optimum number, explaining 26.6% of variability in NHS Ethnography annotations. Figure 2 depicts the space and highlights examples from excerpts in the corpus; an interactive version is available at http://themusiclab.org/nhsplots. (See Text S2.1 for details of the model, including the dimension selection procedure, model diagnostics, a test of robustness, and tests of the potential influence of ethnographer characteristics on model results.) To interpret the space, we examined annotations that load highly on each dimension; to validate this interpretation, we searched for examples at extreme locations and examined their content. Loadings are presented in tables S13 to S15; a selection of extreme examples is given in table S16.

The first dimension (accounting for 15.5% of the variance, including error noise) captures variability in the Formality of a song: Excerpts high along this dimension describe ceremonial events involving adults, large audiences, and instruments; excerpts low on it describe informal events with small audiences and children. The second dimension (accounting for 6.2%) captures variability in Arousal: Excerpts high along this dimension describe lively events with many singers, large audiences, and dancing; excerpts low on it describe calmer events involving fewer people and less overt affect, such as people singing to themselves. The third dimension (4.9%) distinguishes Religious events from secular ones: Passages high along this dimension describe shamanic ceremonies, possession, and funerary songs; passages low on it describe communal events without spiritual content, such as community celebrations.

To validate whether this dimensional space captured behaviorally relevant differences among songs, we tested whether we could reliably recover clusters for four distinctive, easily identifiable, and regularly occurring song types: dance, lullaby, healing, and love (*54*). We searched the NHS Ethnography for keywords and human annotations that matched at least one of the four types (table S17).

Although each song type can appear throughout the space, clear structure is observable (Fig. 2): The excerpts falling into each type cluster together. On average, dance songs (1089 excerpts) occupy the high-Formality, high-Arousal, low-Religiosity region. Healing songs (289 excerpts) cluster in the high-Formality, high-Arousal, high-Religiosity region. Love songs (354 excerpts) cluster in the low-Formality, low-Arousal, low-Religiosity region. Lullabies (156 excerpts) are the sparsest category (although this was likely due to high missingness in variables associated with lullabies, such as one indicating the presence of infant-directed song; see Text S2.1.5) and are located mostly in the low-Formality and low-Arousal regions. An additional 2821 excerpts matched either more than one category or none of the four.

To specify the coherence of these clusters formally rather than just visually, we asked what proportion of song events are closer to the centroid of their own type's location than to any other type (Text S2.1.6). Overall, 64.7% of the songs were located closest to the centroid of their own type; under a null hypothesis that song type is unrelated to location, simulated by randomly shuffling the song labels, only 23.2% would do so ($P < 0.001$ according to a permutation test). This result was statistically significant for three of the four song types (dance, 66.2%; healing, 74.0%; love, 63.6%; Ps < 0.001) although not for lullabies (39.7%, $P = 0.92$). The matrix showing how many songs of each type were near each centroid is in table S18. Note that these analyses eliminated variables with high missingness; a validation model that analyzed the entire corpus yielded similar dimensional structure and clustering (figs. S1 and S2 and Text S2.1.5).

The range of musical behavior is similar across societies

We next examined whether this pattern of variation applies within all societies. Do all societies take advantage of the full spectrum of possibilities made available by the neural, cognitive, and cultural systems that underlie music? Alternatively, is there only a single, prototypical song type that is found in all societies, perhaps reflecting the evolutionary origin of music (love songs, say, if music evolved as a courtship display; or lullabies, if it evolved as an adaptation to infant care), with the other types haphazardly distributed or absent altogether, depending on whether the society extended the prototype through cultural evolution? As a third alternative, do societies fall into discrete typologies, such as a Dance Culture or a Lullaby Culture? As still another alternative, do they occupy sectors of the space, so that there are societies with only arousing songs or only religious songs, or societies whose songs are equally formal and vary only by arousal, or vice versa? The data in Fig. 2, which pool song events across societies, cannot answer such questions.

We estimated the variance of each society's scores on each dimension, aggregated across all ethnographies from that society. This revealed

Downloaded from http://science.sciencemag.org/ on November 28, 2019

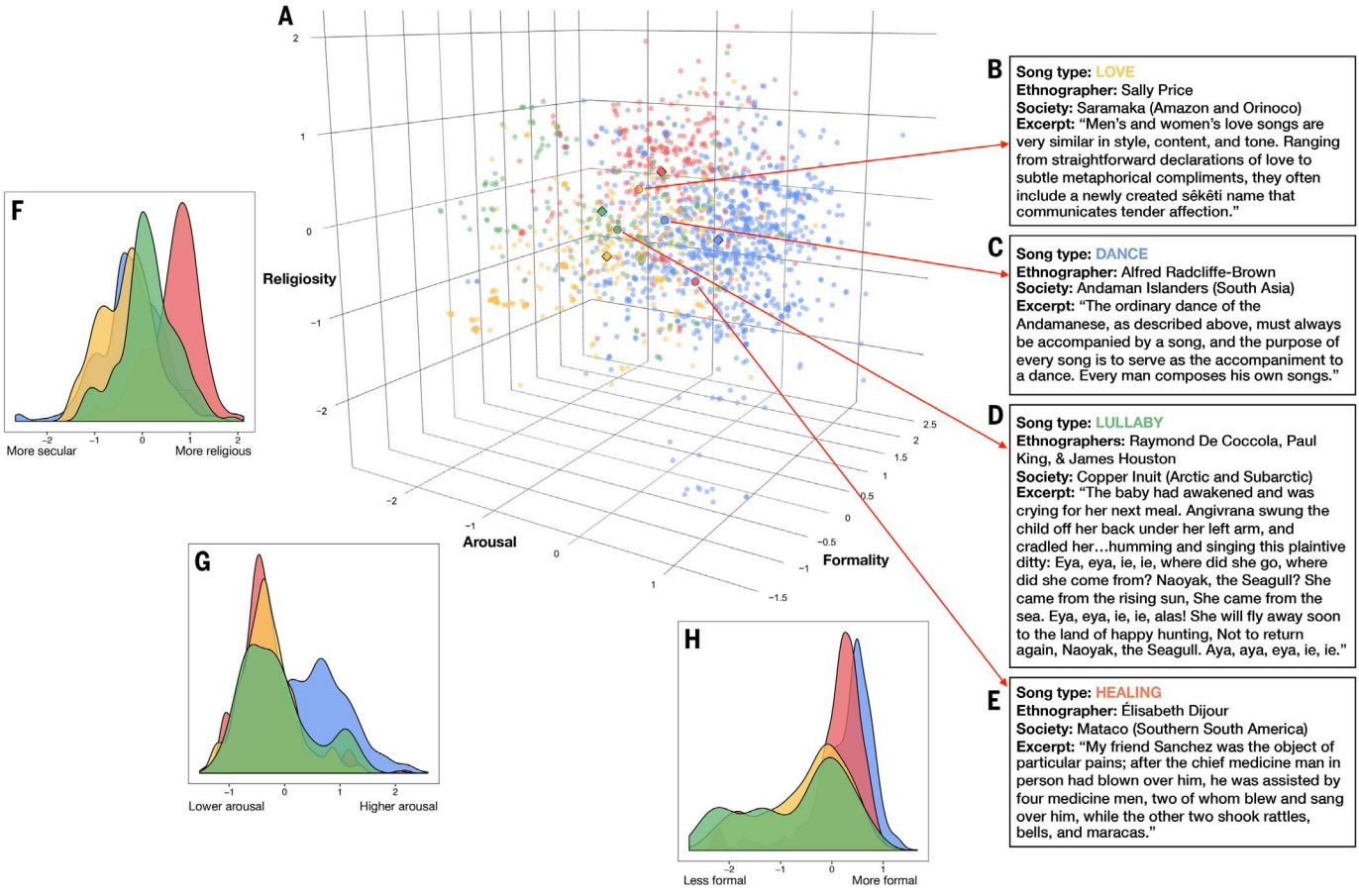


Fig. 2. Patterns of variation in the NHS Ethnography. (**A** to **E**) Projection of a subset of the NHS Ethnography onto three principal components. Each point represents the posterior mean location of an excerpt, with points colored by which of four types (identified by a broad search for matching keywords and annotations) it falls into: dance (blue), lullaby (green), healing (red), or love (yellow). The geometric centroids of each song type are represented by the diamonds. Excerpts that do not match any single search are not plotted but can be viewed in the interactive version of this figure at http://themusiclab.org/nhsplots, along with all text and metadata. Selected examples of each song type are presented here [highlighted circles and (B) to (E)]. (**F** to **H**) Density plots show the differences between song types on each dimension. Criteria for classifying song types from the raw text and annotations are shown in table S17.

that the distributions of each society's observed musical behaviors are remarkably similar (Fig. 3), such that a song with "average formality," "average arousal," or "average religiosity" could appear in any society we studied. This finding is supported by comparing the global average along each dimension to each society's mean and standard deviation, which summarizes how unusual the average song event would appear to members of that society. We found that in every society, a song event at the global mean would not appear out of place: The global mean always falls within the 95% confidence interval of every society's distribution (fig. S3). These results do not appear to be driven by any bias stemming from ethnographer characteristics such as sex or academic field (fig. S4 and Text S2.1.7), nor are they artifacts of a society being related to other societies in the sample by region, subregion, language family, subsistence type, or location in the Old versus New World (fig. S5 and Text S2.1.8).

We also applied a comparison that is common in studies of genetic diversity (*85*) and that has been performed in a recent cultural-phylogenetic study of music (*86*). It revealed that typical within-society variation is approximately six times the between-society variation. Specifically, the ratios of within- to between-society variances were 5.58 for Formality [95% Bayesian credible interval, (4.11, 6.95)]; 6.39 (4.72, 8.34) for Arousal; and 6.21 (4.47, 7.94) for Religiosity. Moreover, none of the 180 mean values for the 60 societies over the three dimensions deviated from the global mean by more than 1.96 times the standard deviation of the principal components scores within that society (fig. S3 and Text S2.1.9).

These findings demonstrate global regularities in musical behavior, but they also reveal that behaviors vary quantitatively across societies, consistent with the long-standing conclusions of ethnomusicologists. For instance, the Kanuri's musical behaviors are estimated to be less formal than those of any other society, whereas those of the Akan are estimated to be the most religious (in both cases, significantly different from the global mean on average). Some ethnomusicologists have attempted to explain such diversity, noting, for example, that more formal song performances tend to be found in more socially rigid societies (*10*).

Despite this variation, a song event of average formality would appear unremarkable in the Kanuri's distribution of songs, as would a song event of average religiosity in the Akan. Overall, we find that for each dimension, approximately one-third of all societies' means significantly differed from the global mean, and approximately half differed from the global mean on at least one dimension (Fig. 3). But despite variability in the societies' means on each dimension, their distributions overlap substantially with one another and with the global mean. Moreover, even the outliers in Fig. 3 appear to represent not genuine idiosyncrasy in some cultures but sampling error: The societies that differ more from the global mean on some dimension are those with sparser documentation in the ethnographic record (fig. S6 and Text S2.1.10). To ensure that these results are not artifacts of the

Downloaded from http://science.sciencemag.org/ on November 28, 2019

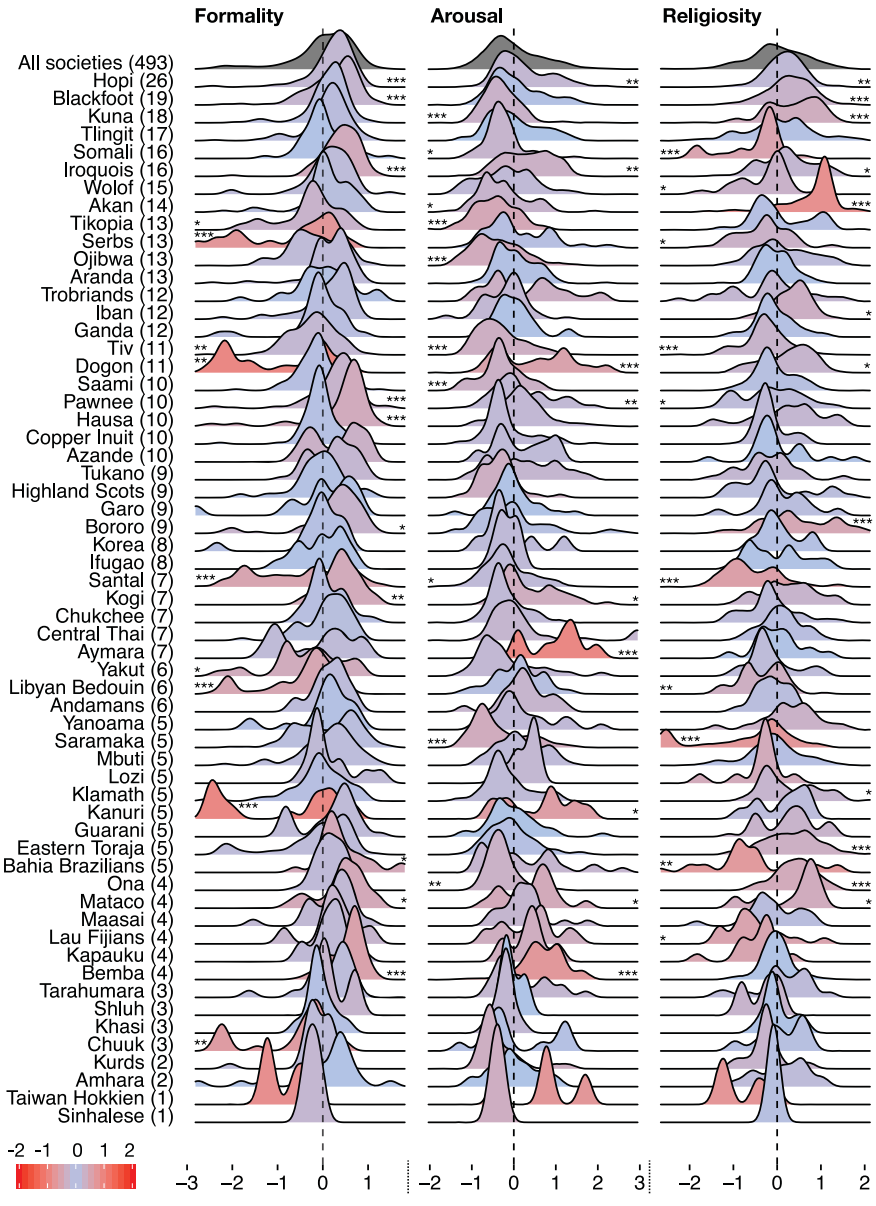


Fig. 3. Society-wise variation in musical behavior. Density plots for each society show the distributions of musical performances on each of the three principal components (Formality, Arousal, Religiosity). Distributions are based on posterior samples aggregated from corresponding ethnographic observations. Societies are ordered by the number of available documents in the NHS Ethnography (the number of documents per society is displayed in parentheses). Distributions are color-coded according to their mean distance from the global mean (in *z*-scores; redder distributions are farther from 0). Although some societies' means differ significantly from the global mean, the mean of each society's distribution is within 1.96 standard deviations of the global mean of 0. One society (Tzeltal) is not plotted because it has insufficient observations for a density plot. Asterisks denote society-level mean differences from the global mean. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

statistical techniques used, we applied them to a structurally analogous dataset whose latent dimensions are expected to vary across countries, namely climate features (for instance, temperature is related to elevation, which certainly is not universal); the results were entirely different from what we found when analyzing the NHS Ethnography (figs. S7 and S8 and Text S2.1.11).

The results suggest that societies' musical behaviors are largely similar to one another, such that the variability within a society exceeds the variability between them (all societies have more soothing songs, such as lullabies; more rousing songs, such as dance tunes; more stirring songs, such as prayers; and other recognizable kinds of musical performance), and that the appearance of uniqueness in the ethnographic record may reflect underreporting.

Associations between song and behavior, corrected for bias

Ethnographic descriptions of behavior are subject to several forms of selective nonreporting: Ethnographers may omit certain kinds of information because of their academic interests (e.g., the author focuses on farming and not shamanism), implicit or explicit biases (e.g., the author reports less information about the elderly), lack of knowledge (e.g., the author is unaware of food taboos), or inaccessibility (e.g., the author wants to report on infant care but is not granted access to infants). We cannot distinguish among these causes, but we can discern patterns of omission in the NHS Ethnography. For example, we found that when the singer's age is reported, the singer is likely to be young, but when the singer's age is not reported, cues that the singer is old are statistically present (such as the fact that a song is ceremonial). Such correlations—between the absence of certain values of one variable and the reporting of particular values of others—were aggregated into a model of missingness (Text S2.1.12) that forms part of the Bayesian principal components analysis reported above. This allowed us to assess variation in musical behavior worldwide, while accounting for reporting biases.

Next, to test hypotheses about the contexts with which music is strongly associated worldwide, in a similarly robust fashion, we compared the frequency with which a particular behavior appears in text describing song with the estimated frequency with which it appears across the board, in all the text written by that ethnographer about that society, which can be treated as the null distribution for that behavior. If a behavior is systematically associated with song, then its frequency in ethnographic descriptions of songs should exceed its frequency in that null distribution, which we estimated by randomly drawing the same number of passages from the same documents [see Text S2.2 for full model details].

We generated a list of 20 hypotheses about universal or widespread contexts for music (Table 1) from published work in anthropology, ethnomusicology, and cognitive science (*4*, *5*, *40*, *54*, *58*–*60*), together with a survey of nearly 1000 scholars that solicited opinions about which behaviors might be universally linked to music (Text S1.4.1). We then designed two sets of criteria for determining whether a given passage of ethnography represented a given behavior in this list. The first used human-annotated identifiers, capitalizing on the fact that every paragraph in the Probability Sample File comes tagged with one of more than 750 identifiers from the Outline of Cultural Materials (OCM),

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Table 1. Cross-cultural associations between song and other behaviors. We tested 20 hypothesized associations between song and other behaviors by comparing the frequency of a behavior in song-related passages to that in comparably-sized samples of text from the same sources that are not about song. Behavior was identified with two methods: topic annotations from the Outline of Cultural Materials ("OCM identifiers") and automatic detection of related keywords ("WordNet seed words"; see table S19). Significance tests compared the frequencies in the passages in the full Probability Sample File containing song-related keywords ("Song freq.") with the frequencies in a simulated null distribution of passages randomly selected from the same documents ("Null freq."). ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, using adjusted P values (*88*); 95% intervals for the null distribution are in parentheses.

Hypothesis	OCM identifier(s)	Song freq.	Null freq.	WordNet seed word(s)	Song freq.	Null freq.
Dance	DANCE	1499***	431 (397, 467)	dance	11,145***	3283 (3105, 3468)
Infancy	INFANT CARE	63*	44 (33, 57)	infant, baby, cradle, lullaby	688**	561 (491, 631)
Healing	MAGICAL AND MENTAL THERAPY; SHAMANS AND PSYCHOTHERAPISTS; MEDICAL THERAPY; MEDICAL CARE	1651***	1063 (1004, 1123)	heal, shaman, sick, cure	3983***	2466 (2317, 2619)
Religious activity	SHAMANS AND PSYCHOTHERAPISTS; RELIGIOUS EXPERIENCE; PRAYERS AND SACRIFICES; PURIFICATION AND ATONEMENT; ECSTATIC RELIGIOUS PRACTICES; REVELATION AND DIVINATION; RITUAL	3209***	2212 (2130, 2295)	religious, spiritual, ritual	8644***	5521 (5307, 5741)
Play	GAMES; CHILDHOOD ACTIVITIES	377***	277 (250, 304)	play, game, child, toy	4130***	2732 (2577, 2890)
Procession	SPECTACLES; NUPTIALS	371***	213 (188, 240)	wedding, parade, march, procession, funeral, coronation	2648***	1495 (1409, 1583)
Mourning	BURIAL PRACTICES AND FUNERALS; MOURNING; SPECIAL BURIAL PRACTICES AND FUNERALS	924***	517 (476, 557)	mourn, death, funeral	3784***	2511 (2373, 2655)
Ritual	RITUAL	187***	99 (81, 117)	ritual, ceremony	8520**	5138 (4941, 5343)
Entertainment	SPECTACLES	44***	20 (12, 29)	entertain, spectacle	744***	290 (256, 327)
Children	CHILDHOOD ACTIVITIES	178***	108 (90, 126)	child	4351***	3471 (3304, 3647)
Mood/emotions	DRIVES AND EMOTIONS	219***	138 (118, 159)	mood, emotion, emotive	796***	669 (607, 731)
Work	LABOR AND LEISURE	137***	60 (47, 75)	work, labor	3500**	3223 (3071, 3378)
Storytelling	VERBAL ARTS; LITERATURE	736***	537 (506, 567)	story, history, myth	2792***	2115 (1994, 2239)
Greeting visitors	VISITING AND HOSPITALITY	360***	172 (148, 196)	visit, greet, welcome	1611***	1084 (1008, 1162)
War	WARFARE	264	283 (253, 311)	war, battle, raid	3154***	2254 (2122, 2389)
Praise	STATUS, ROLE, AND PRESTIGE	385	355 (322, 388)	praise, admire, acclaim	481***	302 (267, 339)
Love	ARRANGING A MARRIAGE	158	140 (119, 162)	love, courtship	1625***	804 (734, 876)
Group bonding	SOCIAL RELATIONSHIPS AND GROUPS	141	163 (141, 187)	bond, cohesion	1582***	1424 (1344, 1508)
Marriage/weddings	NUPTIALS	327***	193 (169, 218)	marriage, wedding	2011	2256 (2108, 2410)
Art/creation	N/A	n/a	n/a	art, creation	905***	694 (630, 757)

such as MOURNING, INFANT CARE, or WARFARE.

The second set of criteria was needed because some hypotheses corresponded only loosely to the OCM identifiers (e.g., "love songs" is only a partial fit to ARRANGING A MARRIAGE and not an exact fit to any other identifier), and still others fit no identifier at all [e.g., "music perceived as art or as a creation" (*59*)]. So we designed a method that examined the text directly. Starting with a small set of seed words associated with each hypothesis (e.g., "religious," "spiritual," and "ritual" for the hypothesis that music is associated with religious activity), we used the WordNet lexical database (*87*) to automatically generate lists of conceptually related terms (e.g., "rite" and "sacred"). We manually filtered the

Downloaded from http://science.sciencemag.org/ on November 28, 2019

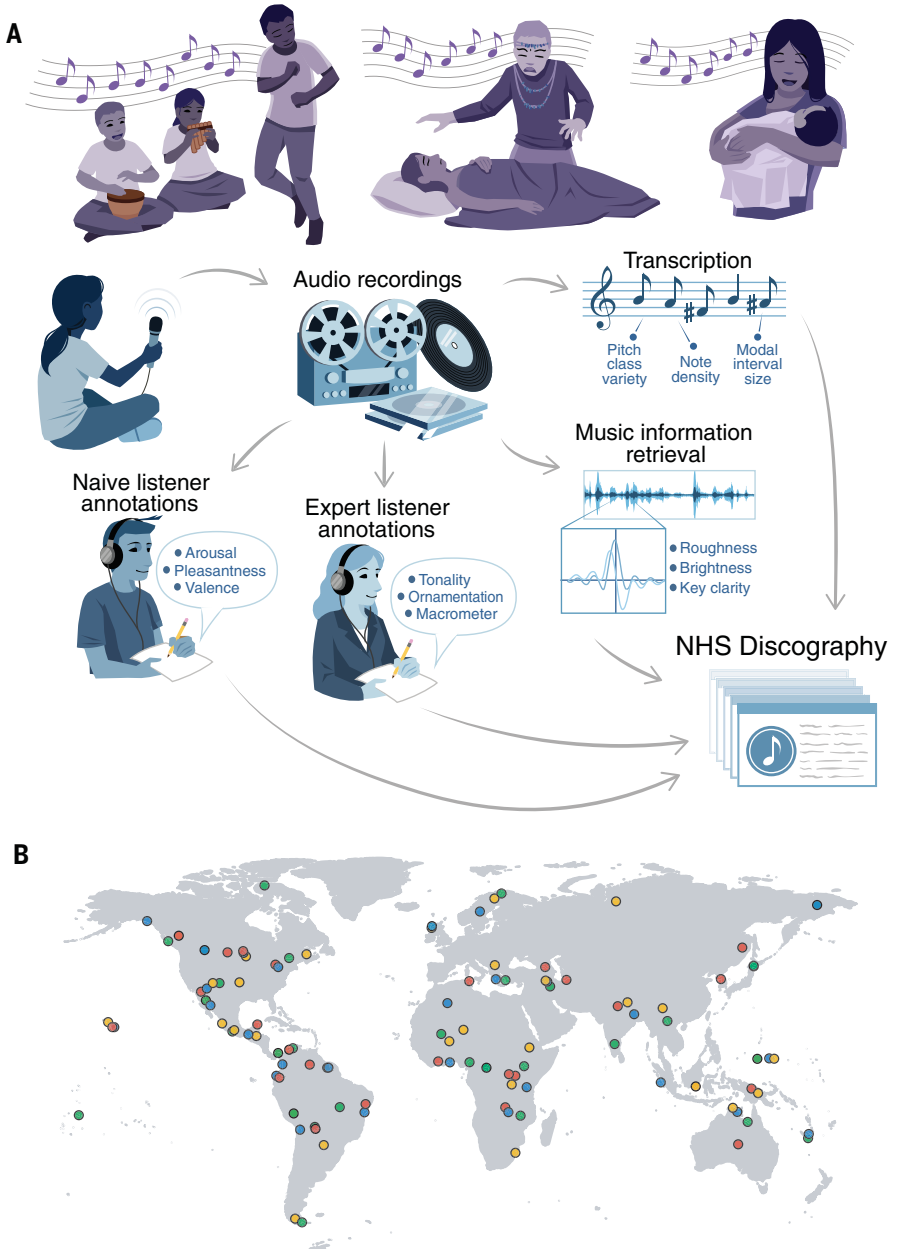


Fig. 4. Design of the NHS Discography. (**A**) Illustration depicting the sequence from acts of singing to the audio discography. People produce songs, which scholars record. We aggregate and analyze the recordings via four methods: automatic music information retrieval, annotations from expert listeners, annotations from naïve listeners, and staff notation transcriptions (from which annotations are automatically generated). The raw audio, four types of annotations, transcriptions, and metadata together form the NHS Discography. (**B**) Plot of the locations of the 86 societies represented, with points colored by the song type in each recording (blue, dance; red, healing; yellow, love; green, lullaby). Codebooks listing all available data are in tables S1 and S7 to S11; a listing of societies and locations from which recordings were gathered is in table S22.

lists to remove irrelevant words and homonyms and add relevant keywords that may have been missed, then conducted word stemming to fill out plurals and other grammatical variants (full lists are in table S19). Each method has limitations: Automated dictionary methods can erroneously flag a passage containing a word that is ambiguous, whereas the human-coded OCM identifiers may miss a relevant passage, misinterpret the original ethnography, or paint with too broad a brush, applying a tag to a whole paragraph or to several pages of text. Where the two methods converge, support for a hypothesis is particularly convincing.

After controlling for ethnographer bias via the method described above, and adjusting the *P* values for multiple hypotheses (*88*), we found support from both methods for 14 of the 20 hypothesized associations between music and a behavioral context, and support from one method for the remaining six (Table 1). To verify that these analyses specifically confirmed the hypotheses, as opposed to being an artifact of some other nonrandom patterning in this dataset, we reran them on a set of additional OCM identifiers matched in frequency to the ones used above [see Text S2.2.2 for a description of the selection procedure]. They covered a broad swath of topics, including DOMESTICATED ANIMALS, POLYGAMY, and LEGAL NORMS that were not hypothesized to be related to song (the full list is in table S20). We find that only one appeared more frequently in song-related paragraphs than in the simulated null distribution (CEREAL AGRICULTURE; see table S20 for full results). This contrasts sharply with the associations reported in Table 1, suggesting that they represent bona fide regularities in the behavioral contexts of music.

Universality of musical forms

We now turn to the NHS Discography to examine the musical content of songs in four behavioral contexts (dance, lullaby, healing, and love; Fig. 4A), selected because each appears in the NHS Ethnography, is widespread in traditional cultures (*59*), and exhibits shared features across societies (*54*). Using predetermined criteria based on liner notes and supporting ethnographic text (table S21), and seeking recordings of each type from each of the 30 geographic regions, we found 118 songs of the 120 possibilities (4 contexts × 30 regions) from 86 societies (Fig. 4B). This coverage underscores the universality of these four types; indeed, in the two possibilities we failed to find (healing songs from Scandinavia and from the British Isles), documentary evidence shows that both existed (*89*, *90*) despite our failure to find audio recordings of the practice. The recordings may be unavailable because healing songs were rare by the early 1900s, roughly when portable field recording became feasible.

The data describing each song comprised (i) machine summaries of the raw audio using automatic music information retrieval techniques, particularly the audio's spectral features (e.g., mean brightness and roughness, variability of spectral entropy) (Text S1.2.1); (ii) general impressions of musical features (e.g., whether its emotional valence was happy or sad) by untrained listeners recruited online from the United States and India (Text S1.2.2); (iii) ratings of additional music-theoretic features such as high-level rhythmic grouping structure [similar in concept to previous rating-scale approaches to analyzing world music (*10*, *53*)] from a group of 30 expert musicians including Ph.D. ethnomusicologists and music theorists (Text S1.2.3); and (iv) detailed manual

Downloaded from http://science.sciencemag.org/ on November 28, 2019

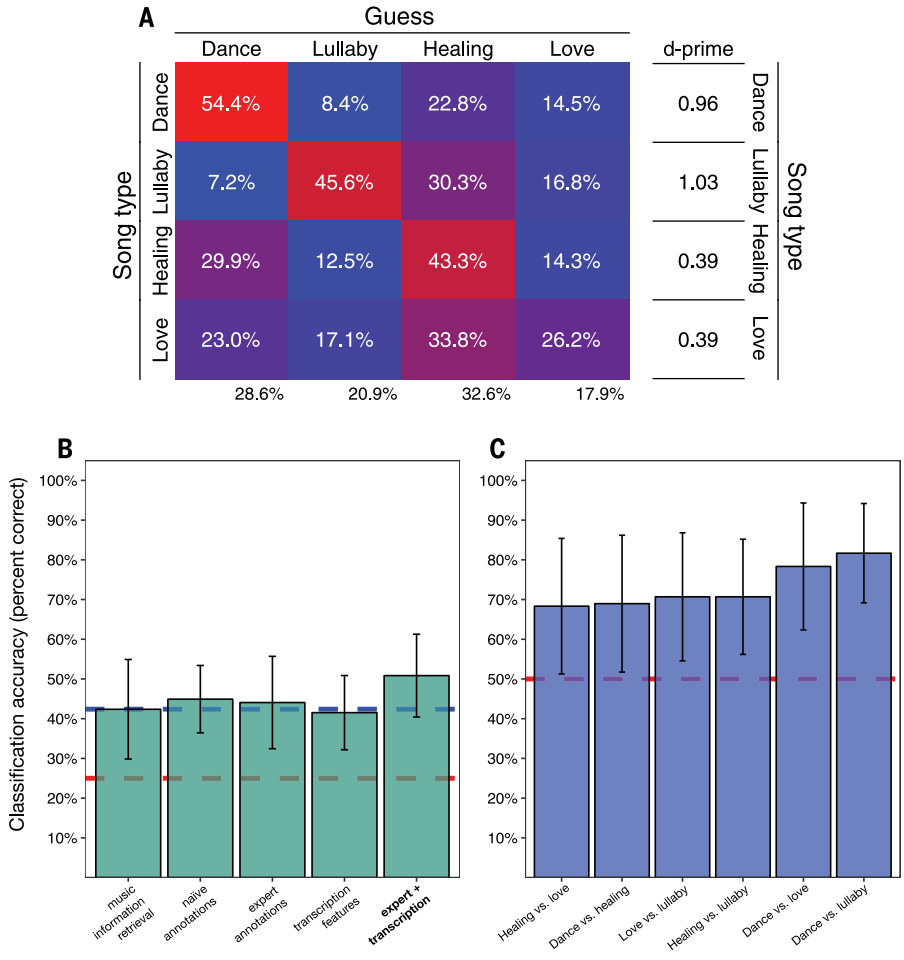


Fig. 5. Form and function in song. (**A**) In a massive online experiment (N = 29,357), listeners categorized dance songs, lullabies, healing songs, and love songs at rates higher than chance level of 25%, but their responses to love songs were by far the most ambiguous (the heat map shows average percent correct, color-coded from lowest magnitude, in blue, to highest magnitude, in red). Note that the marginals (below the heat map) are not evenly distributed across behavioral contexts: Listeners guessed "healing" most often and "love" least often despite the equal number of each in the materials. The *d*-prime scores estimate listeners' sensitivity to the song-type signal independent of this response bias. (**B**) Categorical classification of the behavioral contexts of songs, using each of the four representations in the NHS Discography, is substantially above the chance performance level of 25% (dotted red line) and is indistinguishable from the performance of human listeners, 42.4% (dotted blue line). The classifier that combines expert annotations with transcription features (the two representations that best ignore background sounds and other context) performs at 50.8% correct, above the level of human listeners. (**C**) Binary classifiers that use the expert annotation + transcription feature representations to distinguish pairs of behavioral contexts [e.g., dance from love songs, as opposed to the four-way classification in (B)] perform above the chance level of 50% (dotted red line). Error bars represent 95% confidence intervals from corrected resampled *t* tests (*94*).

transcriptions, also by expert musicians, of musical features (e.g., note density of sung pitches) (Text S1.2.4). To ensure that classifications were driven only by the content of the music, we excluded any variables that carried explicit or implicit information about the context (*54*), such as the number of singers audible on a recording and a coding of polyphony (which indicates the same thing implicitly). This exclusion could be complete only in the manual transcriptions, which are restricted to data on vocalizations; the music information retrieval and naïve listener data are practically inseparable from contextual information, and the expert listener ratings contain at least a small amount, because despite being told to ignore the context, the experts could still hear some of it, such as accompanying instruments. [See Text S2.3.1 for details about variable exclusion.]

Listeners accurately identify the behavioral contexts of songs

In a previous study, people listened to recordings from the NHS Discography and rated their confidence in each of six possible behavioral contexts (e.g., "used to soothe a baby"). On average, the listeners successfully inferred a song's behavioral context from its musical forms: The songs that were actually used to soothe a baby (i.e., lullabies) were rated highest as "used to soothe a baby"; dance songs were rated highly as "used for dancing," and so on (*54*).

We ran a massive conceptual replication (Text S1.4.2) where 29,357 visitors to the citizen-science website http://themusiclab.org listened to songs drawn at random from the NHS Discography and were asked to guess what kind of song they were listening to from among four alternatives (yielding 185,832 ratings, i.e., 118 songs rated about 1500 times each). Participants also reported their musical skill level and degree of familiarity with world music. Listeners guessed the behavioral contexts with a level of accuracy (42.4%) that is well above chance (25%), showing that the acoustic properties of a song performance reflect its behavioral context in ways that span human cultures.

The confusion matrix (Fig. 5A) shows that listeners identified dance songs most accurately (54.4%), followed by lullabies (45.6%), healing songs (43.3%), and love songs (26.2%), all significantly above chance (Ps < 0.001). Dance songs and lullabies were the least likely to be confused with each other, presumably because of their many contrasting features, such as tempo (a possibility we examine below; see Table 2). The column marginals suggest that the raters were biased toward identifying recordings as healing songs (32.6%, above their actual proportion of 23.7%) and away from identifying them as love songs (17.9%), possibly because healing songs are less familiar to Westernized listeners and they were overcompensating in identifying examples. As in previous research (*54*), love songs were least reliably identified, despite their ubiquity in Western popular music, possibly because they span a wide range of styles (for example, the vastly different Elvis Presley hit singles "Love Me Tender" and "Burning Love"). Nonetheless, *d*-prime scores (Fig. 5A), which capture the sensitivity to a signal independently of response bias, show that all behavioral contexts were identified at a rate higher than chance (d' = 0).

Are accurate identifications of the contexts of culturally unfamiliar songs restricted to listeners with musical training or exposure to world music? In a regression analysis, we found that participants' categorization accuracy was statistically related to their self-reported musical skill [$F(4,16245) = 2.57$, $P = 0.036$] and their familiarity with world music [$F(3,16167) = 36.9$, $P < 0.001$; statistics from linear probability models], but with small effect sizes: The largest difference was a 4.7–percentage point advantage for participants

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Table 2. Features of songs that distinguish between behavioral contexts. The table reports the predictive influence of musical features in the NHS Discography in distinguishing song types across cultures, ordered by their overall influence across all behavioral contexts. The classifiers used the average rating for each feature across 30 annotators. The coefficients are from a penalized logistic regression with standardized features and are selected for inclusion using a LASSO for variable selection. For brevity, we only present the subset of features with notable influence on a pairwise comparison (coefficients greater than 0.1). Changes in the values of the coefficients produce changes in the predicted log-odds ratio, so the values in the table can be interpreted as in a logistic regression.

		Coefficient (pairwise comparison)					
Musical feature	**Definition**	**Dance (–) vs. Lullaby (+)**	**Dance (–) vs. Love (+)**	**Healing (–) vs. Lullaby (+)**	**Love (–) vs. Lullaby (+)**	**Dance (–) vs. Healing (+)**	**Healing (–) vs. Love (+)**
Accent	The differentiation of musical pulses, usually by volume or emphasis of articulation. A fluid, gentle song will have few accents and a correspondingly low value.	−0.64	−0.24	−0.85	−0.41	.	−0.34
Tempo	The rate of salient rhythmic pulses, measured in beats per minute; the perceived speed of the music. A fast song will have a high value.	−0.65	−0.51	.	.	−0.76	.
Quality of pitch collection	Major versus minor key. In Western music, a key usually has a "minor" quality if its third note is three semitones from the tonic. This variable was derived from annotators' qualitative categorization of the pitch collection, which we then dichotomized into Major (0) or Minor (1).	.	0.26	0.44	.	−0.37	0.35
Consistency of macrometer	Meter refers to salient repetitive patterns of accent within a stream of pulses. A micrometer refers to the low-level pattern of accents; a macrometer refers to repetitive patterns of micrometer groups. This variable refers to the consistency of the macrometer, in an ordinal scale, from "No macrometer" (1) to "Totally clear macrometer" (6). A song with a highly variable macrometer will have a low value.	−0.44	−0.49	.	.	−0.46	.
Number of common intervals	Variability in interval sizes, measured by the number of different melodic interval sizes that constitute more than 9% of the song's intervals. A song with a large number of different melodic interval sizes will have a high value.	.	0.58	.	.	.	0.62
Pitch range	The musical distance between the extremes of pitch in a melody, measured in semitones. A song that includes very high and very low pitches will have a high value.	.	.	.	−0.49	.	.

continued on next page

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Musical feature	Definition	Coefficient (pairwise comparison)					
		Dance (−) vs. Lullaby (+)	Dance (−) vs. Love (+)	Healing (−) vs. Lullaby (+)	Love (−) vs. Lullaby (+)	Dance (−) vs. Healing (+)	Healing (−) vs. Love (+)
Stepwise motion	Stepwise motion refers to melodic strings of consecutive notes (1 or 2 semitones apart), without skips or leaps. This variable consists of the fraction of all intervals in a song that are 1 or 2 semitones in size. A song with many melodic leaps will have a low value.	.	.	.	.	0.61	−0.20
Tension/release	The degree to which the passage is perceived to build and release tension via changes in melodic contour, harmonic progression, rhythm, motivic development, accent, or instrumentation. If so, the song is annotated with a value of 1.	.	0.27	.	.	.	0.27
Average melodic interval size	The average of all interval sizes between successive melodic pitches, measured in semitones on a 12-tone equal temperament scale, rather than in absolute frequencies. A melody with many wide leaps between pitches will have a high value.	.	−0.46	.	.	.	.
Average note duration	The mean of all note durations; a song predominated by short notes will have a low value.	.	.	.	.	.	−0.49
Triple micrometer	A low-level pattern of accents that groups together pulses in threes.	.	.	.	.	−0.23	.
Predominance of most common pitch class	Variety versus monotony of the melody, measured by the ratio of the proportion of occurrences of the second most common pitch (collapsing across octaves) to the proportion of occurrences of the most common pitch; monotonous melodies will have low values.	.	.	.	.	−0.48	.
Rhythmic variation	Variety versus monotony of the rhythm, judged subjectively and dichotomously. Repetitive songs have a low value.	.	.	.	.	0.42	.
Tempo variation	Changes in tempo: A song that is perceived to speed up or slow down is annotated with a value of 1.	.	.	.	.	.	−0.27
Ornamentation	Complex melodic variation or "decoration" of a perceived underlying musical structure. A song perceived as having ornamentation is annotated with a value of 1.	.	0.25	.	.	.	.

continued on next page

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Musical feature	Definition	Coefficient (pairwise comparison)					
		Dance (−) vs. Lullaby (+)	Dance (−) vs. Love (+)	Healing (−) vs. Lullaby (+)	Love (−) vs. Lullaby (+)	Dance (−) vs. Healing (+)	Healing (−) vs. Love (+)
Pitch class variation	A pitch class is the group of pitches that sound equivalent at different octaves, such as all the Cs, not just middle C. This variable, another indicator of melodic variety, counts the number of pitch classes that appear at least once in the song.	.	.	−0.25	.	.	.
Triple macrometer	If a melody arranges micrometer groups into larger phrases of three, like a waltz, it is annotated with a value of 1.	.	.	0.14	.	.	.
Predominance of most common interval	Variability among pitch intervals, measured as the fraction of all intervals that are the most common interval size. A song with little variability in interval sizes will have a high value.	.	.	.	.	0.12	.

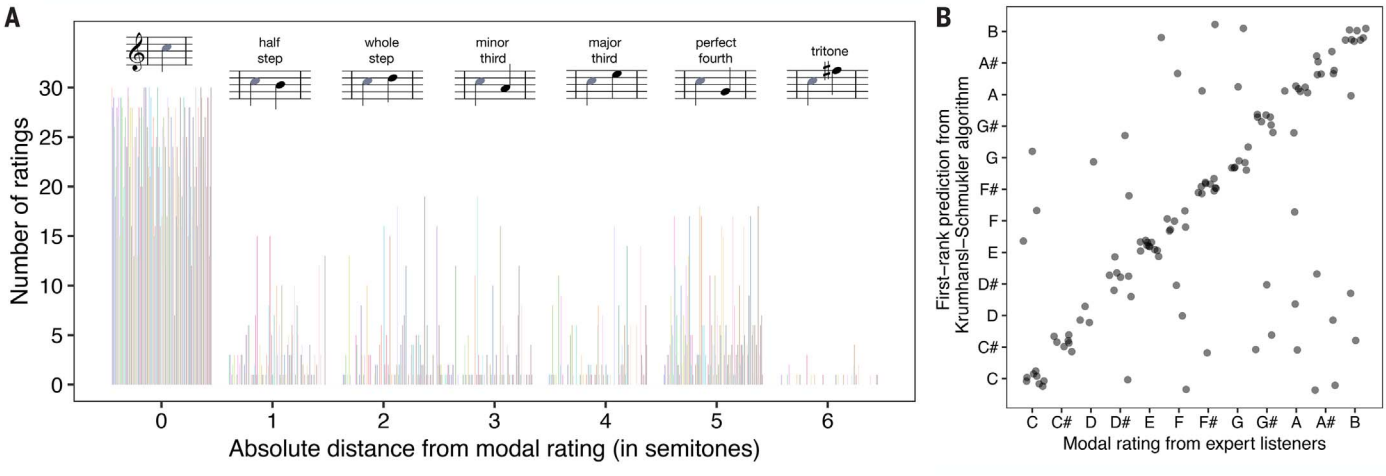


Fig. 6. Signatures of tonality in the NHS Discography. (**A**) Histograms representing 30 expert listeners' ratings of tonal centers in all 118 songs, each song corresponding to a different color, show two main findings: (i) Most songs' distributions are unimodal, such that most listeners agreed on a single tonal center (represented by the value 0). (ii) When listeners disagree, they are multimodal, with the most popular second mode (in absolute distance) five semitones away from the overall mode, a perfect fourth. The music notation is provided as a hypothetical example only, with C as a reference tonal center; note that the ratings of tonal centers could be at any pitch level. (**B**) The scatterplot shows the correspondence between modal ratings of expert listeners with the first-rank predictions from the Krumhansl-Schmuckler key-finding algorithm. Points are jittered to avoid overlap. Note that pitch classes are circular (i.e., C is one semitone away from C# and from B) but the plot is not; distances on the axes of (B) should be interpreted accordingly.

who reported that they were "somewhat familiar with traditional music" relative to those who reported that they had never heard it, and a 1.3–percentage point advantage for participants who reported that they have "a lot of skill" relative to "no skill at all." Moreover, when limiting the dataset to listeners with "no skill at all" or listeners who had "never heard traditional music," mean accuracy was almost identical to the overall cohort. These findings suggest that although musical experience enhances the ability to detect the behavioral contexts of songs from unfamiliar cultures, it is not necessary.

Quantitative representations of musical forms accurately predict behavioral contexts of song

If listeners can accurately identify the behavioral contexts of songs from unfamiliar cultures, there must be acoustic features that universally tend to be associated with these contexts. To identify them, we evaluated the relationship between a song's musical forms [measured in four ways; see Text S1.2.5 and (*12*, *31*, *32*, *91*–*93*) for discussion of how difficult it is to represent music quantitatively] and its behavioral context. We used a cross-validation procedure that determined whether the pattern of correlation between musical forms and context computed from a subset of the

Downloaded from http://science.sciencemag.org/ on November 28, 2019

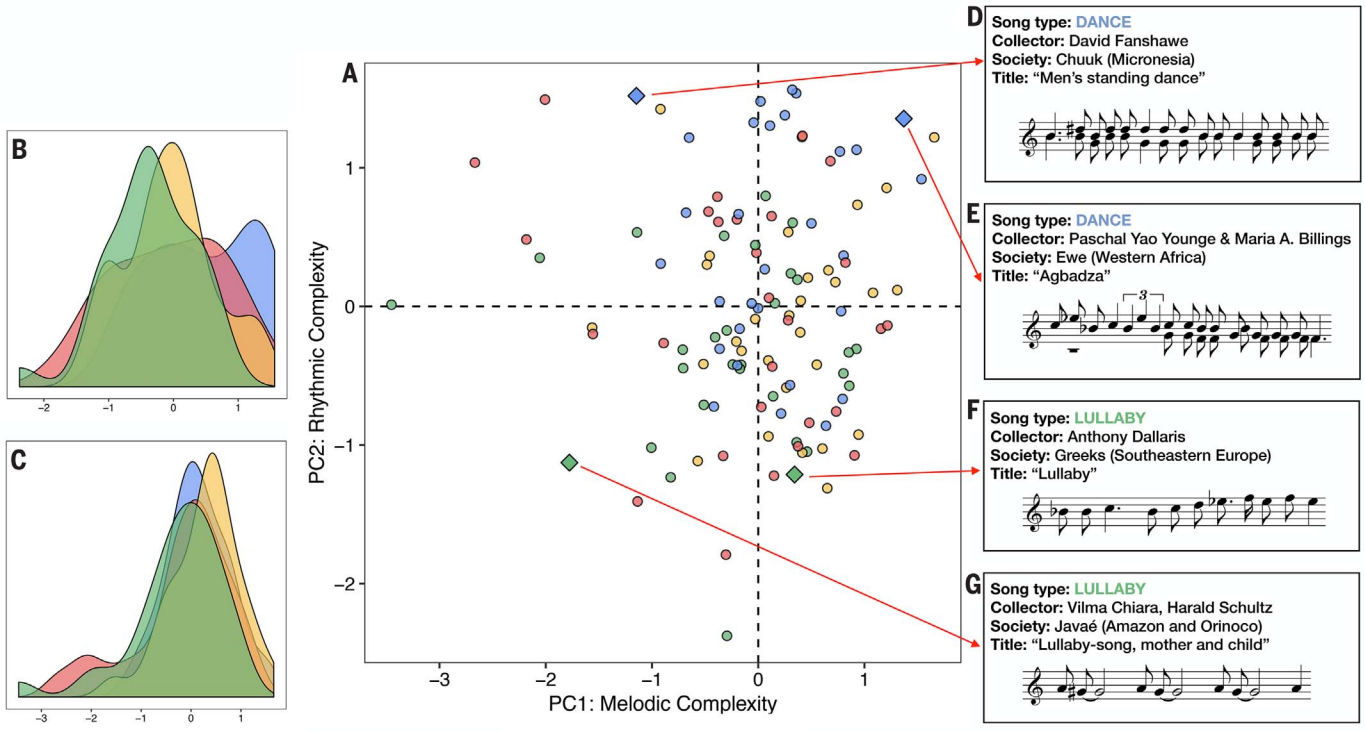


Fig. 7. Dimensions of musical variation in the NHS Discography. (**A**) A Bayesian principal components analysis reduction of expert annotations and transcription features (the representations least contaminated by contextual features) shows that these measurements fall along two dimensions that may be interpreted as rhythmic complexity and melodic complexity. (**B** and **C**) Histograms for each dimension show the differences—or lack thereof—between behavioral contexts. (**D** to **G**) Excerpts of transcriptions from songs at extremes from each of the four quadrants, to validate the dimension reduction visually. The two songs at the high–rhythmic complexity quadrants are dance songs (in blue); the two songs at the low–rhythmic complexity quadrants are lullabies (in green). Healing songs are depicted in red and love songs in yellow. Readers can listen to excerpts from all songs in the corpus at http://osf.io/jmv3q; an interactive version of this plot is available at http://themusiclab.org/nhsplots.

regions could be generalized to predict a song's context in the other regions (as opposed to being overfitted to arbitrary correlations within that subsample). Specifically, we trained a least absolute shrinkage and selection operator (LASSO) multinomial logistic classifier (*94*) on the behavioral context of all the songs in 29 of the 30 regions in the NHS Discography, and used it to predict the context of the unseen songs in the 30th. We ran this procedure 30 times, omitting a different region each time (table S23 and Text S2.3.2). We compared the accuracy of these predictions to two baselines: pure chance (25%) and the accuracy of listeners in the massive online experiment (see above) when guessing the behavioral context from among four alternatives (42.4%).

We found that with each of the four representations, the musical forms of a song can predict its behavioral context (Fig. 5B) at high rates, comparable to those of the human listeners in the online experiment. This finding was not attributable to information in the recordings other than the singing, which could be problematic if, for example, the presence of a musical instrument on a recording indicated that it was likelier to be a dance song than a lullaby (*54*), artificially improving classification. Representations with the least extraneous influence—the expert annotators and the summary features extracted from transcriptions—had no lower classification accuracy than the other representations. And a classifier run on combined expert + transcription data had the best performance of all, 50.8% [95% CI (40.4%, 61.3%), computed by corrected resampled t test (*95*)].

To ensure that this accuracy did not merely consist of patterns in one society predicting patterns in historically or geographically related ones, we repeated the analyses, cross-validating across groupings of societies, including superordinate world region (e.g., "Asia"), subsistence type (e.g., "hunter-gatherers"), and Old versus New World. In many cases, the classifier performed comparably to the main model (table S24), although low power in some cases (i.e., training on less than half the corpus) substantially reduced precision.

In sum, the acoustic form of vocal music predicts its behavioral contexts worldwide (*54*), at least in the contexts of dance, lullaby, healing, and love: All classifiers performed above chance and within 1.96 standard errors of the performance of human listeners.

Musical features that characterize the behavioral contexts of songs across societies

Showing that the musical features of songs predict their behavioral context provides no information about which musical features those are. To help identify them, we determined how well the combined expert + transcription data distinguished between specific pairs of behavioral contexts rather than among all four, using a simplified form of the classifiers described above, which not only distinguished the contexts but also identified the most reliable predictors of each contrast, without overfitting (*96*). This can reveal whether tempo, for example, helps distinguish dance songs from lullabies while failing to distinguish lullabies from love songs.

Performance once again significantly exceeded chance (in this case, 50%) for all six comparisons (Ps < 0.05; Fig. 5C). Table 2 lays out the musical features that drive these successful predictions and thereby characterize the four song types across cultures. Some are consistent with common sense; for instance, dance songs differ from lullabies in tempo, accent, and the consistency of their macrometer (i.e., the superordinate grouping of rhythmic notes). Other distinguishers are

Downloaded from http://science.sciencemag.org/ on November 28, 2019

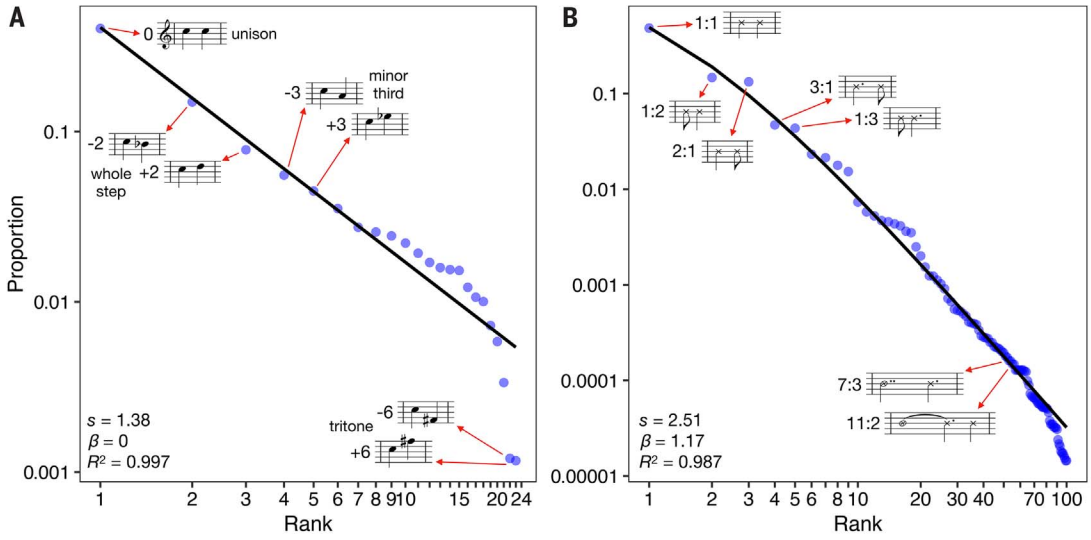


Fig. 8. The distributions of melodic and rhythmic patterns in the NHS Discography follow power laws. (**A** and **B**) We computed relative melodic (A) and rhythmic (B) bigrams and examined their distributions in the corpus. Both distributions followed a power law; the parameter estimates in the inset correspond to those from the generalized Zipf-Mandelbrot law, where *s* refers to the exponent of the power law and β refers to the Mandelbrot offset. Note that in both plots, the axes are on logarithmic scales. The full lists of bigrams are in tables S28 and S29.

subtler: The most common interval of a song occurs a smaller proportion of the time in a dance song than in a healing song, which suggests that dance songs are more melodically variable than healing songs (for explanations of musical terminology, see Table 2). Similarly, it is unsurprising that lullabies and love songs are more difficult to distinguish than lullabies and dance songs (*97*); nonetheless, they may be distinguished by two features: the strength of metrical accents and the size of the pitch range (both larger in love songs).

In sum, four common song categories, distinguished by their contexts and goals, tend to have distinctive musical qualities worldwide. These results suggest that universal features of human psychology bias people to produce and enjoy songs with certain kinds of rhythmic or melodic patterning that naturally go with certain moods, desires, and themes. These patterns do not consist of concrete acoustic features, such as a specific melody or rhythm, but rather of relational properties such as accent, meter, and interval structure.

Of course, classification accuracy that is twice the level of chance still falls well short of perfect prediction; hence, many aspects of music cannot be manifestations of universal psychological reactions. Although musical features can predict differences between songs from these four behavioral contexts, a given song may be sung in a particular context for other reasons, including its lyrics, its history, the style and instrumentation of its performance, its association with mythical or religious themes, and constraints of the culture's musical idiom. And although we have shown that Western listeners, who have been exposed to a vast range of musical styles and idioms, can distinguish the behavioral contexts of songs from non-Western societies, we do not know whether non-Western listeners can do the same. To reinforce the hypothesis of universal associations between musical form and context, similar methods should be tested with non-Western listeners.

Explorations of the structure of musical forms

The NHS Discography can be used to explore world music in many other ways. We present three exploratory analyses here, mindful of the limitation that they may apply only to the four genres the corpus includes.

Signatures of tonality appear in all societies studied

A basic feature of many styles of music is tonality, in which a melody is composed of a fixed set of discrete tones [perceived pitches as opposed to actual pitches, a distinction dating to Aristoxenus's *Elementa Harmonica* (*98*)], and some tones are psychologically dependent on others, with one tone felt to be central or stable (*99*–*101*). This tone (more accurately, perceived pitch class, embracing all the tones one or more octaves apart) is called the tonal center or tonic, and listeners characterize it as a reference point, point of stability, basis tone, "home," or tone that the melody "is built around" and where it "should end." For example, the tonal center of "Row Your Boat" is found in each of the "row"s, the last "merrily," and the song's last note, "dream."

Although tonality has been studied in a few non-Western societies (*102*, *103*), its cross-cultural distribution is unknown. Indeed, the ethnomusicologists who responded to our survey (Text S1.4.1) were split over whether the music of all societies should be expected to have a tonal center: 48% responded "probably not universal" or "definitely not universal." The issue is important because a tonal system is a likely prerequisite for analyzing music, in all its diversity, as the product of an abstract musical grammar (*73*). Tonality also motivates the hypothesis that melody is rooted in the brain's analysis of harmonically complex tones (*104*). In this theory, a melody can be considered a set of "serialized overtones," the harmonically related frequencies ordinarily superimposed in the rich tone produced by an elongated resonator such as the human vocal tract. In tonal melodies, the tonic corresponds to the fundamental frequency of the disassembled complex tone, and listeners tend to favor tones in the same pitch class as harmonics of the fundamental (*105*).

To explore tonality in the NHS Discography, we analyzed the expert listener annotations and the transcriptions (Text S2.4.1). Each of the 30 expert listeners was asked, for each song, whether or not they heard at least one tonal center, defined subjectively as above. The results were unambiguous: 97.8% of ratings were in the affirmative. More than two-thirds of songs were rated as "tonal" by all 30 expert listeners, and 113 of the 118 were rated as tonal by more than 90% of them. The song with the most ambiguous tonality (the Kwakwaka'wakw healing song) still had a majority of raters respond in the affirmative (60%).

If listeners heard a tonal center, they were asked to name its pitch class. Here too, listeners were highly consistent: Either there

Downloaded from http://science.sciencemag.org/ on November 28, 2019

was widespread agreement on a single tonal center or the responses fell into two or three tonal centers (Fig. 6A; the distributions of tonality ratings for all 118 songs are in fig. S10). We used Hartigan's dip test (*106*) to measure the multimodality of the ratings. In the 73 songs that the test classified as unimodal, 85.3% of ratings were in agreement with the modal pitch class. In the remaining 45 songs, 81.7% of ratings were in agreement with the two most popular pitch classes, and 90.4% were in agreement with the three most popular. The expert listeners included six Ph.D. ethnomusicologists and six Ph.D. music theorists; when restricting the ratings to this group alone, the levels of consistency were comparable.

In songs where the ratings were multimodally distributed, the modal tones were often hierarchically related; for instance, ratings for the Ojibwa healing song were evenly split between B (pitch class 11) and E (pitch class 4), which are a perfect fourth (five semitones) apart. The most common intervals between the two modal tones were the perfect fourth (in 15 songs), a half-step (one semitone, in nine songs), a whole step (two semitones, in eight songs), a major third (four semitones, in seven songs), and a minor third (three semitones, in six songs).

We cannot know which features of a given recording our listeners were responding to in attributing a tonal center to it, nor whether their attributions depended on expertise that ordinary listeners lack. We thus sought converging, objective evidence for the prevalence of tonality in the world's music by submitting NHS Discography transcriptions to the Krumhansl-Schmuckler key-finding algorithm (*107*). This algorithm sums the durations of the tones in a piece of music and correlates this vector with each of a family of candidate vectors, one for each key, consisting of the relative centralities of those pitch classes in that key. The algorithm's first guess (i.e., the key corresponding to the most highly correlated vector) matched the expert listeners' ratings of the tonal center 85.6% of the time (measured via a weighted average of its hit rate for the most common expert rating when the ratings were unimodal and either of the two most common ratings when they were multimodal). When we relaxed the criterion for a match to the algorithm's first- and second-ranked guesses, it matched the listeners' ratings on 94.1% of songs; adding its third-ranked estimate resulted in matches 97.5% of the time, and adding the fourth resulted in matches with 98.3% [all Ps < 0.0001 above the chance level of 9.1%, using a permutation test (Text S2.4.1)]. These results provide convergent evidence for the presence of tonality in the NHS Discography songs (Fig. 6B).

These conclusions are limited in several ways. First, they are based on songs from only four behavioral contexts, omitting others such as mourning, storytelling, play, war, and celebration. Second, the transcriptions were created manually and could have been influenced by the musical ears and knowledge of the expert transcribers. (Current music information retrieval algorithms are not robust enough to transcribe melodies accurately, especially from noisy field recordings, but improved ones could address this issue.) The same limitation may apply to the ratings of our expert listeners. Finally, the findings do not show how the people from the societies in which NHS Discography songs were recorded hear the tonality in their own music. To test the universality of tonality perception, one would need to conduct field experiments in diverse populations.

Music varies along two dimensions of complexity

To examine patterns of variation among the songs in the NHS Discography, we applied the same kind of Bayesian principal components analysis used for the NHS Ethnography to the combination of expert annotations and transcription features (i.e., the representations that focus most on the singing, excluding context). The results yielded two dimensions, which together explain 23.9% of the variability in musical features. The first, which we call Melodic Complexity, accounts for 13.1% of the variance (including error noise); heavily loading variables included the number of common intervals, pitch range, and ornamentation (all positively) and the predominance of the most common pitch class, the predominance of the most common interval, and the distance between the most common intervals (all negatively; see table S25). The second, which we call Rhythmic Complexity, accounts for 10.8% of the variance; heavily loading variables included tempo, note density, syncopation, accent, and consistency of macrometer (all positively) and the average note duration and duration of melodic arcs (all negatively; see table S26). The interpretation of the dimensions is further supported in Fig. 7, which shows excerpts of transcriptions at the extremes of each dimension; an interactive version is at http://themusiclab.org/nhsplots.

In contrast to the NHS Ethnography, the principal components space for the NHS Discography does not distinguish the four behavioral contexts of songs in the corpus. We found that only 39.8% of songs matched their nearest centroid (overall $P = 0.063$ from a permutation test; dance: 56.7%, $P = 0.12$; healing: 7.14%, $P > 0.99$; love: 43.3%, $P = 0.62$; lullaby: 50.0%, $P = 0.37$; a confusion matrix is in table S27). Similarly, k-means clustering on the principal components space, asserting $k = 4$ (because there are four known clusters), failed to reliably capture any of the behavioral contexts. Finally, given the lack of predictive accuracy of songs' location in the two-dimensional space, we explored each dimension's predictive accuracy individually, using t tests of each context against the other three, adjusted for multiple comparisons (*88*). Melodic complexity did not predict context (dance, $P = 0.79$; healing, $P = 0.96$; love, $P = 0.13$; lullaby, $P = 0.35$). However, rhythmic complexity did distinguish dance songs (which were more rhythmically complex, $P = 0.01$) and lullabies (which were less rhythmically complex, $P = 0.03$) from other songs; it did not distinguish healing or love songs (Ps > 0.99). When we adjusted these analyses to account for across-region variability, the results were comparable (Text S2.4.2). Thus, although musical content systematically varies in two ways across cultures, this variation is mostly unrelated to the behavioral contexts of the songs, perhaps because complexity captures distinctions that are salient to music analysts but not strongly evocative of particular moods or themes among the singers and listeners themselves.

Melodic and rhythmic bigrams are distributed according to power laws

Many phenomena in the social and biological sciences are characterized by Zipf's law (*108*), in which the probability of an event is inversely proportional to its rank in frequency, an example of a power-law distribution (in the Zipfian case, the exponent is 1). Power-law distributions (as opposed to, say, the geometric distribution) have two key properties: A small number of highly frequent events account for the majority of observations, and there are a large number of individually improbable events whose probability falls off slowly in a thick tail (*109*).

In language, for example, a few words appear with very high frequency, such as pronouns, while a great many are rare, such as the names of species of trees, but any sample will nonetheless tend to contain several rare words (*110*). A similar pattern is found in the distribution of colors among paintings in a given period of art history (*111*). In music, Zipf's law has been observed in the melodic intervals of Bach, Chopin, Debussy, Mendelssohn, Mozart, and Schoenberg (*112*–*116*); in the loudness and pitch fluctuations in Scott Joplin piano rags (*117*); in the harmonies (*118*–*120*) and rhythms of classical music (*121*); and, as Zipf himself noted, in melodies composed by Mozart, Chopin, Irving Berlin, and Jerome Kern (*108*).

We tested whether the presence of power-law distributions is a property of music worldwide by tallying relative melodic bigrams (the number of semitones separating each pair of successive notes) and relative rhythmic bigrams (the ratio of the durations of each pair of successive notes) for all NHS Discography transcriptions (Text S2.4.3). The bigrams overlapped, with the second note of one bigram also serving as the first note of the next.

We found that both the melodic and rhythmic bigram distributions followed power laws

Downloaded from http://science.sciencemag.org/ on November 28, 2019

(Fig. 8), and this finding held worldwide: The fit between the observed bigrams and the best-fitting power function was high within each region (melodic bigrams: median R^2 = 0.97, range 0.92 to 0.99; rhythmic bigrams: median R^2 = 0.98, range 0.88 to 0.99). The most prevalent bigrams were the simplest. Among the melodic bigrams (Fig. 8A), three small intervals (unison, major second, and minor third) accounted for 73% of the bigrams; the tritone (six semitones) was the rarest, accounting for only 0.2%. The prevalence of these bigrams is significant: Using only unisons, major seconds, and minor thirds, one can construct any melody in a pentatonic scale, a scale found in many cultures (*122*). Among the rhythmic bigrams (Fig. 8B), three patterns with simple integer ratios (1:1, 2:1, and 3:1) accounted for 86% of observed bigrams, whereas a large and eclectic group of ratios (e.g., 7:3, 11:2) accounted for fewer than 1%. The distribution is thus consistent with earlier findings that rhythmic patterns with simple integer ratios appear to be universal (*123*). The full lists of bigrams, with their cumulative frequencies, are in tables S28 and S29.

These results suggest that power-law distributions in music are a human universal (at least in the four genres studied here), with songs dominated by small melodic intervals and simple rhythmic ratios and enriched with many rare but larger and more complex ones. Because the specification of a power law is sensitive to sampling error in the tail of the distribution (*124*), and because many generative processes can give rise to a power-law distribution (*125*), we cannot identify a single explanation. Among the possibilities are that control of the vocal tract is biased toward small jumps in pitch that minimize effort, that auditory analysis is biased toward tracking similar sounds that are likely to be emitted by a single sound-maker, that composers tend to add notes to a melody that are similar to ones already contained in it, and that human aesthetic reactions are engaged by stimuli that are power law–distributed, which makes them neither too monotonous nor too chaotic (*116*, *126*, *127*)—"inevitable and yet surprising," as the music of Bach has been described (*128*).

A new science of music

The challenge in understanding music has always been to reconcile its universality with its diversity. Even Longfellow, who declared music to be humanity's universal language, celebrated the many forms it could take: "The peasant of the North … sings the traditionary ballad to his children … the muleteer of Spain carols with the early lark … The vintager of Sicily has his evening hymn; the fisherman of Naples his boat-song; the gondolier of Venice his midnight serenade" (*1*). Conversely, even an ethnomusicologist skeptical of universals in music conceded that "most people make it" (*36*). Music is universal but clearly takes on different forms in different cultures. To go beyond these unexceptionable observations and understand exactly what is universal about music, while circumventing the biases inherent in opportunistic observations, we assembled databases that combine the empirical richness of the ethnographic and musicological record with the tools of computational social science.

The findings allow the following conclusions: Music exists in every society, varies more within than between societies, and has acoustic features that are systematically (albeit probabilistically) related to the behaviors of singers and listeners. At the same time, music is not a fixed biological response with a single, prototypical adaptive function such as mating, group bonding, or infant care: It varies substantially in melodic and rhythmic complexity and is produced worldwide in at least 14 behavioral contexts that vary in formality, arousal, and religiosity. But music does appear to be tied to identifiable perceptual, cognitive, and affective faculties, including language (all societies put words to their songs), motor control (people in all societies dance), auditory analysis (all musical systems have some signatures of tonality), and aesthetics (their melodies and rhythms are balanced between monotony and chaos).

Methods summary

To build the NHS Ethnography, we extracted descriptions of singing from the Probability Sample File by searching the database for text that was tagged with the topic MUSIC and that included at least one of 10 keywords that singled out vocal music (e.g., "singers," "song," "lullaby") (Text S1.1). This search yielded 4709 descriptions of singing (490,615 words) drawn from 493 documents (median 49 descriptions per society). We manually annotated each description with 66 variables to comprehensively capture the behaviors reported by ethnographers (e.g., age of the singer, duration of the song). We also attached metadata about each paragraph (e.g., document publication data; tagged nonmusical topics) using a matching algorithm that located the source paragraphs from which the description of the song was extracted. See Text S1.1 for full details on corpus construction, tables S1 to S6 for annotation types, and table S12 for a list of societies and locations.

Song events from all the societies were aggregated into a single dataset, without indicators of the society they came from. The range of possible missing values was filled in using a Markov chain Monte Carlo procedure that assumes that their absence reflects conditionally random omission with probabilities related to the features that the ethnographer did record, such as the age and sex of the singer or the size of the audience (Text S2.1). For the dimensionality reduction, we used an optimal singular value thresholding criterion (*129*) to determine the number of dimensions to analyze, which we then interpreted by three techniques: examining annotations that load highly on each dimension; searching for examples at extreme locations in the space and examining their content; and testing whether known song types formed distinct clusters in the latent space (e.g., dance songs versus healing songs; see Fig. 2).

To build the NHS Discography, and to ensure that the sample of recordings from each genre is representative of human societies, we located field recordings of dance songs, lullabies, healing songs, and love songs using a geographic stratification approach similar to that of the NHS Ethnography—namely, by drawing one recording representing each behavioral context from each of 30 regions. We chose songs according to predetermined criteria (table S21), studying recordings' liner notes and the supporting ethnographic text without listening to the recordings. When more than one suitable recording was available, we selected one at random. See Text S1.1 for details on corpus construction, tables S1 and S7 to S11 for annotation types, and table S22 for a list of societies and locations.

For analyses of the universality of musical forms, we studied each of the four representations individually (machine summaries, naïve listener ratings, expert listener ratings, and features extracted from manual transcriptions), along with a combination of the expert listener and manual transcription data, which excluded many "contextual" features of the audio recordings (e.g., the sound of an infant crying during a lullaby). For the explorations of the structure of musical forms, we studied the manual transcriptions of songs and also used the Bayesian principal components analysis technique (described above) on the combined expert + transcription data summarizing NHS Discography songs.

Both the NHS Ethnography and NHS Discography can be explored interactively at http://themusiclab.org/nhsplots.

REFERENCES AND NOTES

1. H. W. Longfellow, *Outre-mer: A Pilgrimage Beyond the Sea* (Harper, 1835).
2. L. Bernstein, *The Unanswered Question: Six Talks at Harvard* (Harvard Univ. Press, 2002).
3. H. Honing, C. ten Cate, I. Peretz, S. E. Trehub, Without it no music: Cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. B* **370**, 20140088 (2015). doi: 10.1098/rstb.2014.0088; pmid: 25646511
4. S. A. Mehr, M. M. Krasnow, Parent-offspring conflict and the evolution of infant-directed song. *Evol. Hum. Behav.* **38**, 674–684 (2017). doi: 10.1016/j.evolhumbehav.2016.12.005
5. E. H. Hagen, G. A. Bryant, Music and dance as a coalition signaling system. *Hum. Nat.* **14**, 21–51 (2003). doi: 10.1007/s12110-003-1015-z; pmid: 26189987
6. A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, 1990).

Downloaded from http://science.sciencemag.org/ on November 28, 2019

7. A. S. Bregman, S. Pinker, Auditory streaming and the building of timbre. *Can. J. Psychol.* **32**, 19–31 (1978). doi: 10.1037/h0081664; pmid: 728845
8. S. Pinker, *How the Mind Works* (Norton, 1997).
9. L. J. Trainor, The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philos. Trans. R. Soc. B* **370**, 20140089 (2015). doi: 10.1098/rstb.2014.0089; pmid: 25646512
10. A. Lomax, *Folk Song Style and Culture* (American Association for the Advancement of Science, 1968).
11. A. P. Merriam, *The Anthropology of Music* (Northwestern Univ. Press, 1964).
12. B. Nettl, *The Study of Ethnomusicology: Thirty-Three Discussions* (Univ. of Illinois Press, 2015).
13. N. J. Conard, M. Malina, S. C. Münzel, New flutes document the earliest musical tradition in southwestern Germany. *Nature* **460**, 737–740 (2009). doi: 10.1038/nature08169; pmid: 19553935
14. N. Martínez-Molina, E. Mas-Herrero, A. Rodríguez-Fornells, R. J. Zatorre, J. Marco-Pallarés, Neural correlates of specific musical anhedonia. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7337–E7345 (2016). doi: 10.1073/pnas.1611211113; pmid: 27799544
15. A. D. Patel, Language, music, syntax and the brain. *Nat. Neurosci.* **6**, 674–681 (2003). doi: 10.1038/nn1082; pmid: 12830158
16. D. Perani *et al*., Functional specializations for music processing in the human newborn brain. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4758–4763 (2010). doi: 10.1073/pnas.0909074107; pmid: 20176953
17. J. H. McDermott, A. F. Schultz, E. A. Undurraga, R. A. Godoy, Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* **535**, 547–550 (2016). doi: 10.1038/nature18635; pmid: 27409816
18. B. Nettl, *in The Origins of Music* (MIT Press, 2000), pp. 463–472.
19. J. Blacking, Can musical universals be heard? *World Music* **19**, 14–22 (1977).
20. F. Harrison, Universals in music: Towards a methodology of comparative research. *World Music* **19**, 30–36 (1977).
21. G. Herzog, Music's dialects: A non-universal language. *Indep. J. Columbia Univ.* **6**(10), 1–2 (1939).
22. M. Hood, *The Ethnomusicologist* (McGraw-Hill, 1971).
23. L. B. Meyer, Universalism and relativism in the study of ethnic music. *Ethnomusicology* **4**, 49–54 (1960). doi: 10.2307/924262
24. S. Feld, Sound structure as social structure. *Ethnomusicology* **28**, 383–409 (1984). doi: 10.2307/851232
25. M. Hood, in *Musicology*, F. L. Harrison, M. Hood, C. V. Palisca, Eds. (Prentice-Hall, 1963), pp. 217–239.
26. M. Roseman, The social structuring of sound: The Temiar of peninsular Malaysia. *Ethnomusicology* **28**, 411–445 (1984). doi: 10.2307/851233
27. S. Feld, *Sound and Sentiment: Birds, Weeping, Poetics, and Song in Kaluli Expression* (Duke Univ. Press, 2012).
28. N. Harkness, *Songs of Seoul: An Ethnography of Voice and Voicing in Christian South Korea* (Univ. of California Press, 2014).
29. T. Rose, Orality and technology: Rap music and Afro-American cultural resistance. *Pop. Music Soc.* **13**, 35–44 (1989). doi: 10.1080/03007768908591371
30. S. Feld, A. A. Fox, Music and language. *Annu. Rev. Anthropol.* **23**, 25–53 (1994). doi: 10.1146/annurev.an.23.100194.000325
31. T. Ellingson, in *Ethnomusicology*, H. Myers, Ed. (Norton, 1992), pp. 110–152.
32. T. F. Johnston, The cultural role of Tsonga beer-drink music. *Yearb. Int. Folk Music Counc.* **5**, 132–155 (1973). doi: 10.2307/767499
33. A. Rehding, The quest for the origins of music in Germany circa 1900. *J. Am. Musicol. Soc.* **53**, 345–385 (2000). doi: 10.2307/832011
34. A. K. Rasmussen, Response to "Form and function in human song". *Soc. Ethnomusicol. Newsl.* **52**, 7 (2018).
35. We conducted a survey of academics to solicit opinions about the universality of music. The overall pattern of results from music scholars was consistent with List's claim that music is characterized by very few universals. For instance, in response to the question "Do you think that music is mostly shaped by culture, or do you think that music is mostly shaped by a universal human nature?", the majority of music scholars responded in the "Music is mostly shaped by culture" half of the scale (ethnomusicologists, 71%; music theorists, 68%; other musical disciplines, 62%). See Text S1.4.1 for full details.
36. G. List, On the non-universality of musical perspectives. *Ethnomusicology* **15**, 399–402 (1971). doi: 10.2307/850640
37. N. A. Chomsky, *Language and Mind* (Harcourt Brace, 1968).
38. M. H. Christiansen, C. T. Collins, S. Edelman, *Language Universals* (Oxford Univ. Press, 2009).
39. P. Boyer, *Religion Explained: The Evolutionary Origins of Religious Thought* (Basic Books, 2007).
40. M. Singh, The cultural evolution of shamanism. *Behav. Brain Sci.* **41**, e66 (2018). doi: 10.1017/S0140525X17001893; pmid: 28679454
41. R. Sosis, C. Alcorta, Signaling, solidarity, and the sacred: The evolution of religious behavior. *Evol. Anthropol. Issues News Rev.* **12**, 264–274 (2003). doi: 10.1002/evan.10120
42. D. M. Buss, Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behav. Brain Sci.* **12**, 1–14 (1989). doi: 10.1017/S0140525X00023992
43. B. Chapais, Complex kinship patterns as evolutionary constructions, and the origins of sociocultural universals. *Curr. Anthropol.* **55**, 751–783 (2014). doi: 10.1086/678972
44. A. P. Fiske, *Structures of Social Life: The Four Elementary Forms of Human Relations: Communal Sharing, Authority Ranking, Equality Matching, Market Pricing* (Free Press, 1991).
45. T. S. Rai, A. P. Fiske, Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychol. Rev.* **118**, 57–75 (2011). doi: 10.1037/a0021867; pmid: 21244187
46. O. S. Curry, D. A. Mullins, H. Whitehouse, Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Curr. Anthropol.* **60**, 47–69 (2019). doi: 10.1086/701478
47. J. Haidt, *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (Penguin, 2013).
48. R. W. Wrangham, L. Glowacki, Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers: Evaluating the chimpanzee model. *Hum. Nat.* **23**, 5–29 (2012). doi: 10.1007/s12110-012-9132-1; pmid: 22388773
49. S. Pinker, *The Better Angels of Our Nature: Why Violence Has Declined* (Viking, 2011).
50. A. P. Fiske, T. S. Rai, *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships* (Cambridge Univ. Press, 2015).
51. L. Aarøe, M. B. Petersen, K. Arceneaux, The behavioral immune system shapes political intuitions: Why and how individual differences in disgust sensitivity underlie opposition to immigration. *Am. Polit. Sci. Rev.* **111**, 277–294 (2017). doi: 10.1017/S0003055416000770
52. P. Boyer, M. B. Petersen, Folk-economic beliefs: An evolutionary cognitive model. *Behav. Brain Sci.* **41**, e158 (2018). doi: 10.1017/S0140525X17001960; pmid: 29022516
53. P. E. Savage, S. Brown, E. Sakai, T. E. Currie, Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8987–8992 (2015). doi: 10.1073/pnas.1414495112; pmid: 26124105
54. S. A. Mehr, M. Singh, H. York, L. Glowacki, M. M. Krasnow, Form and function in human song. *Curr. Biol.* **28**, 356–368.e5 (2018). doi: 10.1016/j.cub.2017.12.042; pmid: 29395919
55. T. Fritz *et al*., Universal recognition of three basic emotions in music. *Curr. Biol.* **19**, 573–576 (2009). doi: 10.1016/j.cub.2009.02.058; pmid: 19303300
56. B. Sievers, L. Polansky, M. Casey, T. Wheatley, Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 70–75 (2013). doi: 10.1073/pnas.1209023110; pmid: 23248314
57. W. T. Fitch, The biology and evolution of music: A comparative perspective. *Cognition* **100**, 173–215 (2006). doi: 10.1016/j.cognition.2005.11.009; pmid: 16412411
58. A. Lomax, Universals in song. *World Music* **19**, 117–129 (1977).
59. D. E. Brown, *Human Universals* (Temple Univ. Press, 1991).
60. S. Brown, J. Jordania, Universals in the world's musics. *Psychol. Music* **41**, 229–248 (2013). doi: 10.1177/0305735611425896
61. Human Relations Area Files Inc., eHRAF World Cultures Database; http://ehrafworldcultures.yale.edu/.
62. G. P. Murdock, C. S. Ford, A. E. Hudson, R. Kennedy, L. W. Simmons, J. W. M. Whiting, *Outline of Cultural Materials* (Human Relations Area Files Inc., New Haven, CT, 2008).
63. P. Austerlitz, *Merenge: Dominican Music and Dominican Identity* (Temple Univ. Press, 2007).
64. C. Irgens-Møller, *Music of the Hazara: An Investigation of the Field Recordings of Klaus Ferdinand 1954–1955* (Moesgård Museum, Denmark, 2007).
65. B. D. Koen, *Devotional Music and Healing in Badakhshan, Tajikistan: Preventive and Curative Practices* (UMI Dissertation Services, Ann Arbor, MI, 2005).
66. B. D. Koen, *Beyond the Roof of the World: Music, Prayer, and Healing in the Pamir Mountains* (Oxford Univ. Press, 2011).
67. A. Youssefzadeh, The situation of music in Iran since the revolution: The role of official organizations. *Br. J. Ethnomusicol.* **9**, 35–61 (2000). doi: 10.1080/09681220008567300
68. S. Zeranska-Kominek, The classification of repertoire in Turkmen traditional music. *Asian Music* **21**, 91–109 (1990). doi: 10.2307/834113
69. A. D. Patel, *Music, Language, and the Brain* (Oxford Univ. Press, 2008).
70. D. P. McAllester, Some thoughts on "universals" in world music. *Ethnomusicology* **15**, 379–380 (1971). doi: 10.2307/850637
71. A. P. Merriam, in *Cross-Cultural Perspectives on Music: Essays in Memory of Miczyslaw Kolinski*, R. Falck, T. Rice, M. Kolinski, Eds. (Univ. of Toronto Press, 1982), pp. 174–189.
72. D. L. Harwood, Universals in music: A perspective from cognitive psychology. *Ethnomusicology* **20**, 521–533 (1976). doi: 10.2307/851047
73. F. Lerdahl, R. Jackendoff, *A Generative Theory of Tonal Music* (MIT Press, 1983).
74. Human Relations Area Files Inc., The HRAF quality control sample universe. *Behav. Sci. Notes* **2**, 81–88 (1967). doi: 10.1177/106939716700200203
75. R. O. Lagacé, The HRAF probability sample: Retrospect and prospect. *Behav. Sci. Res.* **14**, 211–229 (1979). doi: 10.1177/106939716700200202
76. R. Naroll, The proposed HRAF probability sample. *Behav. Sci. Notes* **2**, 70–80 (1967). doi: 10.1177/106939716700200202
77. B. S. Hewlett, S. Winn, Allomaternal nursing in humans. *Curr. Anthropol.* **55**, 200–219 (2014). doi: 10.1086/675657; pmid: 24991682
78. Q. D. Atkinson, H. Whitehouse, The cultural morphospace of ritual form. *Evol. Hum. Behav.* **32**, 50–62 (2011). doi: 10.1016/j.evolhumbehav.2010.09.002
79. C. R. Ember, The relative decline in women's contribution to agriculture with intensification. *Am. Anthropol.* **85**, 285–304 (1983). doi: 10.1525/aa.1983.85.2.02a00020
80. D. M. T. Fessler, A. C. Pisor, C. D. Navarrete, Negatively-biased credulity and the cultural evolution of beliefs. *PLOS ONE* **9**, e95167 (2014). doi: 10.1371/journal.pone.0095167; pmid: 24736596
81. B. R. Huber, W. L. Breedlove, Evolutionary theory, kinship, and childbirth in cross-cultural perspective. *Cross-Cultural Res.* **41**, 196–219 (2007). doi: 10.1177/1069397106298261
82. D. Levinson, Physical punishment of children and wifebeating in cross-cultural perspective. *Child Abuse Negl.* **5**, 193–195 (1981). doi: 10.1016/0145-2134(81)90040-5
83. M. Singh, Magic, explanations, and evil: On the origins and design of witches and sorcerers. SocArXiv (2019). doi: 10.31235/osf.io/pbwc7
84. M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis. *J. R. Stat. Soc. B* **61**, 611–622 (1999). doi: 10.1111/1467-9868.00196
85. R. C. Lewontin, in *Evolutionary Biology*, T. Dobzhansky, M. K. Hecht, W. C. Steer, Eds. (Appleton-Century-Crofts, 1972), pp. 391–398.
86. T. Rzeszutek, P. E. Savage, S. Brown, The structure of cross-cultural musical diversity. *Proc. R. Soc. B* **279**, 1606–1612 (2012). doi: 10.1098/rspb.2011.1750; pmid: 22072606
87. Princeton University, WordNet: A lexical database for English (2010); http://wordnet.princeton.edu.
88. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
89. R. Karsten, *The Religion of the Samek: Ancient Beliefs and Cults of the Scandinavian and Finnish Lapps* (Brill, 1955).
90. M. J. Murphy, "A sprain." In Manuscript 1503 (serial #QBA7800250), pp. 237–238 (1957). Available from the Manuscript Collection of the National Folklore Collection, University College Dublin.
91. S. Feld, Linguistic models in ethnomusicology. *Ethnomusicology* **18**, 197–217 (1974). doi: 10.2307/850579
92. S. Arom, *African Polyphony and Polyrhythm: Musical Structure and Methodology* (Cambridge Univ. Press, 2004).

Downloaded from http://science.sciencemag.org/ on November 28, 2019

93. B. Nettl, *Theory and Method in Ethnomusicology* (Collier-Macmillan, 1964).
94. J. Friedman, T. Hastie, R. Tibshirani, Lasso and Elastic-Net Regularized Generalized Linear Models. Rpackage Version 2.0-5 (2016).
95. C. Nadeau, Y. Bengio, Inference for the generalization error. *Mach. Learn.* **52**, 239–281 (2003). doi: 10.1023/A:1024068626366
96. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996). doi: 10.1111/j.2517-6161.1996.tb02080.x
97. S. E. Trehub, A. M. Unyk, L. J. Trainor, Adults identify infant-directed music across cultures. *Infant Behav. Dev.* **16**, 193–211 (1993). doi: 10.1016/0163-6383(93)80017-3
98. A. Barker, *Greek Musical Writings: Harmonic and Acoustic Theory* (Cambridge Univ. Press, 2004).
99. C. L. Krumhansl, The Cognition of Tonality – as We Know it Today. *J. New Music Res.* **33**, 253–268 (2004). doi: 10.1080/0929821042000317831
100. J. H. McDermott, A. J. Oxenham, Music perception, pitch, and the auditory system. *Curr. Opin. Neurobiol.* **18**, 452–463 (2008). doi: 10.1016/j.conb.2008.09.005; pmid: 18824100
101. R. Jackendoff, F. Lerdahl, The capacity for music: What is it, and what's special about it? *Cognition* **100**, 33–72 (2006). doi: 10.1016/j.cognition.2005.11.005; pmid: 16384553
102. M. A. Castellano, J. J. Bharucha, C. L. Krumhansl, Tonal hierarchies in the music of north India. *J. Exp. Psychol. Gen.* **113**, 394–412 (1984). doi: 10.1037/0096-3445.113.3.394; pmid: 6237169
103. C. L. Krumhansl *et al*., Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition* **76**, 13–58 (2000). doi: 10.1016/S0010-0277(00)00068-8; pmid: 10822042
104. H. von Helmholtz, *The Sensations of Tone as a Physiological Basis for the Theory of Music* (Longmans, 1885).
105. D. Cooke, *The Language of Music* (Oxford Univ. Press, 2001).
106. J. A. Hartigan, P. M. Hartigan, The Dip Test of Unimodality. *Ann. Stat.* **13**, 70–84 (1985). doi: 10.1214/aos/1176346577
107. C. L. Krumhansl, *Cognitive Foundations of Musical Pitch* (Oxford Univ. Press, 2001).
108. G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Addison-Wesley, 1949).
109. H. Baayen, *Word Frequency Distributions* (Kluwer Academic, 2001).
110. S. T. Piantadosi, Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **21**, 1112–1130 (2014). doi: 10.3758/s13423-014-0585-6; pmid: 24664880
111. D. Kim, S.-W. Son, H. Jeong, Large-scale quantitative analysis of painting arts. *Sci. Rep.* **4**, 7370 (2014). doi: 10.1073/pnas.87.3.938; pmid: 11607061
112. K. J. Hsü, A. J. Hsü, Fractal geometry of music. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 938–941 (1990). doi: 10.1073/pnas.87.3.938; pmid: 11607061
113. D. H. Zanette, Zipf's law and the creation of musical context. *Music. Sci.* **10**, 3–18 (2006). doi: 10.1177/102986490601000101
114. H. J. Brothers, Intervallic scaling in the Bach cello suites. *Fractals* **17**, 537–545 (2009). doi: 10.1142/S0218348X09004521
115. L. Liu, J. Wei, H. Zhang, J. Xin, J. Huang, A statistical physics view of pitch fluctuations in the classical music from Bach to Chopin: Evidence for scaling. *PLOS ONE* **8**, e58710 (2013). doi: 10.1371/journal.pone.0058710; pmid: 23544047
116. B. Manaris *et al*., Zipf's law, music classification, and aesthetics. *Comput. Music J.* **29**, 55–69 (2005). doi: 10.1162/comj.2005.29.1.55
117. R. F. Voss, J. Clarke, '1/*f* noise' in music and speech. *Nature* **258**, 317–318 (1975). doi: 10.1038/258317a0
118. M. Rohrmeier, I. Cross, Statistical Properties of Tonal Harmony in Bach's Chorales. in *Proceedings of the 10th International Conference on Music Perception and Cognition* (2008), p. 9.
119. F. C. Moss, M. Neuwirth, D. Harasim, M. Rohrmeier, Statistical characteristics of tonal harmony: A corpus study of Beethoven's string quartets. *PLOS ONE* **14**, e0217242 (2019). doi: 10.1371/journal.pone.0217242; pmid: 31170188
120. M. Beltrán del Río, G. Cocho, G. G. Naumis, Universality in the tail of musical note rank distribution. *Physica A* **387**, 5552–5560 (2008). doi: 10.1016/j.physa.2008.05.031
121. D. J. Levitin, P. Chordia, V. Menon, Musical rhythm spectra from Bach to Joplin obey a 1/f power law. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3716–3720 (2012). doi: 10.1073/pnas.1113828109; pmid: 22355125
122. T. Van Khe, Is the pentatonic universal? A few reflections on pentatonism. *World Music* **19**, 76–84 (1977).
123. N. Jacoby, J. H. McDermott, Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* **27**, 359–370 (2017). doi: 10.1016/j.cub.2016.12.031; pmid: 28065607
124. A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-Law Distributions in Empirical Data. *SIAM Rev.* **51**, 661–703 (2009). doi: 10.1137/070710111
125. M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions. *Internet Math.* **1**, 226–251 (2004). doi: 10.1080/15427951.2004.10129088
126. G. D. Birkhoff, *Aesthetic Measure* (Harvard Univ. Press, 2013).
127. B. Manaris, P. Roos, D. Krehbiel, T. Zalonis, J. R. Armstrong, in *Music Data Mining*, T. Li, M. Ogihara, G. Tzanetakis, Eds.(CRC Press, 2012), chapter 6.
128. M. R. Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (Dover, 2009).
129. D. Donoho, M. Gavish, Minimax risk of matrix denoising by singular value thresholding. *Ann. Stat.* **42**, 2413–2440 (2014). doi: 10.1214/14-AOS1257

ACKNOWLEDGMENTS

We thank the hundreds of anthropologists and ethnomusicologists whose work forms the source material for all our analyses; the countless people whose music those scholars reported on; and the research assistants who contributed to the creation of the Natural History of Song corpora and to this research, here listed alphabetically: Z. Ahmad, P. Ammirante, R. Beaudoin, J. Bellissimo, A. Bergson, M. Bertolo, M. Bertuccelli, A. Bitran, S. Bourdaghs, J. Brown, L. Chen, C. Colletti, L. Crowe, K. Czachorowski, L. Dinetan, K. Emery, D. Fratina, E. Galm, S. Gomez, Y.-H. Hung, C. Jones, S. Joseph, J. Kangatharan, A. Keomurjian, H. J. Kim, S. Lakin, M. Laroussini, T. Lee, H. Lee-Rubin, C. Leff, K. Lopez, K. Luk, E. Lustig, V. Malawey, C. McMann, M. Montagnese, P. Moro, N. Okwelogu, T. Ozawa, C. Palfy, J. Palmer, A. Paz, L. Poeppel, A. Ratajska, E. Regan, A. Reid, R. Sagar, P. Savage, G. Shank, S. Sharp, E. Sierra, D. Tamaroff, I. Tan, C. Tripoli, K. Tutrone, A. Wang, M. Weigel, J. Weiner, R. Weissman, A. Xiao, F. Xing, K. Yong, H. York, and J. Youngers. We also thank C. Ember and M. Fischer for providing additional data from the Human Relations Area Files, and for their assistance using those data; S. Adams, P. Laurence, P. O'Brien, A. Wilson, the staff at the Archive of World Music at Loeb Music Library (Harvard University), and M. Graf and the staff at the Archives of Traditional Music (Indiana University) for assistance with locating and digitizing audio recordings; B. Hillers for assistance with information concerning traditional Gaelic music; D. Niles, S. Wadley, and H. Wild for contributing recordings from their personal collections; S. Collins for producing the NHS Ethnography validity annotations; M. Walter for assistance with digital processing of transcriptions; J. Hulbert and R. Clarida for assistance with copyright issues and materials sharing; V. Kuchinov for developing the interactive visualizations; S. Deviche for contributing illustrations; and the Dana Foundation, whose program "Arts and Cognition" led in part to the development of this research. Last, we thank A. Rehding, G. Bryant, E. Hagen, H. Gardner, E. Spelke, M. Tenzer, G. King, J. Nemirow, J. Kagan, and A. Martin for their feedback, ideas, and intellectual support of this work. **Funding:** Supported by the Harvard Data Science Initiative (S.A.M.); the National Institutes of Health Director's Early Independence Award DP5OD024566 (S.A.M.); the Harvard Graduate School of Education/Harvard University Presidential Scholarship (S.A.M.); the Harvard University Department of Psychology (S.A.M. and M.M.K.); a Harvard University Mind/Brain/Behavior Interfaculty Initiative Graduate Student Award (S.A.M. and M.S.); the National Science Foundation Graduate Research Fellowship Program (M.S.); the Microsoft Research postdoctoral fellowship program (D.K.); the Washington University Faculty of Arts and Sciences Dean's Office (C.L.); the Columbia University Center for Science and Society (N.J.); the Natural Sciences and Engineering Research Council of Canada (T.J.O.); Fonds de Recherche du Québec Société et Culture (T.J.O.); and ANR Labex IAST (L.G.). **Author contributions:** S.A.M., M.S., and L.G. created and direct the Natural History of Song project; they oversaw all aspects of this work, including the design and development of the corpora. S.P., M.M.K., and T.J.O. contributed to the conceptual foundation. D.K. designed and implemented all analyses, with support from S.A.M. and C.L. S.A.M., D.K., and M.S. designed the static figures and S.A.M. and D.K. created them. C.L. and S.A.M. designed the interactive figures and supervised their development. S.A.M. recruited and managed all staff, who collected, annotated, processed, and corrected data and metadata. S.A.M., D.M.K., and D.P.-J. transcribed the NHS Discography into music notation. S.A., A.A.E., E.J.H., and R.M.H. provided key support by contributing to annotations, background research, and project management. S.A.M., J.K.H., M.V.J., J.S., and C.M.B. designed and implemented the online experiment at http://themusiclab.org. N.J. assisted with web scraping, music information retrieval, and initial analyses. S.A.M., M.S., and L.G. designed the overall structure of the manuscript; S.A.M., M.S., and S.P. led the writing; and all authors edited it collaboratively. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All Natural History of Song data and materials are publicly archived at http://osf.io/jmv3q, with the exception of the full audio recordings in the NHS Discography, which are available via the Harvard Dataverse at https://doi.org/10.7910/DVN/SESAO1. All analysis scripts are available at http://github.com/themusiclab/nhs. Human Relations Area Files data and the eHRAF World Cultures database are available via licensing agreement at http://ehrafworldcultures.yale.edu; the document- and paragraph-wise word histograms from the Probability Sample File were provided by the Human Relations Area Files under a Data Use Agreement. The Global Summary of the Year corpus is maintained by the National Oceanic and Atmospheric Administration, U.S. Department of Commerce, and is publicly available at www.ncei.noaa.gov/data/gsoy/.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/eaax0868/suppl/DC1
Supplementary Text
Figs. S1 to S15
Tables S1 to S37
References (*130*–*147*)

1 March 2019; accepted 24 October 2019
10.1126/science.aax0868

Downloaded from http://science.sciencemag.org/ on November 28, 2019

RESEARCH ARTICLE

SIGNAL TRANSDUCTION

Structural mechanism of a Rag GTPase activation checkpoint by the lysosomal folliculin complex

Rosalie E. Lawrence[1,2]*, Simon A. Fromm[1]*, Yangxue Fu[1], Adam L. Yokom[1], Do Jin Kim[1], Ashley M. Thelen[1,2], Lindsey N. Young[1], Chun-Yan Lim[1,2], Avi J. Samelson[2,3], James H. Hurley[1,4,5]†, Roberto Zoncu[1,2,4]†

The tumor suppressor folliculin (FLCN) enables nutrient-dependent activation of the mechanistic target of rapamycin complex 1 (mTORC1) protein kinase via its guanosine triphosphatase (GTPase) activating protein (GAP) activity toward the GTPase RagC. Concomitant with mTORC1 inactivation by starvation, FLCN relocalizes from the cytosol to lysosomes. To determine the lysosomal function of FLCN, we reconstituted the human lysosomal FLCN complex (LFC) containing FLCN, its partner FLCN-interacting protein 2 (FNIP2), and the $RagA^{GDP}$:$RagC^{GTP}$ GTPases as they exist in the starved state with their lysosomal anchor Ragulator complex and determined its cryo–electron microscopy structure to 3.6 angstroms. The RagC-GAP activity of FLCN was inhibited within the LFC, owing to displacement of a catalytically required arginine in FLCN from the RagC nucleotide. Disassembly of the LFC and release of the RagC-GAP activity of FLCN enabled mTORC1-dependent regulation of the master regulator of lysosomal biogenesis, transcription factor E3, implicating the LFC as a checkpoint in mTORC1 signaling.

The mechanistic target of rapamycin complex 1 (mTORC1) kinase is at the center of a complex biochemical network that relays information about the availability of nutrients, growth factors, oxygen, and energy to downstream signaling programs. When cells are adequately nourished, these signaling programs drive growth and metabolism and inhibit catabolic pathways such as autophagy (*1*, *2*). Nutrients, including amino acids, glucose, and cholesterol, activate mTORC1 by triggering its translocation from the cytosol to the lysosomal membrane (*3*–*6*). This process is regulated by the heterodimeric Rag guanosine triphosphatases (GTPases) (*3*, *4*, *7*), which localize to the lysosomal membrane via binding to the pentameric Ragulator complex (*4*, *8*, *9*). Rag heterodimers consist of two functionally equivalent pairs: RagA or RagB in complex with RagC or RagD. Nutrients trigger the transition from the "inactive" combination of guanosine diphosphate (GDP)–bound Rag A and guanosine triphosphate (GTP)–bound Rag C ($RagA^{GDP}$:$RagC^{GTP}$) to the "active" $RagA^{GTP}$:$RagC^{GDP}$ state. The active Rag heterodimer binds directly to mTORC1 and recruits it to the lysosome, enabling its subsequent activation (*3*, *4*, *7*, *10*, *11*).

Two GTPase-activating protein (GAP) complexes mediate, in part, the conversion between active and inactive Rag GTPase states. When amino acid concentrations are low, GATOR1 promotes GTP hydrolysis by RagA or RagB in response to a well-characterized upstream nutrient-sensing pathway (*12*–*14*). Conversely, nucleotide hydrolysis on RagC or RagD is stimulated by folliculin (FLCN), the tumor suppressor responsible for Birt-Hogg-Dubé syndrome (BHD) (*15*), in complex with FLCN-interacting protein 1 or 2 (FNIP1 or FNIP2) (*16*–*18*). Consistent with its role in Rag GTPase activation, when FLCN is acutely depleted in cell culture, activation of mTORC1 signaling by amino acids is impaired (*18*, *19*). Yet in cellular and mouse models of BHD, the effects of FLCN loss on mTORC1 signaling range from neutral to activating, leading to a controversial and context-dependent interpretation of FLCN function (*16*, *20*, *21*). A further conundrum concerns the subcellular localization of FLCN:FNIP. FLCN has been reported to localize to the lysosomal membrane in low-nutrient conditions and to relocalize to the cytosol as nutrient levels increase (*18*, *19*, *22*). How these changes in FLCN localization relate to regulation of its GAP activity is not understood. Structural details of FLCN interactions with its partner FNIP, how FLCN:FNIP physically interacts with Rag GTPases, and the key residues for the GAP activity of FLCN:FNIP are unknown.

Assembly of a stable lysosomal FLCN complex lacking GAP activity

Consistent with previous reports (*18*, *19*, *22*), FLCN clustered on LAMP2-positive lysosomes in amino acid–depleted conditions and became dispersed in the cytosol upon refeeding (fig. S1, A and B). FLCN relocalization was controlled by the nucleotide state of the Rag GTPases. Expressing Rag mutants locked in the inactive nucleotide binding state caused FLCN to be constitutively bound to lysosomes irrespective of amino acid conditions. Conversely, the active Rag GTPase mutants caused FLCN to become cytosolic under all conditions (fig. S1, A and B).

To determine the function of lysosomal FLCN in starvation, we monitored the FLCN:FNIP2 GAP reaction in RagC GTPase assays in vitro. To selectively generate Rag heterodimers containing one diphosphate and one triphosphate nucleotide, we used the mutant $RagC^{D181N}$ (D181N: Asp^{181}→Asn), which specifically binds xanthine nucleotides (*8*, *18*) (henceforward referred to simply as RagC). $RagA^{GTP}$:$RagC^{XTP}$ was incubated with a catalytic amount of the appropriate GAP complex (FLCN:FNIP2 or GATOR1) to generate GTP:xanthosine diphosphate (XDP)– or GDP:xanthosine triphosphate (XTP)–bound heterodimers that were confirmed by high-performance liquid chromatography (HPLC) (Fig. 1A). An intrinsic tryptophan fluorescence GTPase assay (*23*) was used to monitor the FLCN:FNIP2-catalyzed RagC-GTPase activity in real time (fig. S2, A and B). FLCN:FNIP2 was added to Rag heterodimers containing a GDP- or GTP-bound RagA, with or without the addition of stoichiometric amounts of Ragulator. We observed similar FLCN:FNIP2 GAP activity for all tested substrate permutations, except that the $RagA^{GDP}$:$RagC^{XTP}$-Ragulator combination greatly reduced the GAP activity of FLCN:FNIP2 (Fig. 1B). An HPLC-based GAP assay confirmed reduction of FLCN:FNIP2 GAP activity in the presence of GDP-bound RagA and Ragulator (Fig. 1C).

Rag heterodimers containing GDP-bound RagA form a tight complex with Ragulator at the lysosomal membrane (*24*). To determine whether $RagA^{GDP}$:$RagC^{XTP}$ and Ragulator form a larger complex that includes FLCN:FNIP2, we combined all three subcomplexes (fig. S3, A to C) and subjected the mixture to size exclusion chromatography (SEC). We observed an early eluting peak when all components were combined, and we confirmed that FLCN:FNIP2, RagA:RagC, and Ragulator coeluted in this fraction (Fig. 1, D and E). Thus, FLCN:FNIP2 forms a stable complex with $RagA^{GDP}$:$RagC^{GTP}$-Ragulator, explaining the localization of FLCN on the lysosomal membrane and suggesting that the FLCN GAP activity is inhibited during starvation. Because this complex corresponds to all properties expected of the lysosomal form of FLCN, we designate it the lysosomal FLCN complex (LFC).

Structure and nucleotide-state dependence of the LFC

We determined the structure of the LFC at a resolution of 3.6 Å by cryo–electron microscopy

[1]Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA 94720, USA. [2]The Paul F. Glenn Center for Aging Research at the University of California, Berkeley, Berkeley, CA 94720, USA. [3]Institute for Neurodegenerative Diseases, University of California at San Francisco, San Francisco, CA 94158, USA. [4]California Institute for Quantitative Biosciences, University of California, Berkeley, Berkeley, CA 94720, USA. [5]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: jimhurley@berkeley.edu (J.H.H.); rzoncu@berkeley.edu (R.Z.)

Downloaded from http://science.sciencemag.org/ on November 28, 2019

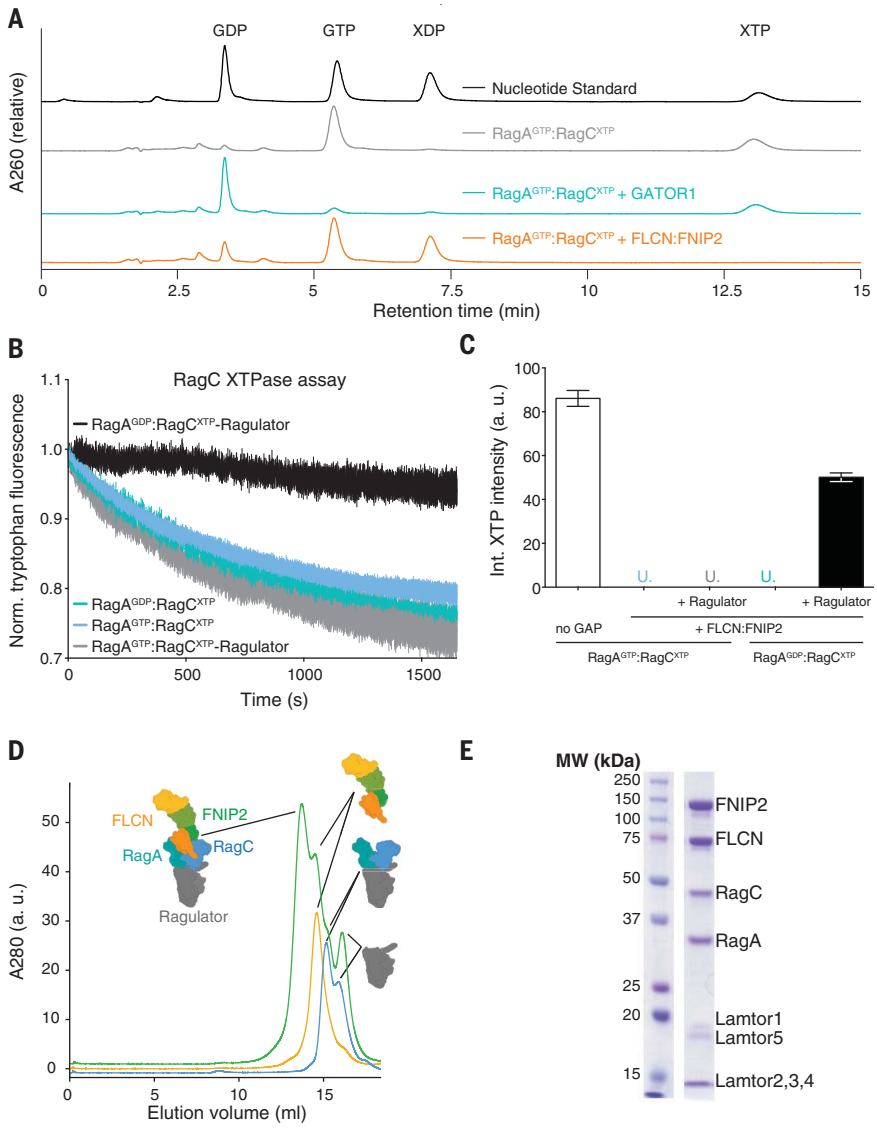


Fig. 1. Absence of FLCN GAP activity in the lysosomal folliculin complex (LFC). (**A**) Nucleotide elution profiles from HPLC analysis of Rag nucleotide binding states. Rag heterodimers were loaded with GTP and XTP to generate RagA^{GTP}:RagC^{XTP} and boiled to release bound nucleotides. The HPLC elution profile is shown (gray). RagA^{GTP}:RagC^{XTP} was incubated with FLCN:FNIP2 (orange) or Gator1 (cyan) and then analyzed by HPLC. A nucleotide standard consisting of 50 μM GDP, GTP, XDP, and XTP is shown (black). A260, absorbance at 260 nm. (**B**) FLCN:FNIP2 RagC-GAP activity visualized by tryptophan fluorescence signal decay. Rags bound to the specified combinations of nucleotides and Ragulator (black, RagA^{GDP}:RagC^{XTP}-Ragulator; cyan, RagA^{GDP}:RagC^{XTP}; blue, RagA^{GTP}:RagC^{XTP}; gray, RagA^{GTP}:RagC^{XTP}-Ragulator) were incubated with FLCN:FNIP2. Plotted are means ± SEM. n = 3 replicates. Norm., normalized. (**C**) HPLC-based GTPase assay comparing FLCN:FNIP2 RagC-GAP activity on the indicated Rag GTPase or Rag GTPase-Ragulator substrates. "No GAP" control was performed on a RagA^{GTP}:RagC^{XTP} substrate. Plotted are means ± SD. n = 3 replicates. "U." indicates that the remaining XTP signal was undetectable by HPLC. Int., intrinsic; a. u., arbitrary units. (**D**) SEC profiles of RagA^{GDP}:RagC^{XTPγS}-Ragulator (cyan), FLCN:FNIP2 (orange), and the full LFC containing RagA^{GDP}:RagC^{XTPγS}-Ragulator-FLCN:FNIP2 (green). (**E**) Coomassie stain of eluted SEC fraction containing assembled LFC. MW, molecular weight.

(cryo-EM) (Fig. 2, A to C; figs. S4 and S5; and table S1). The complex adopts an elongated shape of 240 Å in the longest dimension. The reconstructed cryo-EM density was of sufficient quality to build an atomic coordinate model (Fig. 2C, figs. S6 and S7, and table S2) constituting ~60% of the mass of the complex. The remainder of the molecular mass consists of flexible loops (Fig. 2A) with no corresponding cryo-EM density visible.

FLCN, FNIP1, and FNIP2 are DENN (differentially expressed in normal and neoplastic cells) module proteins composed of an N-terminal longin domain and a C-terminal DENN domain (*16*–*18*, *25*). The known crystal structures (*26*, *27*) of the Ragulator subcomplex and the FLCN DENN domain are positioned at opposite tips of the complex. The Rag GTPases were built on the basis of crystal and cryo-EM structures (*14*, *28*, *29*), and their location was consistent with the crystallographically resolved interaction of their C-terminal roadblock domains with Ragulator (*29*). The remaining structural elements from FLCN and FNIP2 were built on the basis of newly generated homology models (supplementary text; Fig. 2, D to F; and figs. S8 to S10).

The two longin domains of FLCN:FNIP2 form a heterodimer similar to the NPRL2:NPRL3 longin domains of GATOR1 (Fig. 2F and fig. S8C) (*14*). This longin-longin heterodimer is the only FLCN:FNIP2 element to directly interact with the Rag GTPases via their G domains. To understand how the Rag nucleotide state controls LFC formation, we tested whether FLCN:FNIP2 could form a stable complex with Ragulator and Rag GTPases in their active nucleotide binding state (RagA^{GTP}:RagC^{XDP}). Consistent with the cytoplasmic localization of FLCN under nutrient-rich conditions (*18*, *19*, *22*), FLCN:FNIP2 did not form a stable complex with active Rag-Ragulator, as evaluated by SEC (Fig. 3A). To gain structural insight into how the Rag GTPase nucleotide states control LFC assembly, we determined an ~8 to 9 Å cryo-EM structure of the 150-kD Ragulator-RagA^{GTP}:RagC^{XDP} complex (Fig. 3B and table S1). We did not detect any differences in the RagA and RagC G domain orientations between our active Rag-Ragulator structure and the previously reported GATOR1-RagA:RagC structure in a pseudoactive binding state (RagA^{GTP}:RagC^{empty}) (*14*) (Fig. 3B). Alignment of RagA and RagC in the active state with their counterparts in the inactive nucleotide state within the LFC showed that both G domains reorient relative to the roadblock domains (Fig. 3C; fig. S11, A and B; and movie S1). Similar reorientation has been observed for the RagA and RagC orthologs Gtr1 and Gtr2 from *Saccharomyces cerevisiae* (*28*, *30*). As a consequence of this reorientation, the cleft between the RagA and RagC G domains is nearly twice as wide in the inactive (RagA^{GDP}:RagC^{GTP}) state as it is in the active (RagA^{GTP}:RagC^{GDP}) state (Fig. 3D). Without this cleft opening, the FLCN:FNIP2 longin heterodimer would be sterically unable to bind RagA:RagC, as observed in the LFC (fig. S11C). This explains structurally why the LFC does not form when nutrient concentrations are high and the Rag GTPases are in the active nucleotide binding state (Fig. 3E).

Regulation of Rag nucleotide state by the LFC

Investigation of the interfaces between FLCN:FNIP2 and Rag G domains revealed that FLCN:FNIP2 interacts more extensively with RagA than with RagC. This is exemplified by its direct interaction with the clearly visible GDP

Downloaded from http://science.sciencemag.org/ on November 28, 2019

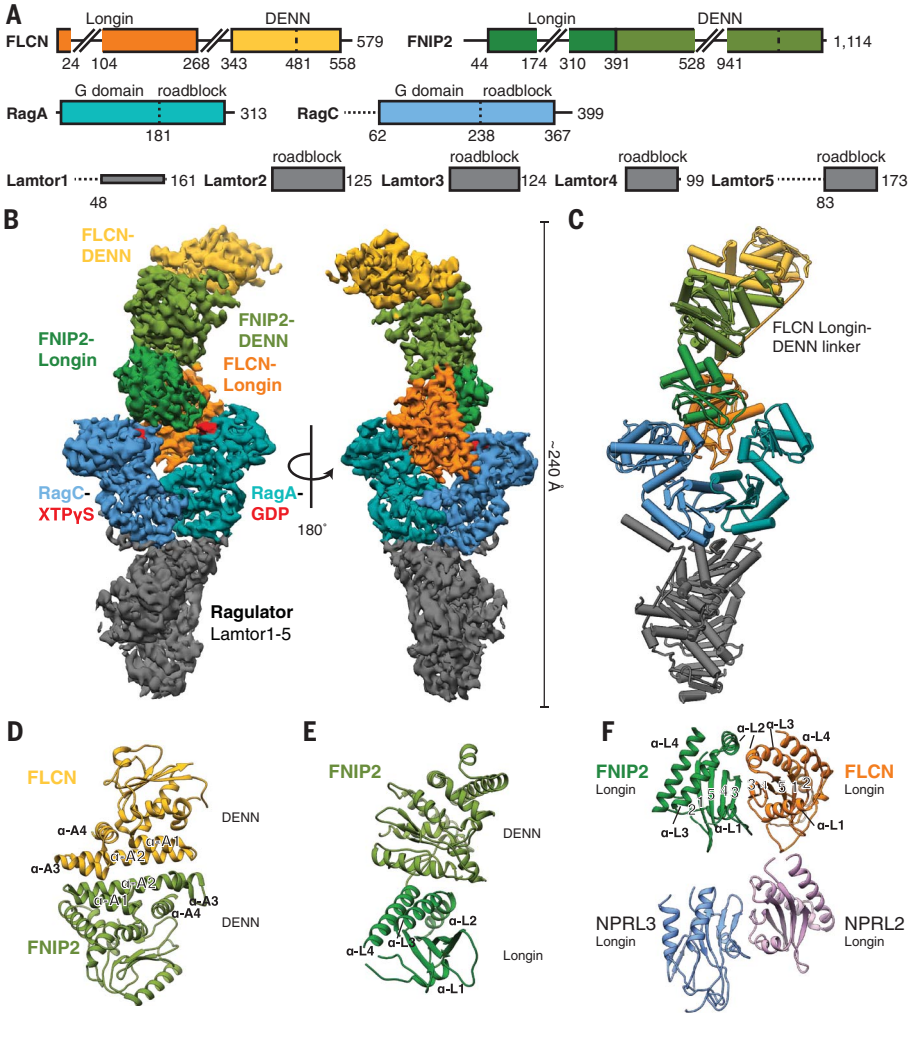


Fig. 2. Cryo-EM structure of the LFC.
(**A**) Domain organization of the LFC proteins. Horizontal dashed lines indicate disordered regions. Extended loops in FLCN and FNIP2 are indicated by diagonal lines. (**B**) Cryo-EM density map of the LFC, with model-based amplitude scaling (a LocScale map was used for all LFC density depictions; contour: 0.15). (**C**) Refined and assembled coordinate model of the LFC represented as pipes (cylinders; α helices) and planks (rectangles; β sheets). Color scheme is as in (A). (**D** to **F**) Ribbon representation of FLCN:FNIP2 domain architecture. (D) DENN-DENN interaction. (E) FNIP2 DENN module arrangement. (F) (Top) Longin-longin domain. (Bottom) The NPRL2:NPRL3 longin-longin domain from GATOR1 is displayed for comparison (PDB 6CES).

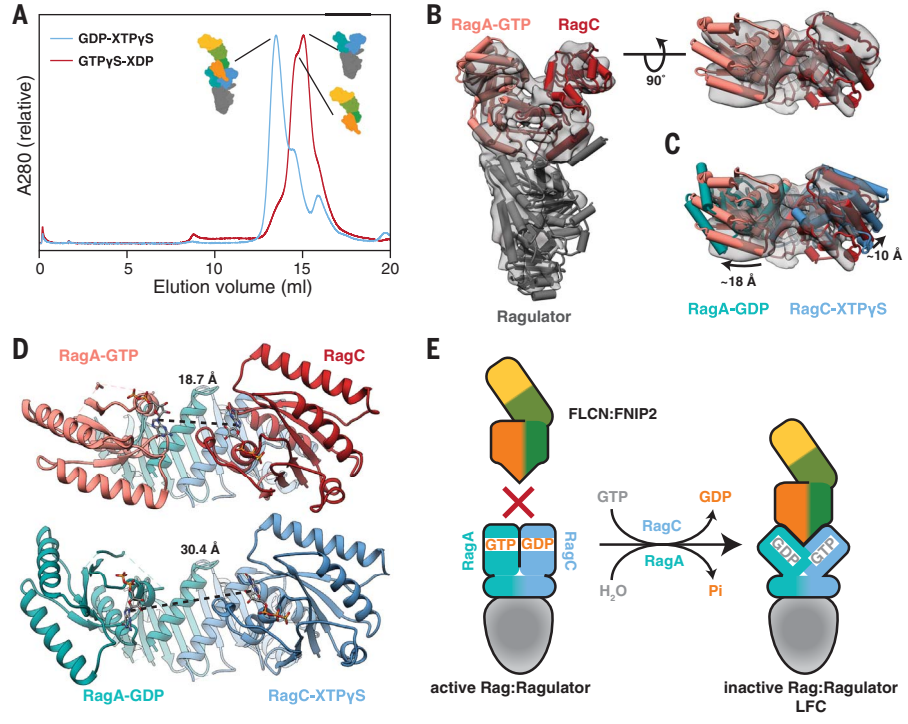


Fig. 3. A nucleotide-driven conformational change of the Rag GTPases controls LFC assembly.
(**A**) SEC profile of FLCN:FNIP2 mixed with Rag-Ragulator in the inactive (blue) and active (red) nucleotide binding state. (**B**) Side (left) and top (right) view of the cryo-EM structure of Ragulator-RagA^{GTP}:RagC^{XDP} (gray and transparent; contour: 0.32). Rigid-body–fitted atomic models for Ragulator (dark gray; PDB 6B9X) and RagA^{GTP}:RagC^{empty} (light red and red; PDB 6CES) are represented as pipes (α helices) and planks (β sheets). (**C**) Top view of the active Rag-Ragulator structure as in (B) with the atomic model of RagA^{GDP}:RagC^{XTPγS} from the LFC overlaid (cyan and blue). Model alignment was based on the constant roadblock domains of RagA:RagC. The extent of the G domain rotations is indicated. (**D**) Nucleotide-nucleotide distance in active (top) and inactive (bottom) RagA:RagC. As RagC in the active RagA:RagC structure is nucleotide-free, the respective nucleotide-nucleotide distance is based on a hypothetical model built by aligning the RagA and RagC G domains of the inactive LFC structure to the G domains of the pseudoactive RagA^{GTP}:RagC^{empty} model. (**E**) Structure-based model for the nucleotide-dependent formation of the LFC.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

and XTPγS nucleotides, respectively (150 versus 20 Å^2 buried surface area) (Figs. 4A and 5A and figs. S7, C and D, and S12). A FLCN:FNIP2 complex containing a FLCN^{F118D} (F118D: Phe^{118}→Asp) mutation located at the FLCN-RagA interface (Fig. 4A and figs. S3D and S13A) did not assemble into a stable LFC, as evaluated by SEC (Fig. 4B). Consistent with a disrupted LFC, GFP-FLCN^{F118D} failed to localize to the lysosomal membrane upon amino acid starvation, establishing the LFC as the structural basis for FLCN localization under low-nutrient conditions (Fig. 4, C and D). Also consistent with its inability to form the LFC, FLCN^{F118D}:FNIP2 had uninhibited GAP activity even in the presence of Ragulator, as judged by both tryptophan fluorescence– and HPLC-based GAP assays (Fig. 4E and fig. S13B).

Compared with other GTPases bound to GDP [e.g., the endoplasmic reticulum–associated small GTPase Sar1 (*31*)], the RagA^{GDP} switch regions are largely disordered in the LFC and do not contribute stabilizing interactions to the bound GDP (Fig. 4A and fig. S6C). This observation prompted us to investigate whether FLCN:FNIP2 binding to the RagA G domain influenced its nucleotide state within the LFC.

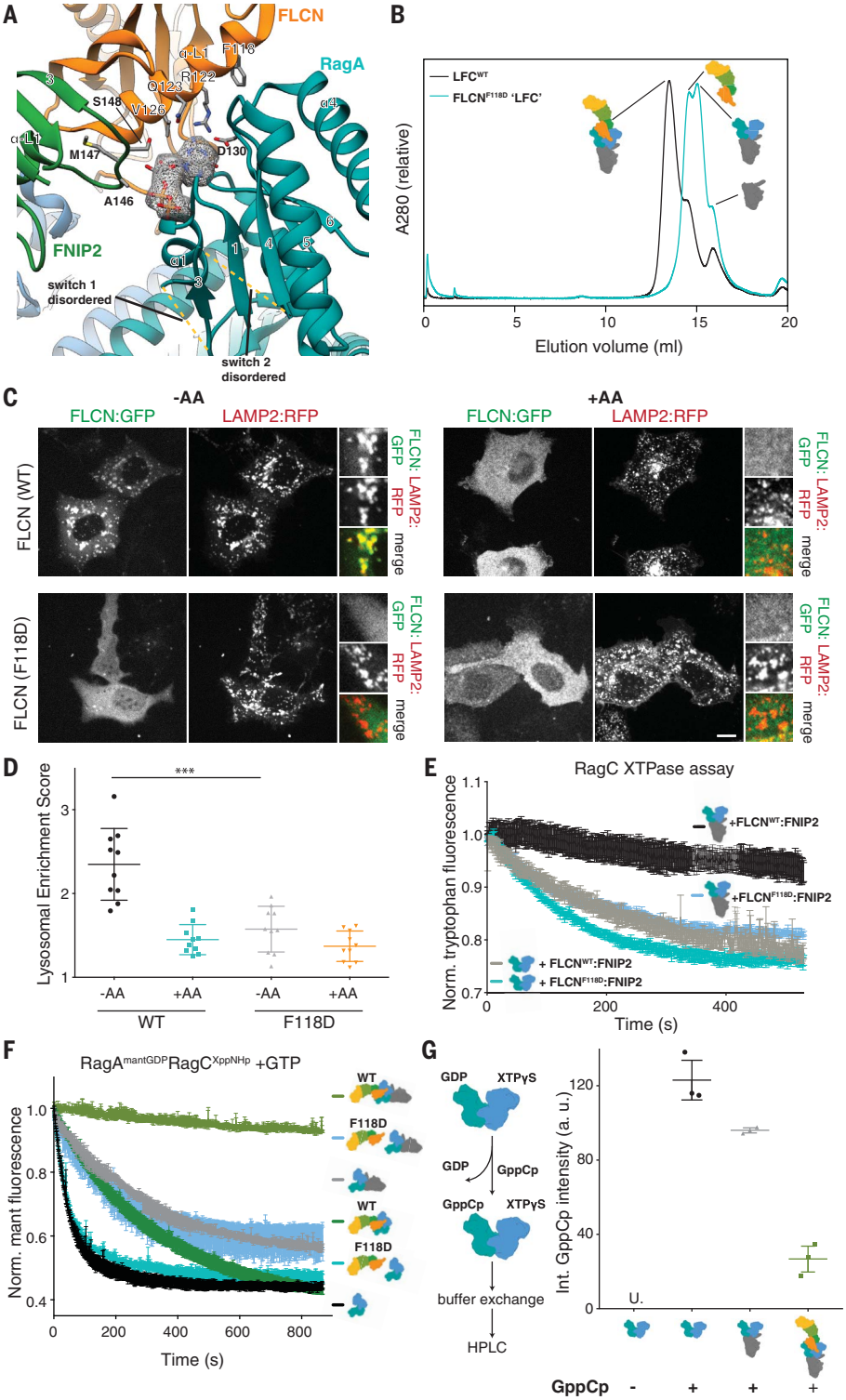


Fig. 4. The LFC stabilizes GDP-bound RagA and is necessary for lysosomal FLCN localization during amino acid starvation. (**A**) Nucleotide binding pocket and interaction with FLCN:FNIP2 (orange and green) of the RagA G domain (cyan) in the LFC. GDP is highlighted by the according cryo-EM density and stick representation. FLCN:FNIP2 residues in the interface are labeled (A, Ala; D, Asp; F, Phe; M, Met; Q, Gln; R, Arg; S, Ser; V, Val), and side chains are shown as sticks. Disordered switch 1 and switch 2 regions of RagA absent from the coordinate model are indicated by dashed lines. (**B**) SEC elution profile of Ragulator-RagA^{GDP}:RagC^{XppNHp}-FLCN:FNIP2 containing FLCN^{WT} (black) or FLCN^{F118D} (cyan). The peak shift indicates that the FLCN^{F118D}-containing complex failed to assemble into a stable LFC. (**C**) Fluorescence images of human embryonic kidney 293A (HEK293A) cells expressing the indicated FLCN:green fluorescent protein (FLCN:GFP) construct along with hemagglutinin:FNIP1 (HA:FNIP1) and the lysosomal marker Lamp2:red fluorescent protein (LAMP2:RFP). Cells were starved for 1 hour for amino acids (-AA) or starved for 50 min and then restimulated with amino acids for 10 min (+AA). Scale bar, 10 μm. (**D**) Quantitation of FLCN lysosomal enrichment score for immunofluorescence images in (C) (means ± SD, n = 10 cells for all conditions). ***P = 0.0001. (**E**) Tryptophan fluorescence XTPase assay with FLCN^{F118D}:FNIP2. RagA^{GDP}:RagC^{XTP} was incubated with FLCN^{WT}:FNIP2 (black and gray) or FLCN^{F118D}:FNIP2 (blue and cyan) in the presence (black and blue) or absence (gray and cyan) of Ragulator, as indicated. Plotted are means ± SEM. n = 3 replicates. (**F**) Mant fluorescence nucleotide exchange assay: Intrinsic mant fluorescence intensity decreases by ~50% when released from the GTPase. Mant RagA nucleotide exchange in response to addition of unlabeled GTP was visualized by mant fluorescence for RagA^{mantGDP}:RagC^{XppNHp} alone (black) or in complex with Ragulator (gray), FLCN^{WT}:FNIP2 (dark green), Ragulator and FLCN^{WT}:FNIP2 (light green), FLCN^{F118D}:FNIP2 (cyan), or Ragulator and FLCN^{118D}:FNIP2 (blue). Plotted are means ± SEM. n = 3 replicates. (**G**) HPLC-based nucleotide exchange assay for RagA. GDP replacement by GppCp (non-hydrolyzable GTP analog) for RagA^{GDP}:RagC^{XTPγS} alone (black), in complex with Ragulator (gray), or in the full LFC (light green) was measured by the GppCp peak in the HPLC elution profile. "U." indicates that the GppCp signal was undetectable. Plotted are individual data points and means ± SD. n = 3 replicates.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

We monitored RagA nucleotide exchange using both HPLC and a fluorescence-based assay in which the RagA:RagC dimer was first bound to (2′ or 3′)-*O*-(*N*-methylanthraniloyl) (mant) GTP and nonhydrolyzable XppNHp, then incubated with GATOR1 to generate $RagA^{mantGDP}$:$RagC^{XppNHp}$ substrate. Nucleotide exchange was monitored as a mantGDP fluorescence decrease upon release from RagA in the presence of excess unlabeled GTP. Rapid and complete replacement of mantGDP with GTP occurred in the absence of other proteins and was confirmed by HPLC (Fig. 4, F and G). In contrast, Sar1 showed negligible mantGDP release under identical conditions (fig. S13C) (*32*). The addition of either Ragulator or FLCN:FNIP2 slowed the kinetics of mantGDP release (Fig. 4F). Notably, when the entire LFC was assembled, RagA nucleotide exchange was barely detectable (Fig. 4, F and G). Consistent with its inability to form the LFC, mutant $FLCN^{F118D}$:FNIP2 did not inhibit mantGDP exchange on RagA either by itself or in the presence of Ragulator (Fig. 4F). These data show that RagA in the context of the isolated Rag dimer, unlike most other small GTPases, is competent to spontaneously exchange GDP for GTP and that the exchange process is blocked in the LFC.

The switch 1 and switch 2 regions of RagC are well defined (Fig. 5A). They adopt a conformation analogous to GTP-bound RagA, Gtr1, and other small GTPases (*14*, *28*, *30*, *33*). In this conformation, RagC Thr^{96} and Gly^{119} stabilize the γ-phosphate of the bound XTPγS nucleotide (Fig. 5A). In line with the finding that FLCN:FNIP2 does not exert GAP activity in the LFC (Fig. 1B), the interface between FLCN:FNIP2 and $RagC^{XTPγS}$ lacks any Arg residue poised to function as the catalytic "arginine finger" characteristic for most GAPs (*34*) (Fig. 5A). The catalytic arginine of GATOR1 is Arg^{78} located in loop β4-β5 of the NPRL2 longin domain (*35*). The longin domains of FLCN and FNIP2 each contain an arginine residue in loop β4-β5 as well, but only Arg^{164} of FLCN is highly conserved from yeast to human (Fig. 5B and figs. S9 and S10). In the LFC, FLCN Arg^{164} is located more than 20 Å away from either nucleotide between the two G domains, with no intersubunit contacts (Fig. 5C). We found that mutant $FLCN^{R164A}$ (R164A: Arg^{164}→Ala) assembles into an LFC that migrates normally on SEC (Fig. 5D and fig. S3E), yet GAP activity is abolished (Fig. 5, E and F). Moreover, the RagC-GAP–incompetent $FLCN^{R164A}$:FNIP2 complex is capable of blocking RagA nucleotide exchange (fig. S13D) and localizes to the lysosomal membrane upon amino acid starvation (fig. S13E). These data establish $FLCN^{R164}$ as a catalytically required arginine for canonical FLCN RagC-GAP activity and explain structurally why FLCN does not function as an active GAP in the LFC.

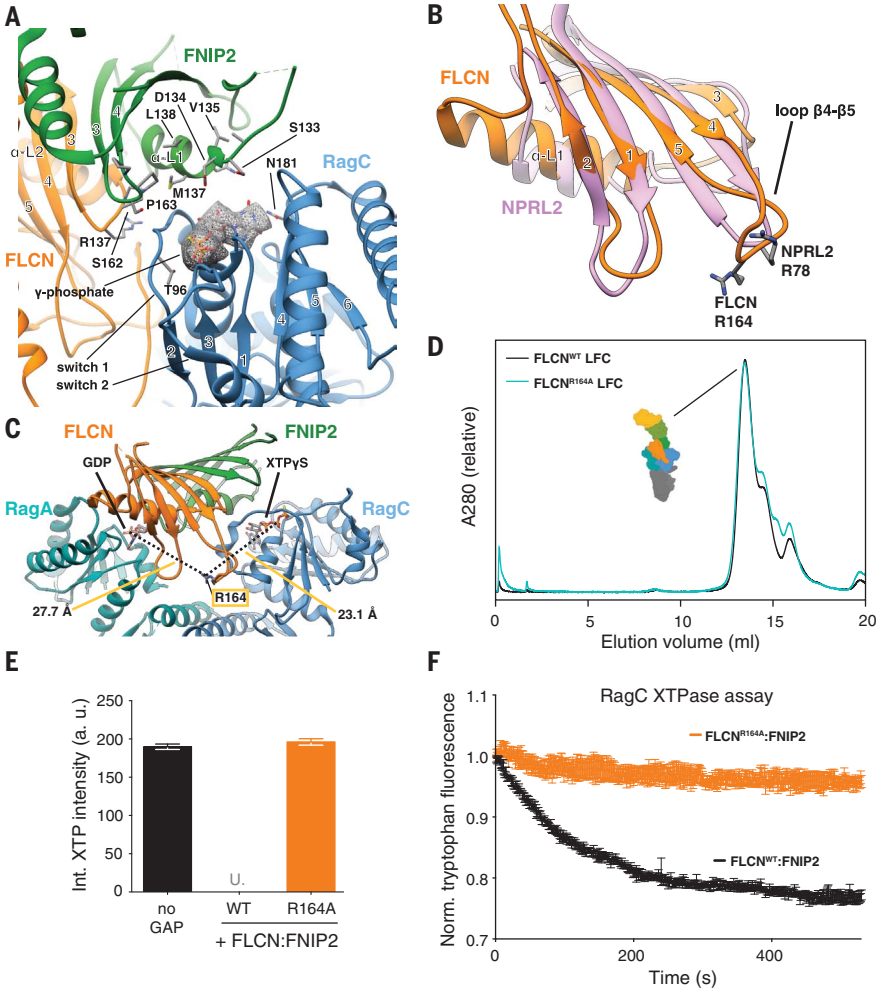


Fig. 5. FLCN Arg^{164} mediates GAP activity and is in a GAP-incompetent position in the LFC. (**A**) Nucleotide-binding pockets and interaction with FLCN:FNIP2 (orange and green) of the RagC G domain (blue) in the LFC. XTPγS is highlighted by the according cryo-EM density and stick representation. FLCN:FNIP2 residues in the interface are labeled (L, Leu; N, Asn; P, Pro; T, Thr), and side chains are shown as sticks. The γ-phosphate of XTPγS locking the switch 1 and switch 2 regions of RagC is indicated. (**B**) Overlay of the FLCN (orange) and NPRL2 (pink) longin domains. α helices α3 and α4 are omitted for clarity. The NPRL2 arginine finger (R78) and FLCN R164 in loop β4-β5 are represented as sticks. (**C**) Position of FLCN R164 in the LFC. Distances to the β-phosphate of GDP (RagA) and XTPγS (RagC) are indicated. (**D**) SEC elution profile of the LFC (Ragulator-$RagA^{GDP}$:$RagC^{XTPγS}$-FLCN:FNIP2) containing $FLCN^{WT}$ (black) or mutant $FLCN^{R164A}$ (cyan). (**E**) HPLC-based GTPase assay assessing GAP activity of $FLCN^{WT}$:FNIP2 or $FLCN^{R164A}$:FNIP2 (orange) on a $RagA^{GTP}$:$RagC^{XTP}$ substrate by integrating the remaining XTP nucleotide signal. Plotted are means ± SD. *n* = 3 replicates. "U." indicates that the remaining XTP signal was undetectable by HPLC. (**F**) Tryptophan fluorescence XTPase assay with FLCN:FNIP2 mutants. $RagA^{GTP}$:$RagC^{XTP}$ is incubated with $FLCN^{WT}$:FNIP2 (black) or $FLCN^{R164}$:FNIP2 (orange). Plotted are means ± SEM. *n* = 3 replicates.

Regulation of MiT/TFE factors by FLCN and LFC

FLCN has emerged as a central controller of the master regulators of lysosomal biogenesis and autophagy, the MiT/TFE family transcription factors TFE3 and TFEB (*19*, *21*, *36*–*39*). TFE3 and TFEB are phosphorylated by mTORC1 to prevent their nuclear translocation and to promote retention in the cytoplasm in nutrient-replete conditions. Upon nutrient starvation and mTORC1 inactivation, the MiT/TFE factors relocalize to the nucleus and activate their gene expression programs (*40*–*43*). In agreement with a key role for FLCN in MiT/TFE regulation, its knockdown using both short hairpin RNA (shRNA) and endonuclease dead Cas9 (dCas9)–mediated silencing led to nuclear localization of TFE3 under full nutrients (fig. S14, A to D). In contrast, FLCN depletion was largely dispensable for phosphorylation of canonical mTORC1 substrates S6K1 and 4E-BP1 (fig. S14, E and F), implicating the LFC as a key component for mTORC1-dependent control of the MiT/TFE factors.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

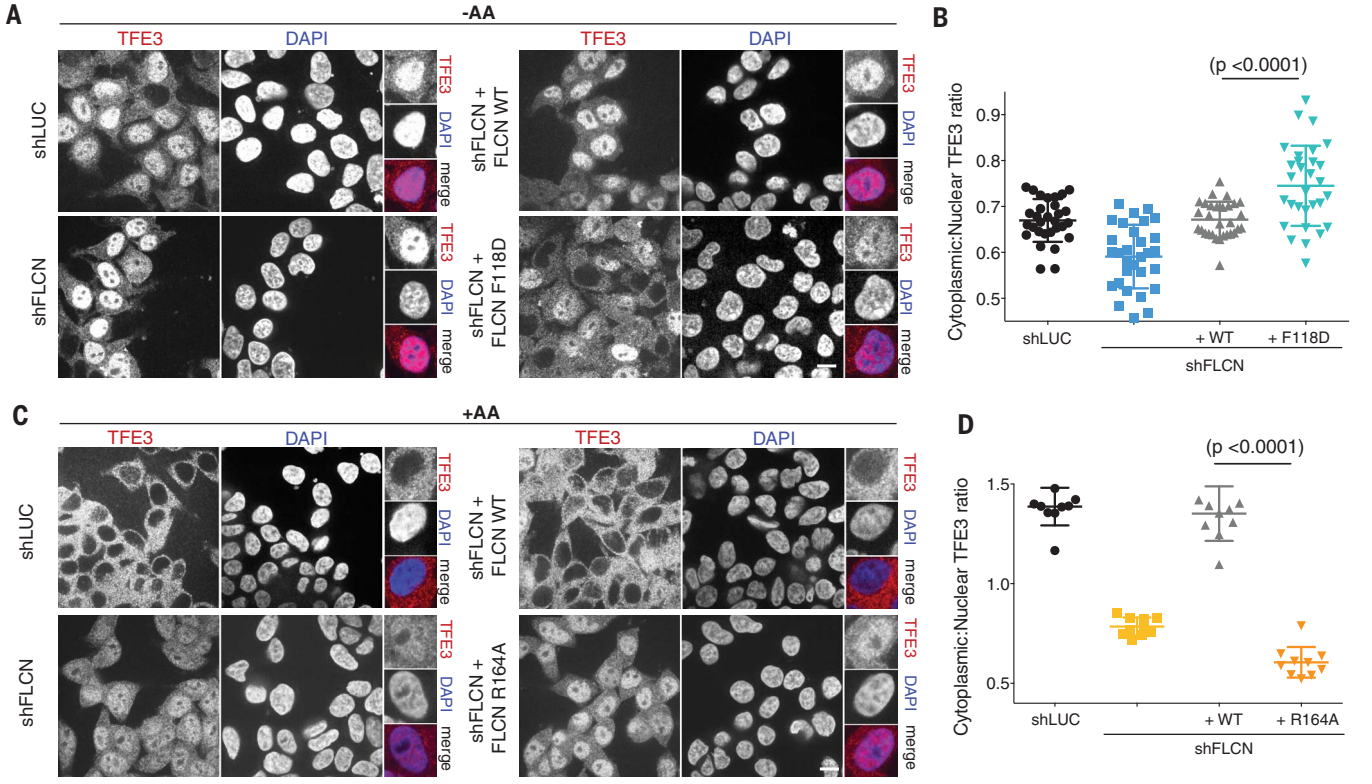


Fig. 6. FLCN GAP activity is necessary for nutrient-induced cytoplasmic TFE3 localization. (**A**) Immunofluorescence images of HEK293T cells expressing the indicated shRNA and FLCN rescue constructs. Cells were starved for 1 hour for amino acids (-AA). Scale bar, 10 μm. DAPI, 4′,6-diamidino-2-phenylindole; LUC, luciferase. (**B**) Quantitation of cytoplasmic:nuclear TFE3 ratio for immunofluorescence images in (A) (means ± SD, *n* = 30 cells for all). (**C**) Immunofluorescence images of HEK293T cells expressing the indicated shRNA and FLCN rescue constructs. Cells were starved for 1 hour and restimulated with amino acids for 2 hours (+AA). Scale bar, 10 μm. (**D**) Quantitation of cytoplasmic:nuclear TFE3 ratio for immunofluorescence images in (C) (means ± SD, *n* = 10 cells for all). (**E**) Model for regulation of Rag GTPase activation by the LFC. In low-nutrient conditions, binding of FLCN:FNIP2 to a stable $RagA^{GDP}$:$RagC^{GTP}$-Ragulator platform leads to LFC formation. Upon nutrient restimulation, RagA nucleotide exchange and LFC disassembly liberate the RagC-GAP activity of FLCN:FNIP2. The resulting active Rag GTPase complex is competent to activate mTORC1 and promote cytoplasmic retention of TFEB and TFE3.

In amino acid–deprived cells in which FLCN was depleted, or depleted and reconstituted with wild-type (WT) FLCN, endogenous TFE3 was largely localized to the nucleus. Relative to the wild-type FLCN reconstituted cells, reconstituting FLCN-depleted cells with LFC-defective $FLCN^{F118D}$, which has uninhibited RagC-GAP activity, caused TFE3 to become less nuclear (Fig. 6, A and B). In FLCN-depleted cells that were restimulated with amino acids, TFE3 failed to relocalize from the nucleus to the cytoplasm (Fig. 6, C and D). Cytoplasmic distribution of TFE3 was fully restored by re-expressing WT FLCN. In contrast, expressing GAP-defective $FLCN^{R164A}$ failed to restore cytoplasmic translocation of TFE3, which therefore remained nuclear despite high nutrient concentrations (Fig. 6, C and D).

Together, these data strongly suggest that the RagC-GAP activity of FLCN enables proper regulation of MiT/TFE nuclear translocation by nutrient inputs (Fig. 6E). Under low-nutrient conditions, the LFC suppresses the RagC-GAP activity of FLCN, maximizing nuclear translocation of the MiT/TFE factors. Conversely, in high-nutrient conditions, disassembly of the LFC liberates the RagC-GAP function of FLCN, which is key for mTORC1-dependent phosphorylation and cytoplasmic retention of the MiT/TFE factors.

Discussion

Our structure of the LFC reveals how FLCN and FNIP2 assemble into one complex by means of an intricate domain arrangement, as well as how they interact with the G domains of the Rag GTPases in their inactive state. It serves as the structural basis for FLCN:FNIP2 localization at the lysosomal membrane under low amino acid concentrations and explains our findings from biochemical reconstitution and in vitro GAP and nucleotide exchange assays. First, we identified Arg^{164} of FLCN as a catalytically required arginine residue for GAP activity. Second, we found that Arg^{164} is located far from the RagC nucleotide pocket within the LFC, and thus the LFC is incompatible with FLCN GAP activity. Third, LFC formation strongly depends on RagA in its GDP-bound conformation, explaining the requirement of GATOR1 for lysosomal FLCN localization (*22*). Fourth, FLCN:FNIP2 interacts with $RagA^{GDP}$ and, together with Ragulator, blocks exchange of GDP with GTP

Downloaded from http://science.sciencemag.org/ on November 28, 2019

in RagA, possibly stabilizing the LFC. Fifth, by controlling the RagC-GAP activity of FLCN, the LFC regulates a subset of mTORC1 substrates, the MiT/TFE transcription factors.

Mutations in the *FLCN* tumor suppressor gene are responsible for BHD, which predisposes patients to kidney tumors (*44*). However, several missense mutations map to the surface of the DENN domain and, on the basis of our structure, seem unlikely to disrupt Rag GTPase interaction or LFC stability (fig. S15). Notably, the MiT/TFE factors, TFEB and TFE*3*, are independently translocated and up-regulated in renal cell carcinoma (*44*–*46*), providing a plausible mechanism for carcinogenesis downstream of FLCN loss.

The ability of FLCN:FNIP2 to contact its client RagC in two states—the inactive, stable LFC and an active, GAP-compatible but transient interaction—is rare among known GAPs. The other notable exception is GATOR1. The cryo-EM structure of GATOR1 in complex with a Rag dimer revealed a stable, GAP-inhibited state in which the noncatalytic subunit of GATOR1 makes direct contact with the G domain of RagA (*14*). FLCN:FNIP2 is the mirror image of the GATOR1 complex in the sense that both contain a dual longin domain–based GAP and both complexes have positive and negative regulatory functions. The multiple configurations of this complex with respect to the Rag GTPases are substantially more elaborate than the single reaction carried out by conventional GAP complexes and suggest that the ability of cells to respond to changing nutrient concentrations may be so important that particularly stringent mechanisms are required.

REFERENCES AND NOTES

1. R. A. Saxton, D. M. Sabatini, *Cell* **168**, 960–976 (2017).
2. R. E. Lawrence, R. Zoncu, *Nat. Cell Biol.* **21**, 133–142 (2019).
3. Y. Sancak *et al.*, *Science* **320**, 1496–1501 (2008).
4. Y. Sancak *et al.*, *Cell* **141**, 290–303 (2010).
5. A. Efeyan *et al.*, *Nature* **493**, 679–683 (2013).
6. B. M. Castellano *et al.*, *Science* **355**, 1306–1311 (2017).
7. E. Kim, P. Goraksha-Hicks, L. Li, T. P. Neufeld, K. L. Guan, *Nat. Cell Biol.* **10**, 935–945 (2008).
8. L. Bar-Peled, L. D. Schweitzer, R. Zoncu, D. M. Sabatini, *Cell* **150**, 1196–1208 (2012).
9. D. Teis, W. Wunderlich, L. A. Huber, *Dev. Cell* **3**, 803–814 (2002).
10. K. B. Rogala *et al.*, *Science* **366**, 468–475 (2019).
11. M. Anandapadamanaban *et al.*, *Science* **366**, 203–210 (2019).
12. L. Bar-Peled *et al.*, *Science* **340**, 1100–1106 (2013).
13. N. Panchaud, M. P. Péli-Gulli, C. De Virgilio, *Sci. Signal.* **6**, ra42 (2013).
14. K. Shen *et al.*, *Nature* **556**, 64–69 (2018).
15. M. L. Nickerson *et al.*, *Cancer Cell* **2**, 157–164 (2002).
16. M. Baba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15552–15557 (2006).
17. H. Hasumi *et al.*, *Gene* **415**, 60–67 (2008).
18. Z. Y. Tsun *et al.*, *Mol. Cell* **52**, 495–505 (2013).
19. C. S. Petit, A. Roczniak-Ferguson, S. M. Ferguson, *J. Cell Biol.* **202**, 1107–1122 (2013).
20. Y. Hasumi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18722–18727 (2009).
21. S. Wada *et al.*, *Genes Dev.* **30**, 2551–2564 (2016).
22. J. Meng, S. M. Ferguson, *J. Cell Biol.* **217**, 2765–2776 (2018).
23. M. R. Ahmadian *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7065–7070 (1999).
24. R. E. Lawrence *et al.*, *Nat. Cell Biol.* **20**, 1052–1063 (2018).
25. D. Zhang, L. M. Iyer, F. He, L. Aravind, *Front. Genet.* **3**, 283 (2012).
26. M. Y. Su *et al.*, *Mol. Cell* **68**, 835–846.e3 (2017).
27. R. K. Nookala *et al.*, *Open Biol.* **2**, 120071 (2012).
28. R. Gong *et al.*, *Genes Dev.* **25**, 1668–1673 (2011).
29. M. E. G. de Araujo *et al.*, *Science* **358**, 377–381 (2017).
30. J. H. Jeong *et al.*, *J. Biol. Chem.* **287**, 29648–29653 (2012).
31. M. Huang *et al.*, *J. Cell Biol.* **155**, 937–948 (2001).
32. C. Barlowe, R. Schekman, *Nature* **365**, 347–349 (1993).
33. I. R. Vetter, A. Wittinghofer, *Science* **294**, 1299–1304 (2001).
34. M. R. Ahmadian, P. Stege, K. Scheffzek, A. Wittinghofer, *Nat. Struct. Biol.* **4**, 686–689 (1997).
35. K. Shen, M. L. Valenstein, X. Gu, D. M. Sabatini, *J. Biol. Chem.* **294**, 2970–2975 (2019).
36. F. Villegas *et al.*, *Cell Stem Cell* **24**, 257–270.e8 (2019).
37. J. C. Kennedy *et al.*, *Hum. Mol. Genet.* 10.1093/hmg/ddz158 (2019).
38. S. B. Hong *et al.*, *PLOS ONE* **5**, e15793 (2010).
39. J. Betschinger *et al.*, *Cell* **153**, 335–347 (2013).
40. C. Settembre *et al.*, *EMBO J.* **31**, 1095–1108 (2012).
41. J. A. Martina, Y. Chen, M. Gucek, R. Puertollano, *Autophagy* **8**, 903–914 (2012).
42. A. Roczniak-Ferguson *et al.*, *Sci. Signal.* **5**, ra42 (2012).
43. G. Napolitano *et al.*, *Nat. Commun.* **9**, 3312 (2018).
44. L. S. Schmidt, W. M. Linehan, *Gene* **640**, 28–42 (2018).
45. R. P. Kuiper *et al.*, *Hum. Mol. Genet.* **12**, 1661–1669 (2003).
46. I. J. Davis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6051–6056 (2003).

ACKNOWLEDGMENTS

We thank H. R. Shin and M.-Y. Su for comments on the manuscript and D. Toso for microscope operational support. Access to the FEI Titan Krios was provided through the BACEM UCB facility. We also thank the J. Thorner laboratory for use of their spectrofluorimeter, L. Yuan and D. Melville from the R. Schekman laboratory for assistance with tryptophan fluorescence experiments and purified Sar1 GTPase, and M. Kampmann for 239T CRISPRi cells. **Funding:** This work was supported by NIH 1DP2CA195761-01 (R.Z.), GM111730 (J.H.H.), CA223029 (L.N.Y.); the Pew-Stewart Scholarship for Cancer Research and Damon Runyon-Rachleff Innovation Award (R.Z.); a University of California Cancer Research Coordinating Committee Predoctoral Fellowship (R.E.L.); an EMBO Long-Term Fellowship (S.A.F.); and a Jane Coffin Childs fellowship (A.L.Y.). **Author contributions:** Conceptualization, R.E.L., S.A.F., J.H.H., and R.Z.; Investigation, R.E.L., S.A.F., Y.F., A.L.Y., L.N.Y., and C.-Y.L.; Resources, D.J.K., A.M.T., and A.J.S.; Supervision, J.H.H. and R.Z.; Writing – Original Draft, R.E.L. S.A.F., J.H.H., and R.Z.; Writing – Review & Editing, all authors. **Competing interests:** R.Z. is a cofounder and stockholder in Frontier Medicines. J.H.H. is a scientific founder of Casma Therapeutics. **Data and materials availability:** EM density maps have been deposited in the Electron Microscopy Data Bank with accession numbers EMD-0554 (LFC) and EMD-0556 (active Rag-Ragulator). Atomic coordinates for the LFC have been deposited in the Protein Data Bank with identification number 6NZD.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/971/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 and S2
References (*47*–*69*)
Movie S1

View/request a protocol for this paper from *Bio-protocol*.

15 February 2019; accepted 23 October 2019
Published online 31 October 2019
10.1126/science.aax0364

Downloaded from http://science.sciencemag.org/ on November 28, 2019

PALEONTOLOGY

Exceptional continental record of biotic recovery after the Cretaceous–Paleogene mass extinction

T. R. Lyson[1]*, I. M. Miller[1], A. D. Bercovici[1,2], K. Weissenburger[1], A. J. Fuentes[3], W. C. Clyde[3], J. W. Hagadorn[1], M. J. Butrim[4], K. R. Johnson[2], R. F. Fleming[1], R. S. Barclay[2], S. A. Maccracken[2,5], B. Lloyd[6], G. P. Wilson[7], D. W. Krause[1,8], S. G. B. Chester[9,10,11]

We report a time-calibrated stratigraphic section in Colorado that contains unusually complete fossils of mammals, reptiles, and plants and elucidates the drivers and tempo of biotic recovery during the poorly known first million years after the Cretaceous–Paleogene mass extinction (KPgE). Within ~100 thousand years (ka) post-KPgE, mammalian taxonomic richness doubled, and maximum mammalian body mass increased to near pre-KPgE levels. A threefold increase in maximum mammalian body mass and dietary niche specialization occurred at ~300 ka post-KPgE, concomitant with increased megafloral standing species richness. The appearance of additional large mammals occurred by ~700 ka post-KPgE, coincident with the first appearance of Leguminosae (the bean family). These concurrent plant and mammal originations and body-mass shifts coincide with warming intervals, suggesting that climate influenced post-KPgE biotic recovery.

The Cretaceous–Paleogene (K–Pg) boundary marks Earth's most recent mass extinction, when >75% of species, including nonavian dinosaurs, went extinct (*1*). In the terrestrial realm, the mass extinction was followed by a radiation of modern clades, particularly placental mammals (*2*), crown birds (*3*), and angiosperms (*4*). The drivers (*5*–*8*) and tempo (*9*, *10*) of the K–Pg mass extinction (KPgE) have been hotly debated, and the patterns of terrestrial recovery in the first million years after the KPgE remain poorly understood. The extinction of all large-bodied vertebrates (*5*) undoubtedly affected the post-KPgE taxonomic, ecologic, and body-mass diversification of various clades, but the lack of a well-preserved fossil record has left the factors influencing ecosystem recovery unknown. Here, we provide a detailed and temporally constrained terrestrial fossil record from this critical interval.

Fossils of terrestrial and freshwater organisms from the first million years after the KPgE are exceedingly rare worldwide, hindering our knowledge of post-KPgE taxonomic and ecological radiations. Thus far, the most fossiliferous sections from this time interval occur in the Williston, San Juan, Hanna, and Denver basins along the eastern margin of the Rocky Mountains in North America (*11*, *12*). In all of these study areas, discontinuous outcrops result in composite stratigraphic sections; plant fossil localities are geographically widely spaced, vertebrate-bearing horizons are sparse and separated by long temporal gaps, complete vertebrate fossils are exceptionally rare, and age control is variable (*10*–*17*). The Williston Basin has the most comprehensive fossil record with excellent age control, but the vertebrate specimens are fragmentary (*10*, *12*, *13*). The San Juan Basin preserves a well-studied early Paleocene vertebrate record but does not record the K–Pg boundary itself (*16*). Moreover, overlying Paleocene rocks only contain two vertebrate fossil–bearing horizons in the first 1 million years post-KPgE (*16*). The Hanna Basin K–Pg section is rich in fragmentary vertebrate fossils but has structurally complex strata and lacks a detailed chronostratigraphic framework (*17*). Finally, the Denver Basin has well-documented Cretaceous and Paleocene strata, a precisely dated K–Pg boundary, and abundant, geographically dispersed plant fossils but, prior to this study, contains only a sparse and fragmentary vertebrate fossil record (*14*, *15*, *18*, *19*).

Corral Bluffs study area, Denver Basin, Colorado, USA

We developed a high-resolution stratigraphic framework in the Corral Bluffs study area (east of Colorado Springs), a single continuous (physically traceable) ~27-km^2 outcrop from the Denver Basin that preserves the biotic recovery of a terrestrial ecosystem in the first million years post-KPgE (*20*) (Fig. 1 and fig. S1). This stratigraphy is tied to the geomagnetic polarity time scale (GPTS 2012) using paleomagnetics and chemical abrasion isotope dilution thermal ionization mass spectrometry (CA-ID-TIMS) U-Pb-dated volcanic ash (*20*). For comparison, ages using an alternative age model based on work in the Denver Basin (*18*) are also provided in data S1 to S14. The study area contains an exceptionally dense vertebrate (299 localities) and megafloral (65 localities) record, with fossils occurring at >150 stratigraphic levels in the ~250-m-thick sequence (Fig. 1). The extensive and nearly continuous outcrop belt spans the last ~100 thousand years (ka) of the Cretaceous and the first ~1 million years (Ma) of the Paleocene. It includes four North American Land Mammal Age (NALMA) interval zones, four palynostratigraphic biozones, three magnetochron boundaries, two U-Pb radiometric dates, and the palynologically defined K–Pg boundary, yielding a locally derived, high-resolution chronostratigraphic framework (Fig. 1, figs. S2 to S5, and supplementary materials) (*20*). Together, these data provide an unprecedented opportunity to assess the biotic recovery of a terrestrial ecosystem after the KPgE.

Vertebrate fossils in the Corral Bluffs succession are unusually complete for this time period, are found in a range of depositional environments, and represent a diversity of taxa and body sizes (Figs. 1 and 2). Most are three-dimensionally preserved in hydroxyapatite concretions and are found in all observed facies, often as articulated skeletons or skulls with intact delicate structures such as middle ear and hyoid elements (Fig. 2). Among vertebrate specimens preserved in concretions, mammalian, turtle, and crocodilian crania (Fig. 2, A to T) and turtle shells (Fig. 2, U to X) are the most common. Individual fossils range in size from ~3 mm^2 (isolated teeth) to larger forms such as 1.5-m-long, articulated crocodilian skeletons. Plant fossils also span the size spectrum across all observed facies, including microscopic palynomorphs, seeds, leaves, roots, branches, in situ saplings, and even large stumps and logs (Fig. 3).

We recognize 16 mammalian taxa, 8 of which are based on cranial remains, including the first occurrence of the late Puercan (Pu3) index taxon *Taeniolabis taoensis* (Fig. 2, K and L) from the Denver Basin. Cranial size and lower first molar area were used to estimate mammalian body mass—an important feature that affects many aspects of the biology and ecology of mammals (Fig. 4) (*21*). Given that there appears to be bias toward large vertebrates

[1]Denver Museum of Nature & Science, 2001 Colorado Boulevard, Denver, CO 80205, USA. [2]National Museum of Natural History, Smithsonian Institution, 10th Street and Constitution Avenue NW, Washington, DC 20560, USA. [3]Department of Earth Sciences, University of New Hampshire, 56 College Road, Durham, NH 03824, USA. [4]Department of Earth and Environmental Sciences, Wesleyan University, Exley Science Center 333, Middletown, CT 06459, USA. [5]Department of Entomology, University of Maryland, College Park, 4291 Fieldhouse Drive, College Park, MD 20742, USA. [6]Department of Geology, Colorado College, 14 E. Cache La Poudre Street, Colorado Springs, CO 80903, USA. [7]Department of Biology, University of Washington, 251 Life Sciences Building, Seattle, WA 98195, USA. [8]Department of Anatomical Sciences, Stony Brook University, 101 Nicolls Road, Stony Brook, NY 11794, USA. [9]Department of Anthropology and Archaeology, Brooklyn College, City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210, USA. [10]Department of Anthropology, The Graduate Center, City University of New York, 365 Fifth Avenue, New York, NY 10016, USA. [11]New York Consortium in Evolutionary Primatology, 200 Central Park West, New York, NY 10024, USA.
*Corresponding author. Email: tyler.lyson@dmns.org

Downloaded from http://science.sciencemag.org/ on November 28, 2019

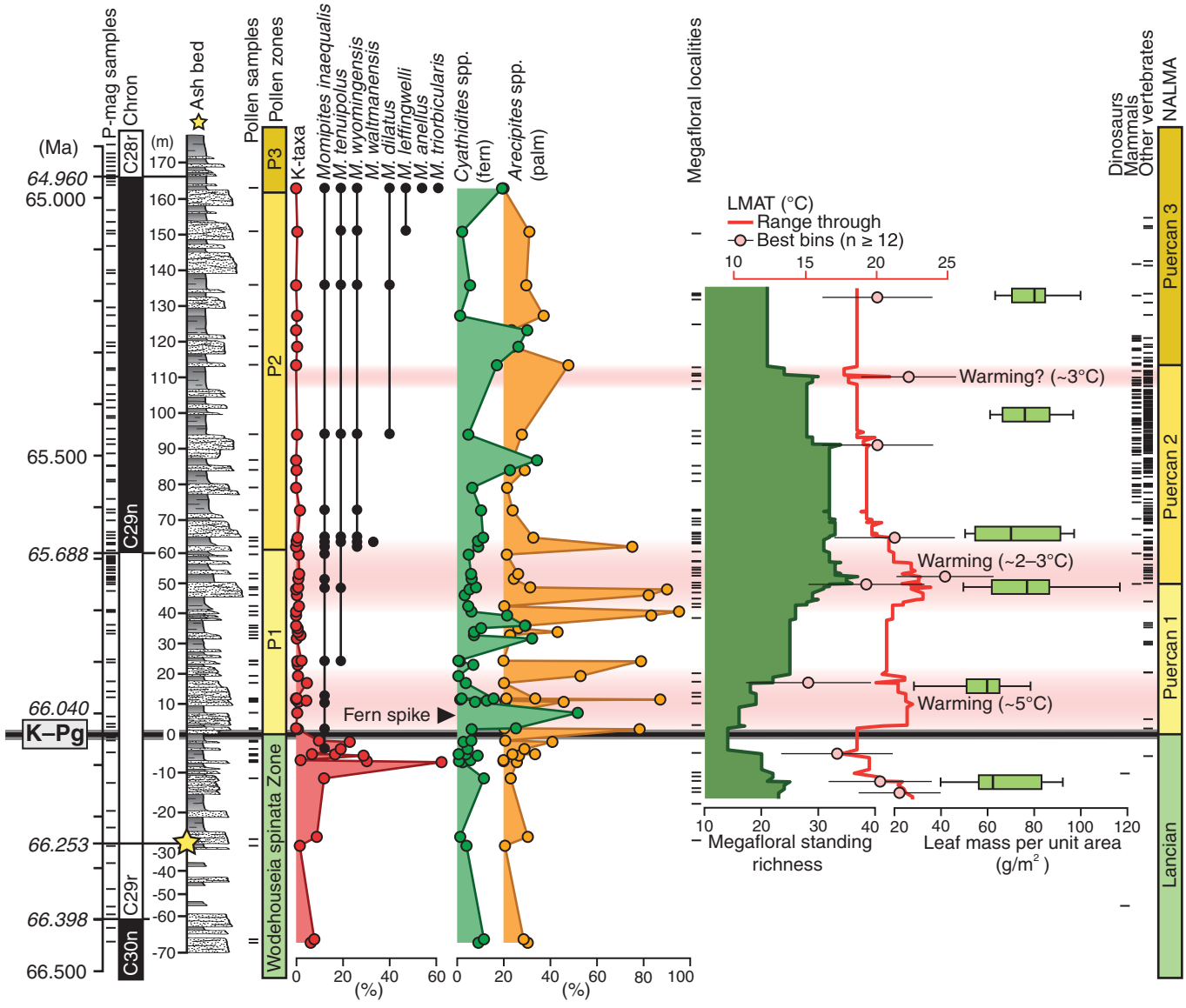


Fig. 1. Temporally calibrated stratigraphic, floral, and faunal data for the K–Pg interval in the Corral Bluffs study area (fig. S1). Stratigraphy is tied to the GPTS 2012 using paleomagnetics (P-mag) and CA-ID-TIMS U-Pb-dated ash (italicized dates) (*20*) (data S1 and figs. S3 and S5). The composite lithostratigraphic log (figs. S2 to S5) is dominated by intercalated mudstone and sandstone, reflecting a variety of fluvial facies. Pollen zones (data S3) are defined by diversification of *Momipites* spp. (fossil juglandaceous pollen) (Fig. 3I). The K–Pg boundary is demarcated by the decrease in abundance of Cretaceous pollen taxa (labeled "K-taxa") without recovery, and the subsequent fern (*Cyathidites* spp.) spike (data S2). Relative abundance (%) of fern (*Cyathidites* spp.) and palm (*Arecipites* spp.) (Fig. 3E) palynomorphs increased considerably post-KPgE (data S2); note that palm pollen percentages are offset from scale by 20%. Standing richness of dicot morphospecies or megafloral standing richness is exclusive of species that occur at a single locality (data S4 to S7). Leaf-estimated mean annual temperature (LMAT) calibrated with East Asian forests (data S8 and fig. S6). Horizontal pink shading indicates hypothesized warming intervals. Estimated leaf mass per unit area (data S9 and S10 and fig. S7) is shown with box plots that represent the distribution of species-site pair means for each 30-m bin (supplementary materials). Box plots are placed along the *y* axis near each bin's stratigraphic midpoint and are repositioned for visibility. See data S11 and supplementary materials for placement of NALMAs. Tick marks for P-mag, pollen zones, megafloral standing richness, and NALMAs show stratigraphic placement of samples and fossil localities (supplementary materials).

in our dataset (supplementary materials and data S11), we focused on maximum mammalian body mass. The largest-bodied mammals disappeared at the K–Pg boundary (*10*) and returned to near pre-KPgE levels within 100 ka after the K–Pg boundary (Fig. 4). Subsequent shifts in maximum mammalian body mass occurred at the Pu1–Pu2 and near the Pu2–Pu3 transitions, ~300 and ~700 ka post-KPgE, respectively (Fig. 4). In addition, the pattern and abundance of vertebrates preserved in all paleoenvironments suggest that by ~700 ka post-KPgE, the largest mammals (25+ kg) were spatially partitioned across the landscape. We observe a strong pattern of association between taxa and facies (Fig. 4), indicating that baenid turtles (Fig. 2, Q to T) and *T. taoensis* (Fig. 2, K and L) lived in or near river channel margins, whereas chelydroid turtles (Fig. 2, W and X) and the large periptychid mammals *Ectoconus ditrigonus* (Fig. 2, C, D, G, and H) and *Carsioptychus coarctatus* (Fig. 2, I and J) primarily occupied distal portions of the floodplain (Fig. 4).

Downloaded from http://science.sciencemag.org/ on November 28, 2019

We recognize 233 plant morphospecies in our study area (supplementary materials). Despite fewer samples from Cretaceous strata (11 Cretaceous localities versus 54 Paleocene localities), richness of dicotyledonous (dicot) leaf morphospecies from raw species counts at localities in the last ~100 ka of the Cretaceous (−18 to 0 m; 7 localities, 777 specimens, most speciose locality $n = 31$) and the first ~100 ka of the Paleocene (0 to 20 m; 6 localities, 1019 specimens, most speciose locality $n = 13$) indicates that earliest Paleocene dicot diversity was less than half that of the latest Cretaceous (fig. S6). Additionally, 46% of Cretaceous dicot leaf morphospecies that occur at more than one site do not occur in any of our Paleocene localities. A comparable study with similar time bins from the Williston Basin estimated

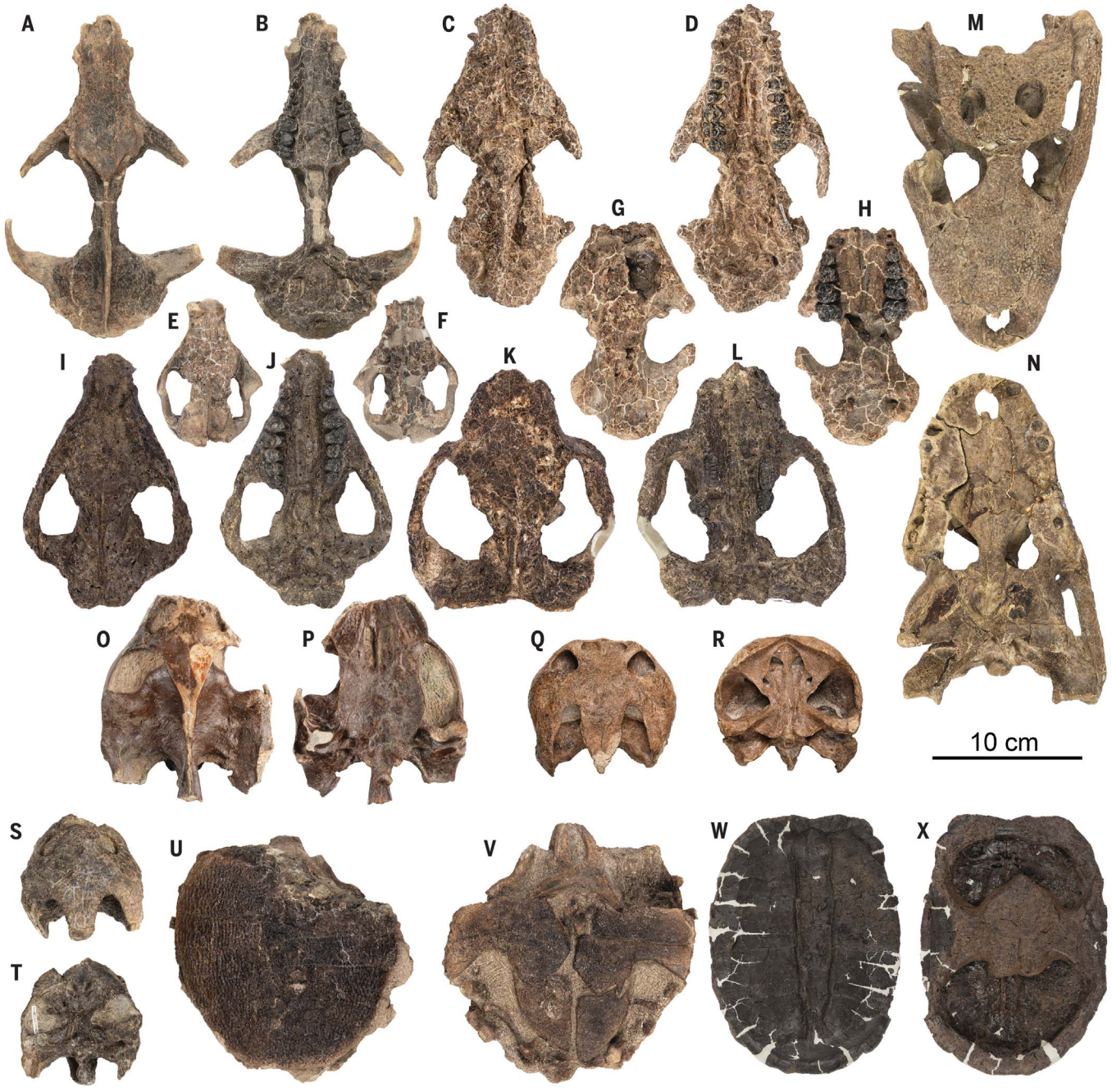


Fig. 2. Representative selection of vertebrate fossils. (**A** to **L**) Mammalian crania in dorsal and ventral views of *Eoconodon coryphaeus* [(A) and (B), DMNH.EPV.130976]; *Ectoconus ditrigonus* [(C) and (D), DMNH.EPV.130985]; *Loxolophus* sp. [(E) and (F), DMNH.EPV.132501]; juvenile *E. ditrigonus* [(G) and (H), DMNH.EPV.132515]; *Carsioptychus coarctatus* [(I) and (J), DMNH.EPV.95283]; and *Taeniolabis taoensis* [(K) and (L), DMNH.EPV.95284]. (**M** and **N**) Crocodilian cranium in dorsal and ventral view of cf. *Navajosuchus* [(M) and (N), DMNH.EPV.48541]. (**O** to **T**) Turtle crania in dorsal and ventral views of *Axestemys infernalis* [(O) and (P), DMNH.EPV.132514]; *Palatobaena* sp. [(Q) and (R), DMNH.EPV 134081]; and [(S) and (T)] *Cedrobaena putorius* (DMNH.EPV.130982). (**U** to **X**) Turtle shells in dorsal and ventral views of *Gilmoremys* sp. [(U) and (V), DMNH.EPV.95454] and *Hoplochelys* sp. [(W) and (X), DMNH.EPV.95453]. All crania and shells are shown to scale, except for (W) and (X), which are scaled 1:2 compared with the other specimens. Scale bar, 10 cm.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

57% extinction in dicot leaf morphospecies at the KPgE (*22*). Leaf mass per area (LMA), a proxy for carbon investment and ecological strategy in plants (*23*), decreased in both maximum and minimum values across the K–Pg boundary (Fig. 1 and fig. S7), which is consistent with a shift to faster growth strategies. Megafloral standing richness and LMA are lowest in the earliest Paleocene but exceed pre-KPgE levels within ~300 ka (Fig. 1 and fig. S7).

After the KPgE, many angiosperm clades diversified (*4*). The Corral Bluffs section preserves the oldest known occurrence of the Leguminosae, or bean family, represented by fossil seedpods and leaflets dated to 65.35 Ma (Fig. 3, J and K). The oldest previously recognized legume (*24*) is based on wood and leaflets (*25*) from early Paleocene rocks of Argentina (*26*), whereas the earliest legume seedpods are not recognized until the late Paleocene (~58 Ma) of Colombia (*27*). Our discovery supports (i) a nearly synchronous first appearance of legumes in North America and southern South America; (ii) a rapid diversification for the group in the earliest Paleocene (*24*); and (iii) their apparent origination in the Western Hemisphere.

Relative changes in leaf-estimated mean annual temperature (LMAT) (Fig. 1, fig. S6, and supplementary materials) from our section track paleotemperature proxies from sections elsewhere in the world. Corral Bluffs experienced a 4.6°C cooling [22.1° ± 2.7°C one standard error (1SE) to 17.5° ± 3.4°C 1SE] during the last ~100 ka of the Cretaceous, comparable to cooling estimates derived from LMAT (*28*) and carbonate-clumped isotopes (*29*) from the Williston Basin, and $\delta^{18}O$ of benthic foraminifera from the South Atlantic (*30*). For the first time, we corroborate (*31*) a warm interval immediately after the K–Pg boundary in a terrestrial section. Here, we observe that a 5.1°C warming event (17.5° ± 3.4°C 1SE to 22.6° ± 3.5°C 1SE) occurred from the K–Pg boundary through the first ~60 ka of the Paleocene, similar to the ~5°C in ~100 ka warming pulse inferred from $\delta^{18}O$ of phosphatic fish scales from the El Kef K–Pg section of Tunisia (*31*). A second ~150-ka interval (65.80–65.65 Ma) shows an initial warming of 2.2°C (21.1° ± 3.3°C 1SE to 23.3° ± 2.9°C 1SE) over ~30 ka, sustained temperatures for ~50 ka, and then 3.0°C cooling (22.7° ± 2.8°C 1SE to 19.7° ± 3.1°C 1SE) over ~70 ka at the top of magnetochron C29r. This event corresponds with the Danian C2 carbon isotopic excursion and inferred warming interval observed in marine (*32*) and terrestrial (*33*) strata. Sampling between these warming intervals is limited, and an alternative hypothesis is a general warming trend from the K–Pg boundary to the magnetochron C29r–C29n boundary. A third 2.9° to 3.2°C warming pulse (18.0° ± 3.3°C 1SE to 20.9° ± 3.0°C 1SE to 17.7° ± 3.5°C 1SE) over ~10 ka is tentatively recognized at ~700 ka post-KPgE.

Paleotemperature and ecosystem recovery

The timing of the early Paleocene warming intervals corresponds with changes in plant richness and taxonomic composition and, likely owing to additional food sources, coincident shifts in mammalian taxonomic composition, ecologic diversification, and expansion in the range of maximum mammalian body mass (Fig. 4). A mammalian taxonomic increase has been documented elsewhere in the Denver Basin, within the first 100 ka of the Paleocene, from 9 species found in the earliest Pu1 faunas to 21 species found in later Pu1 faunas (*34*, *35*). Maximum mammalian body mass increased through this interval to nearly pre-KPgE levels, from the largest known Lancian mammal (~8 kg) to the largest known Pu1 mammal (~6 kg), coincident with the first post-KPgE warming episode (Fig. 4 and figs. S8 and S9). The Pu1–Pu2 transition occurred ~300 ka after the KPgE and was marked by the appearance of varied and large (20+ kg) periptychid mammals. The appearance of larger-bodied periptychid mammals, particularly the herbivorous, hard-object feeder *C. coarctatus* (Fig. 2, I and J) (*36*, *37*), marks a notable dietary niche specialization in the earliest Paleocene, moving from the largely omnivorous or insectivorous diet found in Pu1 mammals (*38*) to a more herbivorous diet found in some Pu2 mammals. This dietary shift is correlated with a threefold increase in maximum mammalian body mass compared with Pu1 faunas (Figs. 1

Downloaded from http://science.sciencemag.org/ on November 28, 2019

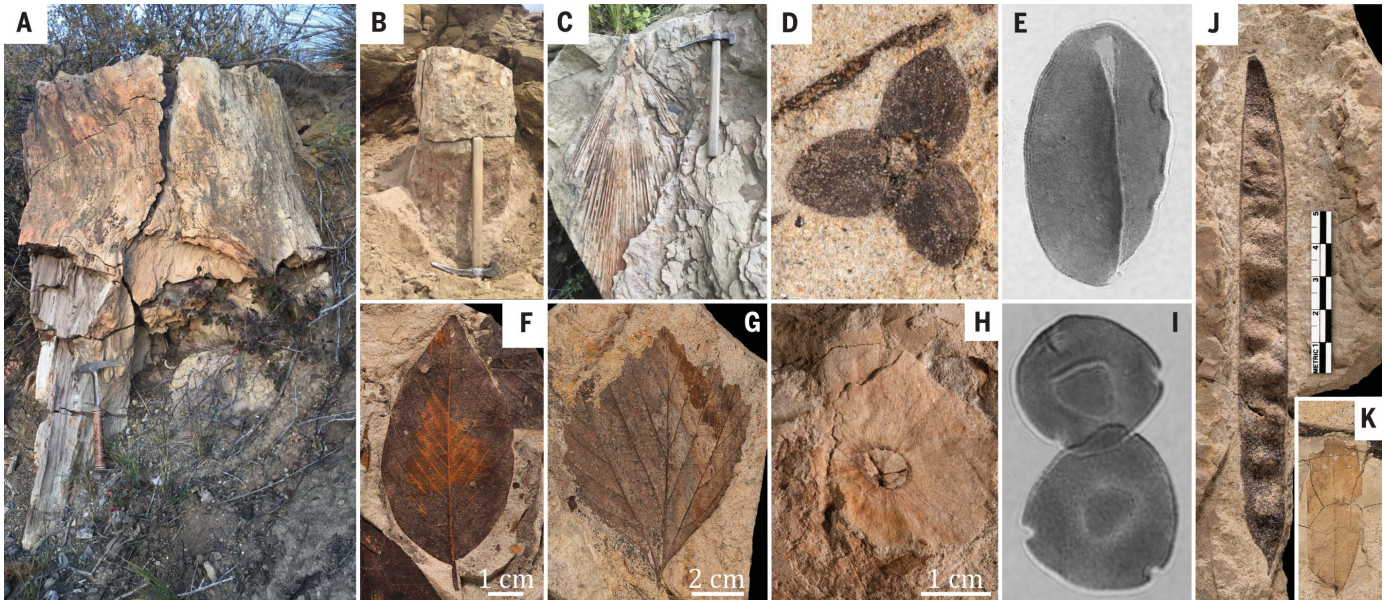


Fig. 3. Representative selection of plant fossils. (**A**) In situ tree stump. (**B** to **E**) Palm fossils, including (B) in situ stump, (C) frond, (D) flower (DMNH.EPI.45594), and (E) *Arecipites* sp. pollen grain. (**F** and **G**) Most common smooth and toothed dicot morphospecies: (F) *"Rhamnus" goldiana* (DMNH.EPI.52262) and (G) *Platanites marginata* (DMNH.EPI.23281). (**H** and **I**) Walnut family fruit and pollen: (H) *Cyclocarya* sp. (DMNH.EPI.52272) and (I) *Momipites tenuipolus* pollen grains preserved as a dyad. (**J** and **K**) Legume: (J) seedpod (DMNH.EPI.45540) and (K) leaflet (DMNH.EPI.45576). Rock hammer handle shown in [(A) to (C)] is 38 cm long; (D) flower is 5 mm wide; (E) pollen grain is 42 μm long; (I) each pollen grain has a 20-μm diameter; leaflet in (K) is scaled 2:1 compared with (J). Scale bar, 5 cm.

and 4 and figs. S8 and S9). The Pu1–Pu2 transition was coincident with the onset of a high plateau in megafloral standing richness, an increase of LMA beyond pre-KPgE levels, a doubling of the diversity of *Momipites* spp. [fossil juglandaceous (walnut family) pollen (Fig. 3I)], and the second early Paleocene warming interval (Figs. 1 and 4). The diversification of Juglandaceae taxa with small, winged seeds to later taxa with larger, wingless seeds is hypothesized to reflect a transition from wind to animal transport (*39*). This hypothesis is supported by the close correlation between diversification reflected in fossil juglandaceous pollen and the appearance of several large herbivorous periptychid mammals whose specialized and enlarged premolars are thought to be for hard-object feeding (*36*, *37*). Finally, the appearance of legumes co-occurred with a shift in maximum mammalian body mass and a tentatively recognized short warming pulse. Specifically, two large-bodied mammals appear at ~700 ka post-KPgE (Fig. 4)—the herbivorous multituberculate *T. taoensis* (~34 kg) and the omnivorous triisodontid archaic ungulate *Eoconodon coryphaeus* (~47 kg) (Fig. 2, A and B). These data suggest that earliest Paleocene warming pulses may have played an important role in post-KPgE ecosystem recovery, perhaps by facilitating immigration and/or in situ coevolution of flora and fauna.

The transition from an ecosystem characterized by a small-bodied mammalian fauna, "postdisaster" ferns, and low-diversity plant communities to one exhibiting a larger-bodied mammalian fauna and more ecologically and taxonomically complex forests mirrors modern postdisaster ecological successions, but on a much longer time scale (typically 10^4 to 10^5 years for recoveries from global mass extinctions versus 10^1 to 10^2 years for modern local-regional ecological recoveries) (*40*). The overall and long-term recovery we observe has recently been described as an aspect of "Earth system succession" (*40*). This concept proposes that global ecological succession after mass extinctions is intrinsically paced by the interactions of the biosphere and geosphere, both of which may be knocked out of equilibrium (*40*). The low-diversity, small-bodied mammalian fauna and low-diversity forests dominated by ferns and palms, often indicative of ecological disequilibrium, suggest that a period of ecosystem disequilibrium lasted for up to ~100 ka post-KPgE in our research area. A period of ecosystem recovery followed ~100 to 300 ka post-KPgE when megafloral diversity steadily increased. At ~300 ka post-KPgE, we see several additional signs of ecosystem recovery, including (i) the increase and then plateau of megafloral standing richness; (ii) LMA exceeding pre-KPgE levels; (iii) diversification of Juglandaceae, a potentially energy-rich food source for mammals; and (iv) the first substantial taxonomic diversification, dietary specialization (e.g., increased herbivory), and increase in maximum body mass of mammals (Pu1–Pu2). Finally, spatial niche partitioning, appearance of several additional large (30+ kg) mammals, and expansion of mammalian body mass disparity continue through ~700 ka at the Pu2–Pu3 boundary, all of which are further indications of ecosystem recovery. These changes are correlated with the arrival of plant taxa (e.g., legumes) that would have offered mammals new calorie-dense food sources. Taken together, our record places time estimates on the patterns of biotic recovery in Earth system succession and demonstrates that several aspects of ecosystem recovery occurred within ~300 ka post-KPgE (Fig. 3).

The pattern of warming pulses correlated with biotic change during the earliest Paleocene demonstrates a strong relationship between

Downloaded from http://science.sciencemag.org/ on November 28, 2019

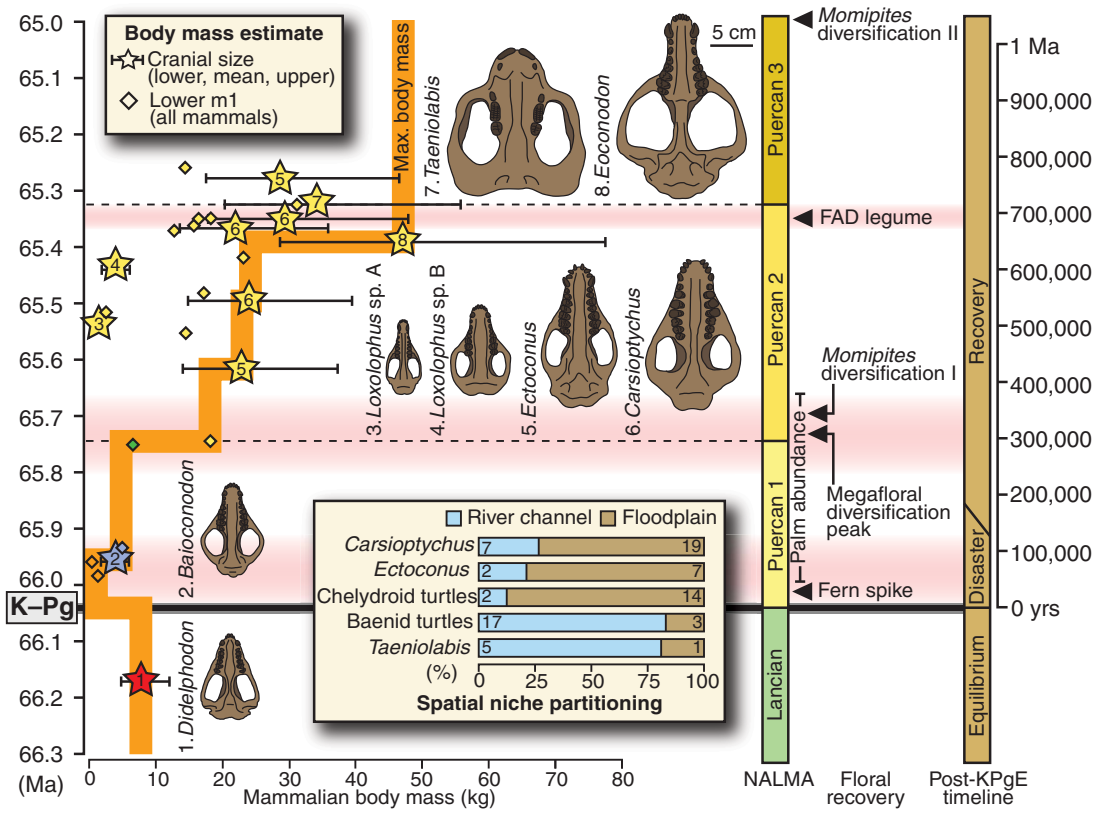


Fig. 4. Timeline of expansion of maximum body mass and niche space in earliest Paleocene mammals correlated with diversification and origination of key plant groups and warming intervals. Post-KPgE "disaster" ecosystems occur for <100 ka, ecosystem "recovery" occurs between ~100 and 300 ka, and overall post-KPgE ecosystem equilibrium occurs within ~300 ka. Mammalian body mass estimated on the basis of cranial and lower first molar dimensions of specimens recovered from Puercan 1 to Puercan 3 (Pu1–Pu3) intervals (data S13 and S14 and figs. S8 and S9). Data from Corral Bluffs study area (yellow) except for Pu1 mammals, which come from adjacent outcrops in the Denver Basin [West Bijou (orange), South Table Mountain (blue), and Alexander Locality (green)] and *Didelphodon* from North Dakota (red) (data S13 and S14 and supplementary materials). Not plotted is the distribution of other large (10 to 100+ kg) vertebrates (e.g., turtles, crocodilians, dinosaurs) found throughout the section (Fig. 1). Horizontal pink shading represents hypothesized warming intervals interpreted from LMAT. Niche partitioning graph showing environmental distribution of vertebrate groups (data S12): *Carsioptychus*, *Ectoconus*, and chelydroid turtles predominantly associated with floodplain and ponded water facies; baenid turtles and *Taeniolabis* predominantly in river channel complexes and proximal to medial crevasse splay facies. FAD, first appearance datum; m1, first molar tooth.

the biosphere and geosphere. The Deccan Traps of the Indian subcontinent represent repeated and voluminous volcanic eruptions (>10^6 km^3 of magma) during the post-KPgE Earth system succession (*6*, *7*). These eruptions might have induced warming pulses via the release of greenhouse gases (e.g., CO_2) (*7*). Recent work on the timing of these eruptions (*6*, *7*) places ~70% of the total volume within the 300-to-400-ka window roughly coincident with the earliest Paleocene warming pulse(s) observed at Corral Bluffs and the temporally correlated shifts in biotic recovery (Figs. 1 and 4). Although not a feedback of the biosphere-geosphere system, Deccan eruptions likely influenced atmospheric chemistry, in turn shaping Earth system succession and post-KPgE ecosystem recovery (Fig. 4). Detailed records of post–mass extinction biotic recovery, such as the one presented here, will provide a critical framework for predicting ecosystem recovery after mass extinction events, including the one we currently face (*41*).

REFERENCES AND NOTES

1. D. M. Raup, J. J. Sepkoski Jr., *Science* **215**, 1501–1503 (1982).
2. M. A. O'Leary *et al*., *Science* **339**, 662–667 (2013).
3. R. O. Prum *et al*., *Nature* **526**, 569–573 (2015).
4. S. Magallón, L. L. Sánchez-Reyes, S. L. Gómez-Acevedo, *Ann. Bot.* **123**, 491–503 (2019).
5. P. Schulte *et al*., *Science* **327**, 1214–1218 (2010).
6. B. Schoene *et al*., *Science* **363**, 862–866 (2019).
7. C. J. Sprain *et al*., *Science* **363**, 866–870 (2019).
8. J. D. Archibald *et al*., *Science* **328**, 973–973, author reply 975–976 (2010).
9. D. A. Pearson, T. Schaefer, K. R. Johnson, D. J. Nichols, J. P. Hunter, in *The Hell Creek Formation and the Cretaceous-Tertiary Boundary in the Northern Great Plains: An Integrated Continental Record of the End of the Cretaceous*, J. H. Hartman, K. R. Johnson, D. J. Nichols, Eds. (Geological Society of America Special Paper, vol. 361, 2002), pp. 145–167.
10. G. P. Wilson, in *Through the End of the Cretaceous in the Type Locality of the Hell Creek Formation in Montana and Adjacent Areas*, G. P. Wilson, W. A. Clemens, J. R. Horner, J. H. Hartman, Eds. (Geological Society of America Special Paper, vol. 503, 2014), pp. 365–392.
11. D. J. Nichols, K. R. Johnson, *Plants and the K-T Boundary* (Cambridge Univ. Press, 2008).
12. D. L. Lofgren, J. A. Lillegraven, W. A. Clemens, P. D. Gingrich, T. E. Williamson, in *Late Cretaceous and Cenozoic Mammals of North America: Biostratigraphy and Geochronology*, M. O. Woodburne, Ed. (Columbia Univ. Press, 2004), pp. 43–105.
13. K. R. Johnson, in *The Hell Creek Formation and the Cretaceous-Tertiary Boundary in the Northern Great Plains: An Integrated Continental Record of the End of the Cretaceous*, J. H. Hartman, K. R. Johnson, D. J. Nichols, Eds. (Geological Society of America Special Paper, vol. 361, 2002) pp. 329–391.
14. R. G. Raynolds, K. R. Johnson, *Rocky Mt. Geol.* **38**, 171–181 (2003).
15. J. J. Eberle, *Rocky Mt. Geol.* **38**, 143–169 (2003).
16. T. E. Williamson, *The Beginning of the Age of Mammals in the San Juan Basin, New Mexico: Biostratigraphy and Evolution of Paleocene Mammals of the Nacimiento Formation* (New Mexico Museum of Natural History and Science Bulletin vol. 8, 1996).
17. J. A. Lillegraven, J. J. Eberle, *J. Paleontol.* **73**, 691–710 (1999).
18. W. C. Clyde, J. Ramezani, K. R. Johnson, S. A. Bowring, M. M. Jones, *Earth Planet. Sci. Lett.* **452**, 272–280 (2016).
19. M. Dechesne, R. G. Raynolds, P. E. Barkmann, K. J. Johnson, "OF-11-01 Notes on the Denver Basin Geologic Maps: Bedrock Geology, Structure, and Isopach Maps of the Upper Cretaceous to Paleogene Strata between Greeley and Colorado Springs, Colorado," Groundwater, Open File Report, Colorado Geological Survey, Division of Minerals and Geology, Department of Natural Resources (2011).
20. A. J. Fuentes *et al*., Constructing a timescale of biotic recovery across the Cretaceous–Paleogene boundary, Corral Bluffs, Denver Basin, Colorado. bioRxiv 636951 [Preprint]. 15 May 2019. https://doi.org/10.1101/636951.
21. S. S. Hopkins, in *Methods in Paleoecology: Reconstructing Cenozoic Terrestrial Environments and Ecological Communities*, D. A. Croft, D. F. Su, S. W. Simpson, Eds. (Springer, 2018), pp. 7–22.
22. P. Wilf, K. R. Johnson, *Paleobiology* **30**, 347–368 (2004).
23. B. Blonder, D. L. Royer, K. R. Johnson, I. Miller, B. J. Enquist, *PLOS Biol.* **12**, e1001949 (2014).
24. M. Lavin, P. S. Herendeen, M. F. Wojciechowski, *Syst. Biol.* **54**, 575–594 (2005).
25. M. Brea, A. B. Zamuner, S. D. Matheos, A. Iglesias, A. F. Zucol, *Alcheringa* **32**, 427–441 (2008).
26. E. E. Comer *et al*., *Palaios* **30**, 553–573 (2015).
27. S. L. Wing *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18627–18632 (2009).
28. P. Wilf, K. R. Johnson, B. T. Huber, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 599–604 (2003).
29. T. S. Tobin, G. P. Wilson, J. M. Eiler, J. H. Hartman, *Geology* **42**, 351–354 (2014).
30. J. S. K. Barnet *et al*., *Geology* **46**, 147–150 (2017).
31. K. G. MacLeod, P. C. Quinton, J. Sepúlveda, M. H. Negra, *Science* **360**, 1467–1469 (2018).
32. F. Quillévéré, R. D. Norris, D. Kroon, P. A. Wilson, *Earth Planet. Sci. Lett.* **265**, 600–615 (2008).
33. I. Gilmour *et al*., *Geology* **41**, 783–786 (2013).
34. E. L. Dahlberg, J. J. Eberle, J. J. Sertich, I. M. Miller, *Rocky. Mt. Geo.* **51**, 1–22 (2016).
35. M. D. Middleton, E. W. Dewar, in *Paleogene Mammals*, S. G. Lucas, K. E. Zeigler, P. E. Kondrashov, Eds. (New Mexico Museum of Natural History and Science Bulletin, vol. 26, 2004), pp. 59–80.
36. J. D. Archibald, in *Evolution of Tertiary Mammals of North America: Volume 1: Terrestrial Carnivores, Ungulates, and Ungulate Like Mammals*, C. M. Janis, K. M. Scott, L. L. Jacobs, Eds. (Cambridge Univ. Press, 1998), pp. 292–331.
37. K. D. Rose, *The Beginning of the Age of Mammals* (Johns Hopkins Univ. Press, 2006).
38. E. W. Dewar, *PaleoBios* **23**, 1–19 (2003).
39. B. H. Tiffney, *Ann. Mo. Bot. Gard.* **71**, 551–576 (1984).
40. P. Hull, *Curr. Biol.* **25**, R941–R952 (2015).
41. A. D. Barnosky *et al*., *Nature* **471**, 51–57 (2011).

ACKNOWLEDGMENTS

We thank Norwood Properties, City of Colorado Springs, Waste Management, Aztec Family Raceway, J. Hawkins, J. Hilaire, J. Carner, W. Pendleton, the Bishop family, and H. Kunstle for land access; the State of Colorado, Office of the State Archaeologist, for issuing collection permits; J. Alicia, S. Begin, J. P. Cavigelli, H. Cochard, J. Englehorn, J. Groenke, F. Koether, L. Lacey, A. Lujan, B. Masek, B. Pittman, and N. Toth for preparation of specimens; K. Getty, R. Hess, R. Lavie, S. Milito, Y. Rollot, P. Sullivan, J. Wyman, and L. Taylor for field assistance; F. Cochard, L. Dougan, S. Luallin, R. Wicker, J. Wood, and the USGS National Unmanned Aircraft Systems team for photography; K. MacKenzie and N. Neu-Yagle for collections assistance; and B. Snellgrove for logistics. **Funding:** Funding was provided by The Lisa Levin Appel Family Foundation, M. Cleworth, Lyda Hill Philanthropies, David B. Jones Foundation, M. L. and S. R. Kneller, T. and K. Ryan, and J. R. Tucker as part of the Denver Museum of Nature & Science (DMNS) No Walls Community Initiative. **Author contributions:** T.R.L. and I.M.M. led the project. T.R.L. wrote and edited the manuscript in collaboration with I.M.M., A.D.B., K.W., and S.G.B.C. were primary project participants. All authors collected and/or analyzed data and samples, interpreted results, and edited the manuscript. **Competing interests:** None declared. **Data and materials availability:** Fossil specimens and pollen slides are all deposited at the DMNS. All data are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/977/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Table S1
References (*42*–*123*)
Data S1 to S14

3 June 2019; accepted 15 October 2019
Published online 24 October 2019
10.1126/science.aay2268

Downloaded from http://science.sciencemag.org/ on November 28, 2019

REPORT

2D MAGNETISM

Direct observation of van der Waals stacking–dependent interlayer magnetism

Weijong Chen[1], Zeyuan Sun[1], Zhongjie Wang[1], Lehua Gu[1], Xiaodong Xu[2], Shiwei Wu[1,3]*, Chunlei Gao[1,3]*

Controlling the crystal structure is a powerful approach for manipulating the fundamental properties of solids. In van der Waals materials, this control can be achieved by modifying the stacking order through rotation and translation between the layers. Here, we observed stacking-dependent interlayer magnetism in the two-dimensional (2D) magnetic semiconductor chromium tribromide ($CrBr_3$), which was enabled by the successful growth of its monolayer and bilayer through molecular beam epitaxy. Using in situ spin-polarized scanning tunneling microscopy and spectroscopy, we directly correlate the atomic lattice structure with the observed magnetic order. Although the individual monolayer $CrBr_3$ is ferromagnetic, the interlayer coupling in bilayer depends on the stacking order and can be either ferromagnetic or antiferromagnetic. Our observations pave the way for manipulating 2D magnetism with layer twist angle control.

The type of van der Waals (vdW) stacking has been extensively recognized as a critical component in determining the properties of layered vdW materials. In particular, weak interlayer vdW interactions allow for the control over the rotational and translational degrees of freedom between layers, creating a host of new materials with distinct stacking symmetries and functionalities (*1*–*3*). Whereas most previous work has focused on the electronic and optical properties associated with the vdW stacking (*4*–*7*), the recent discovery of magnetism in two-dimensional (2D) materials, achieved through both mechanical exfoliation (*8*–*13*) and molecular beam epitaxy (MBE) (*14*, *15*), provides an exciting opportunity to explore the effects of stacking order on a material's magnetic properties.

Among the 2D magnetic materials discovered so far, the family of chromium trihalides, CrX_3 (where X = Cl, Br, or I), has garnered special interest (*8*, *16*). For instance, ferromagnetism perpendicular to the 2D plane persists in monolayer CrI_3, whereas its bilayer exhibits interlayer antiferromagnetism. These magnetic structures lead to a number of emerging phenomena, such as giant tunneling magnetoresistance in spin-filter magnetic tunnel junctions (*17*–*19*), the electrical control of 2D magnetism (*20*–*22*), and giant nonreciprocal optical second-harmonic generation (*23*). In contrast to CrI_3, recent tunneling measurements suggest that the interlayer coupling in atomically thin $CrBr_3$ is ferromagnetic (*24*). This was confirmed through polar reflectance magnetic circular dichroism (RMCD) measurements of mechanically exfoliated bilayers of $CrBr_3$, which reveal magnetic hysteresis behavior centered around zero applied magnetic field, distinct from the three-step staircase pattern observed in bilayer CrI_3 (fig. S1). Given that CrI_3 and $CrBr_3$ are isostructural (*25*), such a substantial difference in the magnetic properties between their bilayers prompts a thorough investigation of the mechanisms governing the interlayer magnetic coupling (*26*).

Here, we used in situ spin-polarized scanning tunneling microscopy and spectroscopy (*27*, *28*) to establish a direct correlation between the interlayer magnetic coupling and the stacking structures in bilayer $CrBr_3$. The $CrBr_3$ films were grown on freshly cleaved, highly oriented pyrolytic graphite (HOPG) substrates by MBE. During the growth, the sample surface was monitored in situ by reflection high-energy electron diffraction (RHEED). Figure 1, A and B, respectively, show the RHEED patterns before and after growing $CrBr_3$ with the diffraction orders indicated. The stripe-like RHEED pattern from the $CrBr_3$ indicates the formation of a 2D crystalline thin film, confirmed by scanning tunneling microscopy (STM) to be a monolayer $CrBr_3$ (Fig. 1C). Further deposition of $CrBr_3$ led to the appearance of bilayer $CrBr_3$ islands (Fig. 1D) [for details about the MBE growth, see (*29*)].

Figure 1E shows the atomically resolved STM image of the monolayer $CrBr_3$, revealing periodically spaced triangular cluster protrusions. To understand this image, we consider the monolayer crystal structure of $CrBr_3$. As shown in Fig. 1F, Cr atoms are arranged in a honeycomb lattice structure, and each atom is surrounded by an octahedron of six Br atoms (*16*). Within a single honeycomb formed by six Cr atoms, there are three Br atoms at the top and bottom surfaces, marked by solid and dotted green triangles with opposite orientations, respectively. Thus, the monolayer $CrBr_3$ has three-fold rotational symmetry, with the rotational axis across either the Cr atoms or the center of the honeycomb. By overlaying the atomic structure on top of the STM image in Fig. 1E, we determined that the observed triangular cluster is formed by the three top Br atoms (solid green triangle in Fig. 1E). The measured in-plane lattice constant is 6.3 Å, consistent with the bulk lattice constant (6.26 Å) (*25*). The monolayer thickness is determined to be 6.5 Å by atomic force microscopy (AFM), as shown in Fig. 1G. Structural domains and their boundaries are also clearly seen in the image. Both the large-scale topography and the atomically resolved STM images demonstrate the high-quality growth of the monolayer $CrBr_3$ films.

To measure the magnetic properties, we performed in situ spin-polarized STM measurements with an applied out-of-plane magnetic field at 5 K [for spin-polarized STM measurements, see (*29*)]. We used antiferromagnetic Cr-coated tungsten tips in which magnetization at the tip apex remained stable while the external magnetic field was swept (*30*). Figure 2A shows the tunneling spectra (dI/dV) from a monolayer $CrBr_3$ with oppositely applied fields (B = ±0.3 T), measured on the center of a Br triangular cluster. Although both dI/dV curves show semiconducting behavior near zero bias voltage, they clearly differ at some bias voltage V_b, which results from the magnetization flip of the monolayer $CrBr_3$ under the reversal of the external magnetic field. To further confirm the existence of ferromagnetism, we measured a series of dI/dV spectra while sweeping the magnetic field back and forth. Figure 2B plots the hysteresis loop at V_b = 1.4 V. Additional data at different bias voltages are shown in fig. S2. A rectangular hysteresis loop was observed with a coercive field of ~30 mT, with the two dI/dV plateaus corresponding to different magnetization states (spin up or spin down). This observation suggests that epitaxial $CrBr_3$ monolayers grown on HOPG are semiconducting ferromagnets with a well-defined, out-of-plane easy axis. The dI/dV plateaus have sensitive dependence on bias voltages, atomic sites, and Cr tips (see figs. S2 and S3) because the dI/dV signal is determined by the spin- and energy-dependent local density of states of the tip and the sample (*29*).

[1]State Key Laboratory of Surface Physics, Key Laboratory of Micro and Nano Photonic Structures (MOE), Department of Physics, and Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China. [2]Department of Physics and Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, USA. [3]Collaborative Innovation Center of Advanced Microstructures, Nanjing 210093, China.
*Corresponding author. Email: clgao@fudan.edu.cn (C.G.); swwu@fudan.edu.cn (S.W.)

Downloaded from http://science.sciencemag.org/ on November 28, 2019

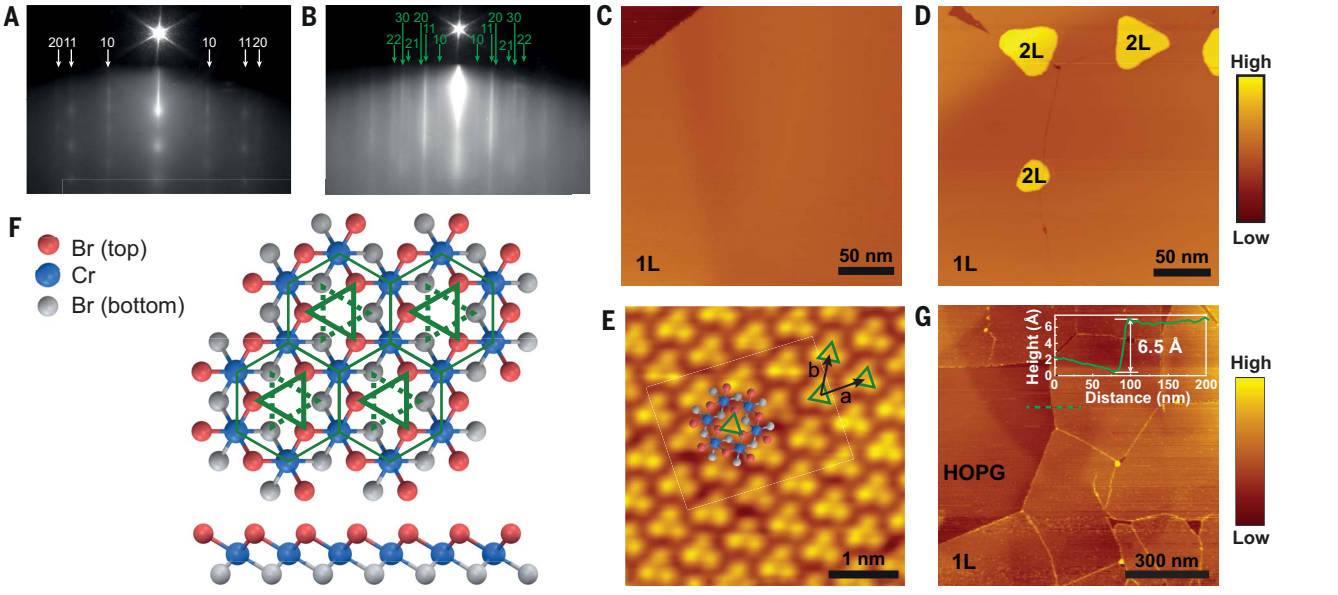


Fig. 1. MBE Growth of $CrBr_3$ monolayer and bilayer on HOPG. (**A** and **B**) RHEED patterns with indicated diffraction orders of (A) the bare HOPG substrate and (B) the MBE-grown $CrBr_3$ film. (**C** and **D**) STM images of (C) the $CrBr_3$ monolayer with (D) bilayer islands. The scan parameters were as follows: V_b = 1.1 V, I = 100 pA, T = 5 K for (C) and V_b = 1.5 V, I = 100 pA, T = 5 K for (D). (**E**) Atomically resolved image of a monolayer $CrBr_3$ with an overlaid atomic structure. The scan parameters were as follows: V_b = 1.5 V, I = 500 pA, T = 5 K. The lattice constants were determined to be 6.3 Å for the primitive vectors **a** and **b**, consistent with the bulk values. (**F**) Illustrations of the top and side views of the monolayer $CrBr_3$ atomic structure. The Cr atoms form a honeycomb lattice sandwiched by Br atoms. Within the Cr honeycomb lattice, the top and bottom surfaces of Br atoms form single triangles but with opposite orientation, indicated by solid and dotted green lines, respectively. (**G**) AFM image of monolayer $CrBr_3$ with partial coverage. A line-cut profile across the monolayer and bare substrate is shown with a monolayer height of ~6.5 Å.

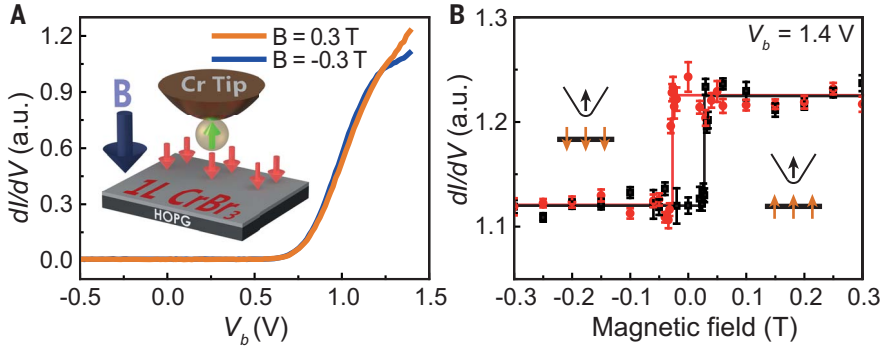


Fig. 2. Spin-polarized tunneling of monolayer $CrBr_3$. (**A**) Spin-polarized tunneling spectra under positive and negative out-of-plane magnetic fields (±0.3 T). The inset illustrates the experimental geometry. The magnetization at the tip apex is assumed to be spin up, and the same assumption applies to the following figures [see discussion in (*29*)]. (**B**) *dI/dV* signal as a function of the magnetic field. V_b was fixed at 1.4 V. The out-of-plane magnetic field was swept upward (black data) and downward (red data). The ferromagnetic hysteresis loop is outlined as rectangular solid lines. Insets sketch the two configurations of the magnetization alignment between the Cr tip and the monolayer $CrBr_3$ film. The in-plane component of magnetization at the Cr tip apex, if any, does not contribute to the magnetic contrast in *dI/dV*.

With the atomic structure of monolayer $CrBr_3$ determined and its ferromagnetism confirmed, we now focus on bilayer $CrBr_3$. In general, there are two types of stacking structures with the unit cell size the same as that of the monolayer: the R-type, with both layers aligned to the same orientation, and the H-type, with a 180° rotation between the layers. For R-type stacking, two polytypes for $CrBr_3$ bulk crystals were reported: rhombohedral at low temperature and monoclinic at high temperature (>420 K) (*16*). These two structures mainly differ in the relative translation between the two monolayers. However, common to vdW materials (*4*, *5*, *31*), the energy differences between different stacking structures can be relatively small, even if the stacking structure is absent in bulk crystal. Therefore, the possible stacking structures in the bilayer case could be far more complex than the two reported in the literature. By only considering the high-symmetry configurations (i.e., hollow, bridge, and top sites) between the halogen atoms at the interface of adjacent layers, there are already six and five polytypes for R-type stacking (table S1) and H-type stacking (table S2), respectively.

We found that the H- and R-type stacking structures can both form in the MBE-grown bilayers and they give rise to distinct interlayer magnetic coupling. The results corresponding to the H-type stacking structure are given in Fig. 3. Figure 3A shows the STM image across the step edge between a bilayer $CrBr_3$ island and a monolayer $CrBr_3$ film that is an extension of the bilayer's bottom layer. Their atomically resolved STM images are shown in Fig. 3, B and C, which show individual triangular clusters formed by the top three Br atoms in the unit cell of $CrBr_3$, similar to those in Fig.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

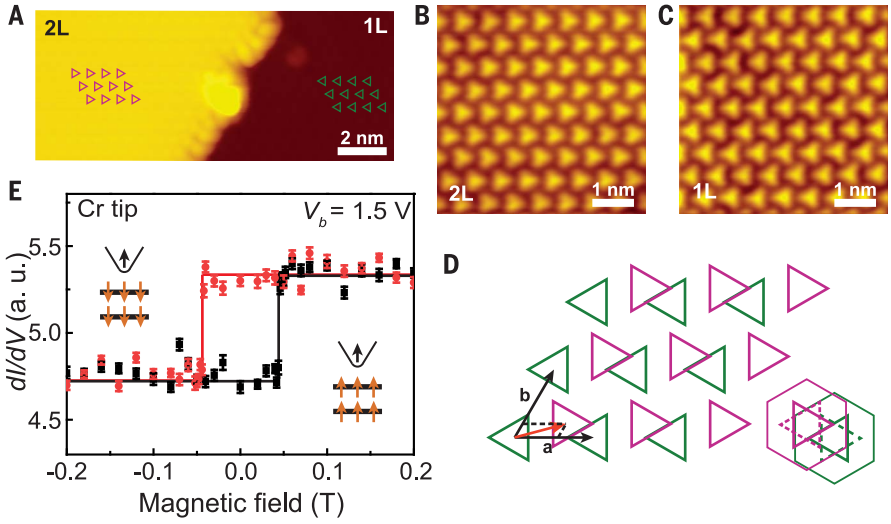


Fig. 3. Interlayer ferromagnetic coupling in an H-type stacked bilayer $CrBr_3$. (**A**) STM image of a $CrBr_3$ film with both a monolayer (1L) region and a bilayer (2L) island. (**B** and **C**) Magnified, atomically resolved images of (B) the bilayer region and (C) its extended bottom monolayer at V_b = 1.9 V, indicating that the top and bottom layers in the bilayer are anti-aligned, or rotated by 180° (H-type stacking). (**D**) Atomic structure of the bilayer $CrBr_3$, as determined from atomically resolved STM images (for further details, see fig. S4). The unit cells of the top and bottom layers are represented by the magenta and green solid triangles, respectively, corresponding to the top surface of Br atoms in each monolayer sheet. These magenta and green solid triangles are also overlaid on the monolayer and bilayer in (A). The unit cell of the top layer (magenta) is a translation by 0.55**a** + 0.20**b** of the bottom layer (green). For comparison with the structures in table S2, the stacking structure is also shown with the bottom surface of Br atoms of each monolayer sheet as dotted triangles and the Cr atoms as solid hexagons. (**E**) Spin-polarized tunneling on the bilayer $CrBr_3$ as a function of magnetic field with a Cr tip at V_b = 1.5 V. The out-of-plane magnetic field was swept upward (black data) and downward (red data). Like that of the monolayer $CrBr_3$, a rectangular ferromagnetic hysteresis loop was observed with a coercive field of ~45 mT. Insets depict two configurations of the magnetization alignment between the tip and the sample.

1E. The orientation of the triangular clusters from the monolayer region, which has the same crystal structure as the bottom layer of the bilayer, is exactly opposite to the top layer in the bilayer, suggesting H-type stacking.

To determine the translational degree of freedom between the two layers, we overlaid the two $CrBr_3$ lattices, represented by the magenta and green triangles, on the top and bottom layers of the bilayer island shown in Fig. 3A (for further details, see fig. S4). We found that the top layer is translated to 0.55**a** + 0.20**b** with respect to the bottom layer (Fig. 3D). Given that the experimental uncertainty of translation is ~±7% of the lattice constants **a** and **b**, the closest stacking structure is "bridge I" of H-type stacking listed in table S2. We also observed the equivalent bridge I structure, in which the top layer is translated to 0.76**a** + 0.38**b** of the bottom layer (fig. S5). In terms of both rotational and translational degrees of freedom, this observed stacking structure is distinct from the reported structure models for CrX_3 in the literature (*16*, *25*). Further spin-polarized STM measurement on this H-type stacked bilayer $CrBr_3$ showed a rectangular magnetic hysteresis loop with a coercive magnetic field of ~45 mT (Fig. 3E for V_b = 1.5 V and fig. S6 for other bias voltages), similar to the hysteresis observed in $CrBr_3$ monolayers (Fig. 2B). Thus, in H-type stacked bilayer $CrBr_3$, the interlayer coupling is ferromagnetic.

The second kind of stacking structure is depicted in Fig. 4, where the triangular clusters from the adjacent monolayer and bilayer regions are oriented in the same direction (Fig. 4, A to C), suggesting R-type stacking. The translational degree of freedom between the two layers is determined in the same way as in Fig. 3. In R-type stacking, the top layer is translated to 0.48**a** + 0.48**b** with respect to the bottom layer (Fig. 4D; for further analysis details, see fig. S7). In one R-type–stacking sample, we also observed that the top layer was translated to 0.21**a** + 0.78**b** of the bottom layer (fig. S8). Examining their stacking structures, we found that they are close to the structures named "special" and "bridge I" listed in table S1. As illustrated in fig. S9, both structures share the same lattice symmetry (C_{2h}) as the monoclinic structure (*16*), but with a translation along the mirror plane.

For the R-type stacked bilayer, we observed four plateaus in the hysteresis loop when a Cr tip was used (Fig. 4E and fig. S10), distinct from the two plateaus observed on the $CrBr_3$ monolayer and the H-type stacked bilayer. This four-plateau behavior suggests that the R-type stacked bilayer is antiferromagnetically coupled in its ground states, giving rise to two additional magnetization configurations. Above +0.5 T or below −0.5 T, the magnetization direction of the bilayer $CrBr_3$ is aligned to the same direction by the external magnetic field. As in the ferromagnetic $CrBr_3$ monolayer and bilayer, the parallel or antiparallel alignment between the tip and the sample yields two different dI/dV plateaus. Within ±0.5 T, the individual ferromagnetic layers are antiferromagnetic coupled, which creates an additional magnetic tunnel junction across the vdW gap. Depending on how the spin direction of the top layer aligns with the Cr tip, two new magnetization configurations are formed (inset of Fig. 4E).

The interlayer antiferromagnetism in R-type stacked bilayer $CrBr_3$ is further corroborated with a nonmagnetic W tip under external magnetic field. Figure 4F shows such a measurement on another R-type stacked bilayer at V_b = 1.5 V (see fig. S11 for data with different bias voltages). Starting with zero applied magnetic field, sweeping the external magnetic field in either direction results in a plateau in the dI/dV curve. Beyond magnetic fields of ±0.5 T, the dI/dV abruptly changes (Fig. 4F). Although the relative magnitude of dI/dV plateaus depends on bias voltage, the two-plateau behavior demonstrates the magnetic field–driven transition from antiferromagnetic to ferromagnetic interlayer coupling. Such tunneling magnetoresistance behavior was recently discovered in spin-filter magnetic tunnel junctions using mechanically exfoliated bilayer CrI_3 (*17*–*19*). By contrast, when a nonmagnetic W tip was used for H-type stacked bilayers that were ferromagnetically coupled, no magnetic field–dependent dI/dV features at any bias voltage could be observed (fig. S12). In addition to the dI/dV magnetic contrast, the antiferromagnetic ground state in this R-type stacked bilayer could be captured in the contrast of STM height driven under external magnetic field, as shown in fig. S13.

The distinct interlayer magnetism, from antiferromagnetic coupling in R-type stacking to ferromagnetic coupling in H-type stacking, in our MBE-grown bilayer $CrBr_3$ clearly demonstrates the substantial tunability of 2D magnetism through the stacking order. The interlayer coupling in bilayer $CrBr_3$ is mediated through a superexchange interaction, which is controlled by the directional hybridization

Downloaded from http://science.sciencemag.org/ on November 28, 2019

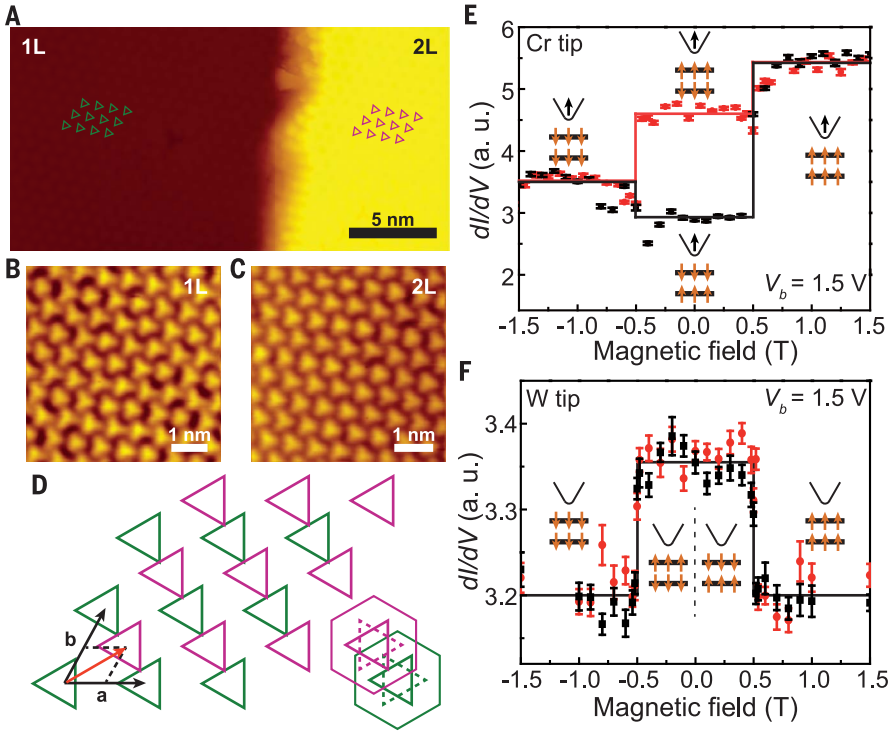


Fig. 4. Interlayer antiferromagnetic coupling in an R-type stacked bilayer $CrBr_3$. (**A**) STM image of a $CrBr_3$ film with both a monolayer (1L) region and a bilayer (2L) island. (**B** and **C**) Atomically resolved images of (B) the monolayer and (C) the bilayer. V_b = 1.9 V. The stacking configuration in the bilayer is identified as R-type, i.e., the top and bottom layers are in the same orientation. (**D**) Atomic structure of the bilayer $CrBr_3$, as determined from atomically resolved STM images (see fig. S7 for further details). The representation style follows that in Fig. 3. The unit cell of the top layer (magenta) is translated by 0.48**a** + 0.48**b** from the bottom layer (green). (**E**) Spin-polarized tunneling on an R-type stacked bilayer $CrBr_3$ with a Cr tip at V_b = 1.5 V. The insets show four magnetization configurations, including the Cr tip and bilayer $CrBr_3$, corresponding to different magnetic field–dependent *dI*/*dV* plateaus. The out-of-plane magnetic field was swept upward (black data) and downward (red data). (**F**) Spin-dependent tunneling on the bilayer $CrBr_3$ in (C) with a nonmagnetic W tip at V_b = 1.5 V. Abrupt decrease of the *dI*/*dV* signal was observed at magnetic fields of ~±0.5 T, suggesting an interlayer antiferromagnetic coupling within ±0.5 T.

between the Br p-orbitals and Cr d-orbitals (*32*, *33*). Because the bond angles and the bond distances of the Cr–Br–Br–Cr exchange path are strongly dependent on the stacking order, the interlayer magnetism is expected to also depend on the specific stacking structure, including the interlayer distance and atomic site positions. In fact, recent theoretical studies have investigated the interlayer magnetism of the rhombohedral and monoclinic structures in bilayer CrI_3 (*34*–*37*).

Although the exact growth mechanism remains to be investigated, our observations illustrate the polytypism in vdW materials and its importance to 2D magnetism. Our work also calls for a close examination for the stacking structures in mechanically exfoliated CrX_3 samples, as the exfoliated bilayer CrI_3 and $CrBr_3$ exhibit distinct interlayer magnetic coupling. We envision that this working principle could be used to manipulate 2D magnetism, for example, by engineering spatially dependent spin textures uniquely enabled by twisted bilayers and heterostructures (*38*).

REFERENCES AND NOTES

1. A. K. Geim, I. V. Grigorieva, *Nature* **499**, 419–425 (2013).
2. K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. Castro Neto, *Science* **353**, aac9439 (2016).
3. R. Ribeiro-Palau *et al.*, *Science* **361**, 690–693 (2018).
4. W. Bao *et al.*, *Nat. Phys.* **7**, 948–952 (2011).
5. C. H. Lui, Z. Li, K. F. Mak, E. Cappelluti, T. F. Heinz, *Nat. Phys.* **7**, 944–947 (2011).
6. Y. Shan *et al.*, *Sci. Adv.* **4**, eaat0074 (2018).
7. T. Jiang *et al.*, *Nat. Nanotechnol.* **9**, 825–829 (2014).
8. B. Huang *et al.*, *Nature* **546**, 270–273 (2017).
9. C. Gong *et al.*, *Nature* **546**, 265–269 (2017).
10. Z. Fei *et al.*, *Nat. Mater.* **17**, 778–782 (2018).
11. Y. Deng *et al.*, *Nature* **563**, 94–99 (2018).
12. J. U. Lee *et al.*, *Nano Lett.* **16**, 7433–7438 (2016).
13. X. Wang *et al.*, *2D Mater.* **3**, 031009 (2016).
14. D. J. O'Hara *et al.*, *Nano Lett.* **18**, 3125–3131 (2018).
15. M. Bonilla *et al.*, *Nat. Nanotechnol.* **13**, 289–293 (2018).
16. M. A. McGuire, H. Dixit, V. R. Cooper, B. C. Sales, *Chem. Mater.* **27**, 612–620 (2015).
17. T. Song *et al.*, *Science* **360**, 1214–1218 (2018).
18. D. R. Klein *et al.*, *Science* **360**, 1218–1222 (2018).
19. Z. Wang *et al.*, *Nat. Commun.* **9**, 2516 (2018).
20. S. Jiang, J. Shan, K. F. Mak, *Nat. Mater.* **17**, 406–410 (2018).
21. B. Huang *et al.*, *Nat. Nanotechnol.* **13**, 544–548 (2018).
22. S. Jiang, L. Li, Z. Wang, K. F. Mak, J. Shan, *Nat. Nanotechnol.* **13**, 549–553 (2018).
23. Z. Sun *et al.*, *Nature* **572**, 497–501 (2019).
24. D. Ghazaryan *et al.*, *Nature Electronics* **1**, 344–349 (2018).
25. M. McGuire, *Crystals (Basel)* **7**, 121 (2017).
26. L. Thiel *et al.*, *Science* **364**, 973–976 (2019).
27. R. Wiesendanger, *Rev. Mod. Phys.* **81**, 1495–1550 (2009).
28. M. Bode, *Rep. Prog. Phys.* **66**, 523–582 (2003).
29. See supplementary materials for additional information.
30. R. Wiesendanger *et al.*, *J. Magn. Magn. Mater.* **272**, 2115–2120 (2004).
31. M. Chhowalla *et al.*, *Nat. Chem.* **5**, 263–275 (2013).
32. J. L. Lado, J. Fernández-Rossier, *2D Mater.* **4**, 035002 (2017).
33. W. B. Zhang, Q. Qu, P. Zhua, C. H. Lam, *J. Mater. Chem. C Mater. Opt. Electron. Devices* **3**, 12457–12468 (2015).
34. P. Jiang *et al.*, *Phys. Rev. B* **99**, 144401 (2019).
35. D. Soriano, C. Cardoso, J. Fernández-Rossier, *Solid State Commun.* **299**, 113662 (2019).
36. N. Sivadas, S. Okamoto, X. Xu, C. J. Fennie, D. Xiao, *Nano Lett.* **18**, 7658–7664 (2018).
37. S. W. Jang, M. Y. Jeong, H. Yoon, S. Ryee, M. J. Han, *Phys. Rev. Mater.* **3**, 031001(R) (2019).
38. Q. Tong, F. Liu, J. Xiao, W. Yao, *Nano Lett.* **18**, 7194–7199 (2018).
39. W. Chen *et al.*, Replication data for: Direct observation of van der Waals stacking-dependent interlayer magnetism, version 1.0, Zenodo (2019).

ACKNOWLEDGMENTS

We thank D. Xiao and T. Cao for insightful discussion, B. Huang for proofreading the paper, and D. Feng for lending the AFM. **Funding:** The work at Fudan was supported by the National Key Research and Development Program of China (grant nos. 2016YFA0300904 and 2016YFA0301002), the National Basic Research Program of China (grant no. 2014CB921601), and the National Natural Science Foundation of China (grant nos. 11427902 and 11674063). X.X. is supported by the Department of Energy, Basic Energy Sciences, Materials Sciences and Engineering Division (DE-SC0018171). **Author contributions:** C.G. and S.W. conceived and supervised the project. W.C. grew the sample and conducted the STM experiment with the assistance from Z.W. Z.S. conducted the RMCD measurement. L.G. and W.C. conducted the AFM measurement. W.C., Z.S., X.X., S.W., and C.G. analyzed the data and wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data discussed in this paper are available at Zenodo (*39*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/983/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 and S2
References (*40*–*42*)

23 August 2018; resubmitted 20 February 2019
Accepted 25 October 2019
10.1126/science.aav1937

Downloaded from http://science.sciencemag.org/ on November 28, 2019

CORRELATED ELECTRONS

Strange metallicity in the doped Hubbard model

Edwin W. Huang[1,2]*, Ryan Sheppard[1], Brian Moritz[2], Thomas P. Devereaux[2,3]*

Strange or bad metallic transport, defined by incompatibility with the conventional quasiparticle picture, is a theme common to many strongly correlated materials, including high-temperature superconductors. The Hubbard model represents a minimal starting point for modeling strongly correlated systems. Here we demonstrate strange metallic transport in the doped two-dimensional Hubbard model using determinantal quantum Monte Carlo calculations. Over a wide range of doping, we observe resistivities exceeding the Mott-Ioffe-Regel limit with linear temperature dependence. The temperatures of our calculations extend to as low as 1/40 of the noninteracting bandwidth, placing our findings in the degenerate regime relevant to experimental observations of strange metallicity. Our results provide a foundation for connecting theories of strange metals to models of strongly correlated materials.

Strongly correlated materials are renowned for their rich phase diagrams containing intertwined orders (*1*, *2*). In many of these materials, the high-temperature disordered phase has anomalous properties, such as the DC resistivity that exceeds the Mott-Ioffe-Regel (MIR) criterion with no sign of a crossover or saturation, signaling the absence of well-defined quasiparticles (*3*–*5*); these phases are referred to as strange or bad metals. For many such systems, the resistivity is also characterized by linear temperature dependence up to the highest experimentally accessible temperatures. The incompatibility of these behaviors with conventional Fermi liquid theory poses a fundamental challenge to our understanding of strongly correlated electron systems. In particular, for the long-standing problem of high-temperature superconductivity, it was recognized early on that transition temperatures in hole-doped cuprates are maximal where resistivity is most T-linear, suggesting that unconventional pairing is deeply connected to the nature of the strange metal.

The Hubbard model on a square lattice, containing only a local Coulomb interaction, is perhaps the most studied model of correlated electrons. Although motivated in part by its believed relevance to cuprate superconductors, the model is generically important to the theoretical understanding of strong correlation effects owing to its simple and plausibly realistic form. Lacking an analytical solution in two dimensions, the Hubbard model has been studied through a variety of numerical approaches focusing primarily on the nature of its ground state upon doping. Its transport properties remain relatively unexplored.

Here we demonstrate and study strange metallic transport in the normal state of the Hubbard model using determinantal quantum Monte Carlo (DQMC) calculations at finite temperatures (*6*, *7*) combined with series expansions at infinite temperature (*8*–*11*). The Hubbard model Hamiltonian is $H = -\sum_{ij\sigma} t_{ij} c^{\dagger}_{i\sigma} c_{j\sigma} + U \sum_{i} c^{\dagger}_{i\uparrow} c_{i\uparrow} c^{\dagger}_{i\downarrow} c_{i\downarrow}$, where $c^{\dagger}_{i\sigma}$ is the creation operator for an electron on site i with spin σ. The hopping energy t_{ij} between sites i and j equals t for nearest neighbors and t' for next nearest neighbors. We choose $t'/t = -0.25$ and an intermediate interaction strength $U/t = 6$, and simulate 8×8 square clusters with periodic boundaries. Our simulations encompass a range of hole dopings from $p = 0$ to $p = 0.3$ and temperatures down to $T/t = 0.2$, or 1/40 of the noninteracting bandwidth $W = 8t$.

Our principal results are based on DQMC measurements of the current-current correlation function $\Lambda(\tau) = \langle \mathbf{j}(\tau)\mathbf{j} \rangle$, where $\mathbf{j} = i \sum_{ij\sigma} t_{ij}(\mathbf{r}_i - \mathbf{r}_j) c^{\dagger}_{i\sigma} c_{j\sigma}$ is the current operator at momentum $\mathbf{q} = 0$ and τ is imaginary time. For the square clusters that we study, it is sufficient to consider only the xx component of $\Lambda(\tau)$. The optical conductivity $\sigma(\omega)$ relates to the imaginary time current-current correlation function through $\Lambda(\tau) = \int \frac{d\omega}{\pi} \frac{\omega e^{-\tau\omega}}{1-e^{-\beta\omega}} \sigma(\omega)$. We adopt the standard maximum entropy method of analytic continuation to extract the optical conductivity given DQMC measurements of the current correlator in imaginary time (*12*, *13*). Further details are provided in (*14*), including data from larger cluster simulations indicating negligible finite size effects.

We first discuss the qualitative temperature dependence of optical conductivity (Fig. 1) for hole dopings $p = 0$, 0.1, and 0.2. Although we are concerned primarily with the metallic state of the doped system, it is important to establish the insulating nature of the undoped, half-filled model to verify strong correlation effects for our set of model parameters. The optical conductivity at half-filling, shown in Fig. 1A, demonstrates insulating behavior below roughly $T \sim t$, where cooling leads to a decrease of DC conductivity and formation of an optical gap. This behavior contrasts with the metallic properties of the doped case (Fig. 1, B and C), where a Drude-like peak at zero frequency is present and the conductivity increases with lowering temperature. In the metallic regime, the increase in conductivity is primarily associated with narrowing of the $\omega = 0$ peak. Below $T \sim t$, relatively little spectral weight is transferred to or from the Hubbard peak at $\omega \approx U = 6t$, which contains roughly the same spectral weight over a decade of temperature.

The metallic behavior at high temperatures is markedly distinct. For $T \gtrsim t$, the optical conductivity and its temperature evolution are similar for all dopings, including half-filling. Broad peaks are present at $\omega = 0$ and $\omega \approx U = 6$t. In this high-temperature regime, the heights of both peaks scale linearly with the inverse temperature $\beta = 1/T$. In contrast to the lower-temperature metallic regime, here the width of the $\omega = 0$ peak does not evolve with temperature, and the overall profile of the optical conductivity remains fixed.

Having explored the qualitative doping and temperature trends of the optical conductivity, we now focus on the Hubbard model's DC transport properties. The resistivity in natural units of h/e^2, where h is the reduced Planck constant and e is the elementary charge, is plotted versus temperature in Fig. 2. The Mott-Ioffe-Regel (MIR) limit tends to be of order unity in natural units. Evidently in our data, no saturation related to the MIR criterion is present. In particular, the resistivity for lightly doped systems greatly exceeds the MIR limit even at our lowest accessible temperature.

A clear distinction is present between temperatures below and above $T \sim 1.5t$. As discussed previously, in the half-filled model, this temperature scale marks an onset of insulating behavior. In Fig. 1, we additionally saw that in the doped, metallic cases, $T \sim 1.5t$ separates two regimes of qualitatively different temperature dependences in the optical conductivity. Here in Fig. 2, we see that the high- and low-temperature regimes differ also in the temperature and doping dependence of DC resistivity. The T-linear resistivity of the infinite temperature limit extends down to $T \sim 2t$ with little doping dependence. This terminates in the crossover region $t \lesssim T \lesssim 2t$, where the contrasting behaviors of the half-filled insulator and the doped metal start to appear. Going to lower temperature, T-linearity re-emerges in the doped metal for $T \lesssim t$, but with strong doping dependence. From $p = 0.1$ to $p = 0.3$, the temperature coefficient of resistivity decreases by roughly a factor of

[1]Department of Physics, Stanford University, Stanford, CA 94305, USA. [2]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory and Stanford University, Menlo Park, CA 94025, USA. [3]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: edwinwhuang@gmail.com (E.W.H.); tpd@stanford.edu (T.P.D.)

Downloaded from http://science.sciencemag.org/ on November 29, 2019

3 for low temperatures, whereas it remains nearly constant for $T \gtrsim 2t$. For all considered dopings, the resistivity appears T-linear and uninfluenced by MIR, thus indicating that strange metallic transport is present through a substantial portion of the Hubbard model's phase diagram.

To assess the relevance of model calculations to material physics, it is instructive to consider the infinite temperature limit. For a generic nonintegrable model with a bounded energy spectrum, it is expected that $T\sigma(\omega)$ converges to a limit for temperatures above the largest energy scales of the model (*15*), namely the ultrahigh-temperature limit. An immediate consequence is that a large, linear-T resistivity violating the MIR limit is expected for sufficiently high temperature. Although such behavior nominally reflects bad metallic transport, it is less relevant to experimental realizations of bad metals: Generally, both bad metals and saturating metals show their typical behavior at temperatures that are considerably smaller than the Fermi temperature or interaction energy scales. In our calculations of the Hubbard model, we have seen that properties expected in the ultrahigh-temperature limit extend down to $T \sim 2t$ before crossing over to a low-temperature regime with distinct properties. The fact that the Hubbard model already violates MIR and displays T-linear resistivity in this low-temperature regime suggests that its bad metallic transport is of a similar nature to that in strongly correlated materials.

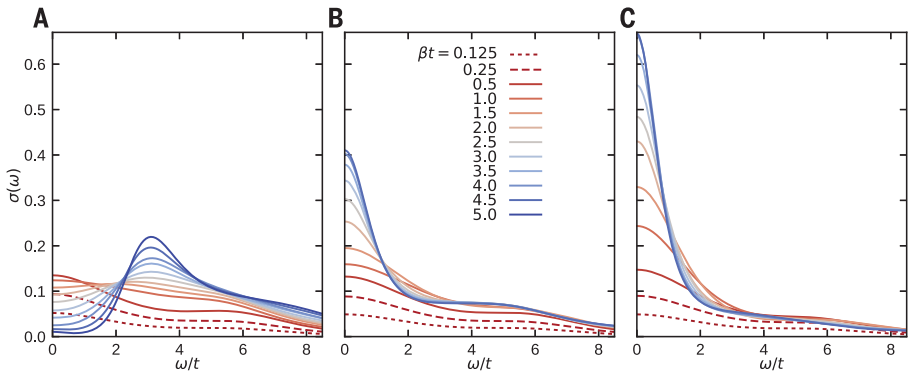


Fig. 1. Optical conductivity of the Hubbard model. Optical conductivity obtained through DQMC and MaxEnt analytic continuation for the Hubbard model with parameters U/t = 6, t'/t = −0.25. Shown are the data for various temperatures T = 1/β, where we have set Boltzmann's constant k_B = 1. Hole doping level is (**A**) p = 0.0, (**B**) 0.1, and (**C**) 0.2. Simulation cluster size is 8 × 8; see (*14*) for comparison against simulations on larger clusters.

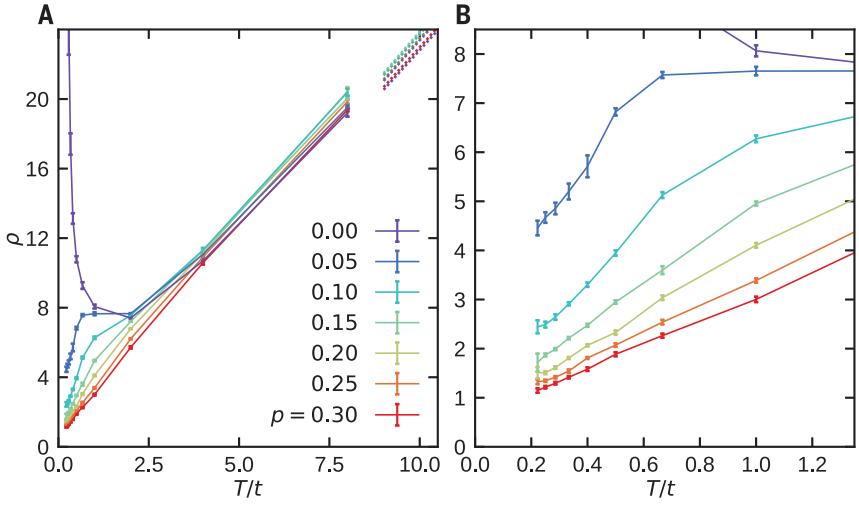


Fig. 2. DC resistivity extracted by analytic continuation. (**A**) DC resistivity as a function of temperature and hole doping, obtained from analytically continued optical conductivity as shown in Fig. 1. Solid lines through DQMC data points are guides to the eye. Dotted lines are results from moments expansions up to 18th order in the high-temperature limit (*14*). (**B**) Close-up view of the lowest-temperature data of (A). Error bars represent random sampling errors, determined by bootstrap resampling (*14*).

Besides analyzing analytically continued optical conductivity, DC transport properties may be estimated through imaginary time proxies: simple functions of the imaginary time current correlator that converge to the true DC resistivity in the low-temperature limit. Intuitively, one expects low-frequency properties to be most strongly related to data at large imaginary times. Specifically, $\tau = \beta/2$ is the "largest" imaginary time (because $\Lambda(\beta-\tau) = \Lambda(\tau)$). We first consider the proxy $\rho_1 = \pi T^2 \Lambda(\beta/2)^{-1}$, where $\Lambda(\beta/2) = \int d\omega f(\omega)\sigma(\omega)$. $f(\omega) = \frac{\omega}{2\pi}/\sinh(\beta\omega/2)$ is a bell-shaped function with width approximately $8T$ that becomes a delta function for $T \to 0$ (*16*). ρ_1 thus approaches the true DC resistivity if the optical conductivity is featureless over the width of $f(\omega)$. In Fig. 1, we have seen that the zero frequency peak can be sharper than $8T$, especially with increased doping. Owing to this, ρ_1, plotted in Fig. 3A, deviates from the analytically continued data of Fig. 2.

The shortcomings of ρ_1 can be compensated by incorporating information of the curvature of the current correlator at $\tau = \beta/2$ (*17*). In particular, $\rho_2 = \Lambda''(\beta/2)/(2\pi\Lambda(\beta/2)^2)$ provides a more robust estimate of resistivity when the Drude-like peak is more narrow than $8T$. As an example, if the optical conductivity consists of a Lorentzian peak at $\omega = 0$ with width Γ, the ratio of the proxy to the DC resistivity ranges from $\rho_2/\rho_{DC} = 1$ for $\Gamma \gg T$ to $\rho_2/\rho_{DC} = 1/2$ for $\Gamma \ll T$. Plotting ρ_2 for our DQMC data in Fig. 3B, we see that ρ_2 captures many of the same features present in Fig. 2. Although there may be differences in the precise value, in part owing to limitations of this simple proxy, the trends and the decrease of the temperature coefficient with doping compare well with analytically continued results and corroborate the presence of strange metallicity in the Hubbard model.

To further analyze transport properties of the Hubbard model, we consider the Nernst-Einstein relation, which connects conductivity to charge compressibility and diffusivity: $\sigma = \chi\mathcal{D}$. In the context of correlated materials, because compressibility is nearly constant at experimentally relevant temperatures, the T-linearity of resistivity derives from the diffusivity, which has been argued to be a more fundamental transport property (*18*, *19*). In Fig. 4A, we plot the inverse compressibility, obtained in DQMC without analytic continuation. Qualitatively similar trends in doping dependence are present in the resistivity and inverse compressibility, which are somewhat cancelled out when combined to calculate the diffusivity (Fig. 4B). Because both resistivity and inverse compressibility scale linearly in temperature, the inverse diffusivity approaches a constant at high

Downloaded from http://science.sciencemag.org/ on November 29, 2019

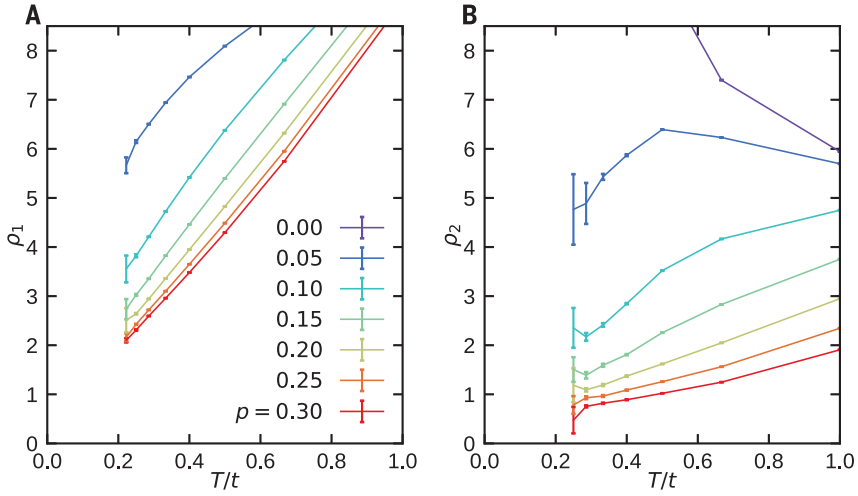


Fig. 3. DC resistivity estimated by imaginary time proxies. Proxies of DC resistivity (**A**) $\rho_1 = \pi T^2/\Lambda(\beta/2)$ and (**B**) $\rho_2 = \Lambda''(\beta/2)/(2\pi\Lambda(\beta/2)^2)$. Error bars are ±1 SEM, determined by bootstrap resampling.

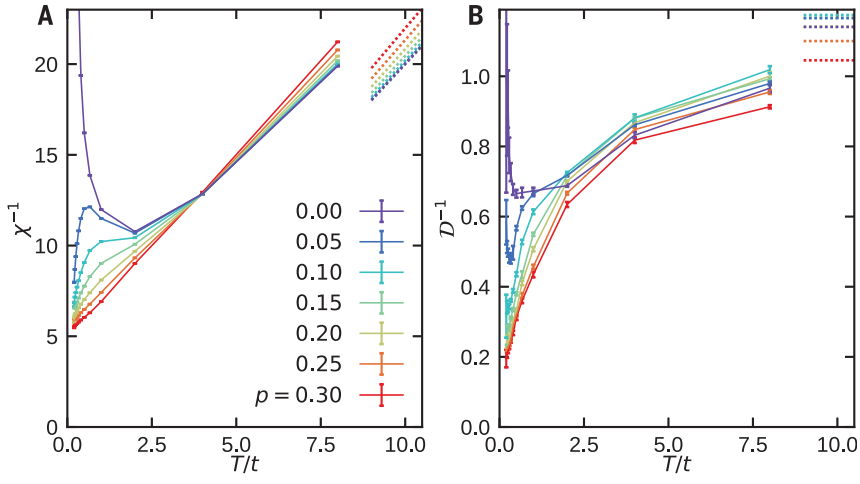


Fig. 4. Compressibility and diffusivity. (**A**) Inverse charge compressibility $\chi^{-1} = \left(\frac{\partial\langle n\rangle}{\partial\mu}\right)^{-1}$ calculated by DQMC simulations (without analytic continuation). Solid lines are guides to the eye; dotted lines are the high-temperature limit $\chi = \frac{1-p^2}{2T}$. (**B**) Inverse diffusivity obtained by applying the Nernst-Einstein relation $\sigma = \chi\mathcal{D}$ to the data of (A) and Fig. 2.

temperatures. Conversely, at low temperatures, the compressibility approaches a limiting constant value. We thus see in Fig. 4 that the temperature dependence of resistivity crosses over from being dominated by compressibility (*20*) to being controlled by diffusivity when lowering temperature. Similar crossover behavior has been observed in a recent study of an extended Hubbard model in the $t/U \to 0$ limit (*21*).

The presence of strange metallicity in the Hubbard model at temperatures small compared to the energy scales of model parameters provides promising evidence that the fundamental physics of correlated materials may be approached through studying simplified model Hamiltonians. In this regard, we view thorough numerical results as presented here to be an important benchmark for testing theoretical descriptions of strange metals and approximate approaches to the Hubbard model (*22*–*25*). A recent development involves measurement of transport properties in the Hubbard model by cold atoms experiments (*26*–*28*), with broadly similar findings to our results. Both in this field and in finite-temperature numerical approaches, studying the normal state down to temperatures proximate to ordering temperatures for superconductivity and other emergent phases remains a major challenge.

Although ground state calculations of the Hubbard model have revealed intertwined orders with analogies to experimental phase diagrams (*29*–*31*), important questions remain concerning their emergence from the normal state. Controlled approaches to the Hubbard model at finite temperature, such as our DQMC calculations where there is a sign problem, currently are unable to directly access these phases. Whether superconductivity in the Hubbard model follows directly from the strange metal as temperatures are lowered, or if coherent quasiparticles may emerge in between the strange metal and the ground state, remain intriguing open questions. Answers may be found through extending our measurements of dynamical quantities, including resistivity, by developing new numerical techniques or by improved quantum simulations.

REFERENCES AND NOTES

1. E. Fradkin, S. A. Kivelson, J. M. Tranquada, *Rev. Mod. Phys.* **87**, 457–482 (2015).
2. B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, J. Zaanen, *Nature* **518**, 179–186 (2015).
3. V. J. Emery, S. A. Kivelson, *Phys. Rev. Lett.* **74**, 3253–3256 (1995).
4. O. Gunnarsson, M. Calandra, J. E. Han, *Rev. Mod. Phys.* **75**, 1085–1099 (2003).
5. N. E. Hussey, K. Takenaka, H. Takagi, *Philos. Mag.* **84**, 2847–2864 (2004).
6. R. Blankenbecler, D. J. Scalapino, R. L. Sugar, *Phys. Rev. D Part. Fields* **24**, 2278–2286 (1981).
7. S. R. White *et al.*, *Phys. Rev. B Condens. Matter* **40**, 506–516 (1989).
8. I. Khait, S. Gazit, N. Y. Yao, A. Auerbach, *Phys. Rev. B* **93**, 224205 (2016).
9. O. A. Starykh, A. W. Sandvik, R. R. P. Singh, *Phys. Rev. B Condens. Matter* **55**, 14953–14967 (1997).
10. N. H. Lindner, A. Auerbach, *Phys. Rev. B Condens. Matter Mater. Phys.* **81**, 054512 (2010).
11. E. Perepelitsky *et al.*, *Phys. Rev. B* **94**, 235115 (2016).
12. M. Jarrell, J. E. Gubernatis, *Phys. Rep.* **269**, 133–195 (1996).
13. O. Gunnarsson, M. W. Haverkort, G. Sangiovanni, *Phys. Rev. B Condens. Matter Mater. Phys.* **82**, 165125 (2010).
14. See supplementary materials.
15. S. Mukerjee, V. Oganesyan, D. Huse, *Phys. Rev. B Condens. Matter Mater. Phys.* **73**, 035113 (2006).
16. N. Trivedi, R. T. Scalettar, M. Randeria, *Phys. Rev. B Condens. Matter* **54**, R3756–R3759 (1996).
17. S. Lederer, Y. Schattner, E. Berg, S. A. Kivelson, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4905–4910 (2017).
18. S. A. Hartnoll, *Nat. Phys.* **11**, 54–61 (2015).
19. T. Hartman, S. A. Hartnoll, R. Mahajan, *Phys. Rev. Lett.* **119**, 141601 (2017).
20. J. Kokalj, *Phys. Rev. B* **95**, 041110 (2017).
21. C. H. Mousatov, I. Esterlis, S. A. Hartnoll, *Phys. Rev. Lett.* **122**, 186601 (2019).
22. T. Pruschke, M. Jarrell, J. Freericks, *Adv. Phys.* **44**, 187–210 (1995).
23. D. Bergeron, V. Hankevych, B. Kyung, A.-M. S. Tremblay, *Phys. Rev. B Condens. Matter Mater. Phys.* **84**, 085128 (2011).
24. X. Deng *et al.*, *Phys. Rev. Lett.* **110**, 086401 (2013).
25. W. Xu, K. Haule, G. Kotliar, *Phys. Rev. Lett.* **111**, 036401 (2013).
26. W. Xu, W. McGehee, W. Morong, B. DeMarco, Bad-metal relaxation dynamics in a Fermi lattice gas. arXiv:1606.06669 [cond-mat.quant-gas] (2016).
27. P. T. Brown *et al.*, *Science* **363**, 379–382 (2019).
28. M. A. Nichols *et al.*, *Science* **363**, 383–387 (2019).
29. H.-C. Jiang, T. P. Devereaux, *Science* **365**, 1424–1428 (2019).
30. B.-X. Zheng *et al.*, *Science* **358**, 1155–1160 (2017).

Downloaded from http://science.sciencemag.org/ on November 29, 2019

31. E. W. Huang, C. B. Mendl, H.-C. Jiang, B. Moritz, T. P. Devereaux, *npj Quant. Mat.* **3**, 22 (2018).
32. E. W. Huang, Source code for "Strange metallicity in the doped Hubbard model," Zenodo (2019). https://doi.org/10.5281/zenodo.2650371.
33. E. W. Huang, Data for "Strange metallicity in the doped Hubbard model," Zenodo (2019). https://doi.org/10.5281/zenodo.2650374.

ACKNOWLEDGMENTS

We acknowledge helpful discussions with E. Berg, L. Delacrétaz, B. Goutéraux, S. Hartnoll, S. Kivelson, Y. Schattner, and J. Zaanen. **Funding:** This work was supported by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Division of Materials Sciences and Engineering. Computational work was performed on the Sherlock cluster at Stanford University and on resources of the National Energy Research Scientific Computing Center, supported by the U.S. DOE under Contract no. DE-AC02-05CH11231. **Author contributions:** E.W.H. conceived the study, performed numerical simulations, and analyzed the data. R.S. performed initial simulations and data analysis. E.W.H., B.M., and T.P.D. wrote the manuscript. **Competing interests:** The authors declare no competing interest. **Data and materials availability:** Source code for the simulations, including the MaxEnt analytic continuation code, is available at (*32*). Data used in generating the main figures are available at (*33*). Raw output data of the DQMC simulations are stored on the Sherlock cluster at Stanford University and are available from the corresponding author upon request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/987/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Table S1
References (*34*–*37*)

8 July 2018; resubmitted 28 December 2018
Accepted 25 October 2019
10.1126/science.aau7063

Downloaded from http://science.sciencemag.org/ on November 29, 2019

ORGANIC CHEMISTRY

Highly enantioselective carbene insertion into N–H bonds of aliphatic amines

Mao-Lin Li, Jin-Han Yu, Yi-Hao Li, Shou-Fei Zhu*, Qi-Lin Zhou*

Aliphatic amines strongly coordinate, and therefore easily inhibit, the activity of transition-metal catalysts, posing a marked challenge to nitrogen-hydrogen (N–H) insertion reactions. Here, we report highly enantioselective carbene insertion into N–H bonds of aliphatic amines using two catalysts in tandem: an achiral copper complex and chiral amino-thiourea. Coordination by a homoscorpionate ligand protects the copper center that activates the carbene precursor. The chiral amino-thiourea catalyst then promotes enantioselective proton transfer to generate the stereocenter of the insertion product. This reaction couples a wide variety of diazo esters and amines to produce chiral α-alkyl α–amino acid derivatives.

Chiral amines are ubiquitous in natural products, pharmaceuticals, and agrochemicals. Approximately 43% of the top 200 prescription medicines in 2016 contain an aliphatic amine moiety (Fig. 1A) (*1*, *2*). The development of highly enantioselective transition-metal–catalyzed reactions that form C–N bonds is thus of long-standing interest in synthetic chemistry (*3*–*5*). Transition-metal–catalyzed carbenoid insertion into N–H bonds has proven a straightforward method in this respect, benefitting from mild reaction conditions, good functional group tolerance, and readily available reactants (*6*, *7*). Recently, chiral transition-metal catalysts have been successfully applied to enantioselective N–H insertion reactions in the synthesis of natural or unnatural chiral α–amino acid derivatives (*8*, *9*). However, these reactions have been restricted to aromatic amines (*10*–*15*) or amides (*16*–*18*) (Fig. 1B). Aliphatic amines are comparatively stronger Lewis bases and thus poison the metal catalysts by strong coordination, interfering with generation of the metal carbenoid (*19*, *20*). Moreover, excess aliphatic amines can displace the ylide from metal-ylide intermediates, leading to racemic product formation from the free ylide (Fig. 1C, upper). We envisioned that a combination of two catalysts (*21*, *22*) might address these challenges: An achiral transition-metal catalyst compatible with aliphatic amines would generate the ylide intermediate, and a separate chiral catalyst would then promote enantioselective proton transfer. After exploring various transition-metal catalysts and chiral H-bonding catalysts in the N–H insertion reaction of α-diazobutanoates with benzylamine (tables S1 to S5), we report here the success of this approach, pairing the homoscorpionate-coordinated copper complex Tp*Cu [Tp* is hydrotris(3,5-dimethylpyrazolyl) borate] (*23*–*25*) with a chiral amino-thiourea (CAT) bearing a pyrrolidine motif (*26*–*29*) (Fig. 1C, lower). The reaction provides efficient, highly enantioselective access to chiral α-alkyl α–amino acid derivatives bearing secondary and tertiary amino substituents, which are difficult to prepare by other methods.

Under the optimal reaction conditions, a broad range of aliphatic amines was then investigated for N–H insertion with 2-phenylpropan-2-yl α-diazobutyrate **2** (Fig. 2A). The benzylic primary amines underwent the N–H insertion reaction smoothly to afford the corresponding α-aminobutanoic acid derivatives (**3** to **9**) in high yields (81 to 95%) with high enantioselectivities [88 to 92% enantiomeric excess (ee)], though 2-phenethylamine and *n*-butylamine gave moderate enantioselectivities (**10** and **11**). Secondary amines were also suitable substrates for the reaction but required longer reaction times and excess diazo compounds for satisfactory outcomes. Piperidine derivatives generally exhibited high enantioselectivities, and the introduction of electron-withdrawing groups (CO_2Me and CN) at the 4-position led to higher yields (71 and 86%, respectively) and better enantioselectivities (90 and 94% ee, respectively) (**12** to **14**). Morpholine, substituted piperazines, and thiomorpholine also underwent the N–H insertion and gave the desired products (**15** to **19**) in high yield with 87 to 97% ee. Fused heterocyclic amines also afforded N–H insertion products in satisfactory yields and enantioselectivities (**20** and **21**). However, azepane and *N*-methyl-1-phenylmethanamine gave lower enantioselectivities (73 and 77% ee, respectively, **22** and **23**). The N–H insertion reactions with chiral drugs, such as amoxapine, trimetazidine, and vortioxetine, proceeded smoothly to afford corresponding α-aminobutanoic acid derivatives (**24** to **26**) in high yields with good enantioselectivities (Fig. 2B). The scope with respect to the diazo reactant in the N–H insertion of morpholine was investigated next (Fig. 2C). Diazo esters with linear or branched α-alkyl chains afforded the desired products in good to high yields (66 to 99%) with excellent enantioselectivities (94 to 96% ee) (**27** to **31**). Various functional groups (alkenyl, ester, ether, amide) appended to the alkyl chain were tolerated, giving high yields (86 to 99%) and enantioselectivities (87 to 96% ee) (**32** to **38**). Furthermore, α-aryl diazoacetates also afforded the corresponding arylglycine derivatives (**39** to **44**) in high yields (>95%) with good enantioselectivities (72 to 89% ee).

To demonstrate the further synthetic utility of the N–H insertion reaction, several transformations of the insertion products were performed. The product (*R*)-**3** was reduced by $LiAlH_4$ to afford (*R*)-2-benzylamino-butanol [(*R*)-**45**], an intermediate for the synthesis of γ-secretase inhibitors (*30*) and PDK1 inhibitors (*31*) (Fig. 3A). The product **15**, which could be prepared at gram scale from N–H insertion of morpholine and diazo ester **2**, was hydrolyzed to acid **46**, an intermediate for the synthesis of hyperproliferative disorder (HPD) treatment agents (Fig. 3B) (*32*). The (*R*)-2-morpholinopropanoic acid (**47**), which is prepared by hydrolysis of the product **27**, is a key intermediate for the synthesis of the phosphatidylinositol 3-kinase δ (PI3Kδ) inhibitors (*33*), as well as DNA-dependent protein kinase (DNA-PK) inhibitors (*34*) (Fig. 3C).

To gain deeper insight into the mechanism of the N–H insertion reaction, we performed kinetic analyses by using in situ infrared (IR) spectroscopy. To accelerate the kinetics, the initial rates of the reaction were measured at various concentrations of the components at 40°C (figs. S1 to S5). The rate showed a first-order dependence on concentrations of Tp*Cu and diazo compound **2** (Fig. 4A), which indicates that the formation of metal carbenoid through Tp*Cu-catalyzed decomposition of diazo ester **2** is the likely rate-limiting step. The negative first-order dependence on CAT is consistent with a pre-equilibrium formation of a resting-state complex between the thiourea catalyst CAT and Tp*Cu, which would suppress the copper-catalyzed decomposition of the diazo ester. However, benzylamine, generally coordinating with the metal catalyst and suppressing the formation of metal carbenoid, showed a zero-order kinetic effect in the reaction, which suggests that the coordination of CAT to Tp*Cu is much stronger than that of benzylamine, and the inhibition by benzylamine can be ignored (fig. S6). We posit that the negative Tp* ligand renders the copper catalyst a softer Lewis acid that favors interaction with a soft base like sulfur. Further evidence for the stronger interaction between CAT and Tp*Cu includes observations

State Key Laboratory and Institute of Elemento-Organic Chemistry, College of Chemistry, Nankai University, Tianjin 300071, China.
*Corresponding author. Email: sfzhu@nankai.edu.cn (S.-F.Z.); qlzhou@nankai.edu.cn (Q.-L.Z.)

Downloaded from http://science.sciencemag.org/ on November 29, 2019

of the changes in nuclear magnetic resonance (NMR) and ultraviolet (UV)–visible spectra after adding CAT into a mixture of Tp*Cu and benzylamine (figs. S7 to S11). Taken together, the kinetic, NMR, and UV studies are consistent with the Tp*Cu•CAT complex, rather than the Tp*Cu•$BnNH_2$ complex, as the resting state of the catalyst in the reaction. Although the Tp*Cu•CAT complex is the main resting state of the copper, free Tp*Cu is still evident under the reaction conditions (fig. S10) and can react with the diazo compound.

Density functional theory (DFT) calculations indicated that the copper catalyst in the intermediate Cu-ylide could be recoordinated by thiourea to release a free ylide or its more stable tautomer free enol (Fig. 4B). Even without the added chiral catalyst, the proton transfer of these intermediates is still a very rapid process and can be mediated by trace water, enol intermediates, and even substrates themselves (*18*, *35*). The similar pK_a values (where K_a is the acid dissociation constant) (figs. S12 to S14 and table S8) of the Brønsted acidic site and protonated basic site of the thiourea promote concerted proton transfer, whereby the thiourea protonates the newly formed stereogenic center, and the amino group deprotonates the free enol. Computational modeling of the Tp*Cu•CAT complex using DFT revealed a minimum energy structure in which bonding of the Tp*Cu to the sulfur atom of the thiourea enhances Brønsted acidity (Fig. 4B).

The enantio-determining proton-transfer step was studied computationally by using DFT with the Tp*Cu•CAT complex as a catalyst. The structures corresponding to the lowest-energy transition states for the major and minor enantiomers of product are presented in Fig. 4C. The optimal transition state TSRaCu-I is only 2.8 kcal/mol higher in free energy than the free enol, implying extremely efficient proton-transfer catalysis. In accord with experimental observations, the calculated energy of transition state TSRaCu-I was 4.9 kcal/mol lower than that of the TSSaCu-I transition state leading to the unfavored (*S*)-product of N–H insertion. Besides the different hydrogen-bonding interactions in the transition states TSRaCu-I and TSSaCu-I, the S–Cu bond in TSRaCu-I is markedly shorter than that in TSSaCu-I. The shorter S–Cu bond indicates stronger coordination of copper by thiourea and likely higher Brønsted acidity of the thiourea catalyst, which would promote proton transfer to the substrate.

Several other tris(pyrazolyl)borate (Tp) ligands bearing different substituents on the pyrazol rings were also evaluated under the standard reaction conditions (Fig. 4D). Despite a large fluctuation in the yield, the in situ IR studies showed that all tested Tp ligands promoted high conversions and that the major by-product was 2-phenylpropan-2-yl but-2-enoate, resulting from the β-H migration of the metal carbenoid (figs. S15 to S19). Modifying the Tp ligands also influenced the enantioselectivity when the same chiral thiourea catalyst was used, indicating involvement of the copper catalyst in the enantio-determining step. By contrast, upon tuning the electronic properties of the arene ring of the chiral thiourea catalyst, the enantioselectivity decreased precipitously, whereas the yield remained almost unchanged (table S7). We again hypothesize that copper coordination enhances the Brønsted acidity of the thiourea catalyst while minimally influencing the distant site of enantioinduction (Fig. 4C).

On the basis of the aforementioned mechanistic studies, a catalytic cycle is proposed in Fig. 4E. The Tp*Cu•CAT complex serves as a resting state of the catalyst and dissociates to release Tp*Cu, which catalyzes transformation of the diazo ester into the metal carbenoid in the rate-determining step. Nucleophilic attack on the metal carbenoid by the aliphatic amine

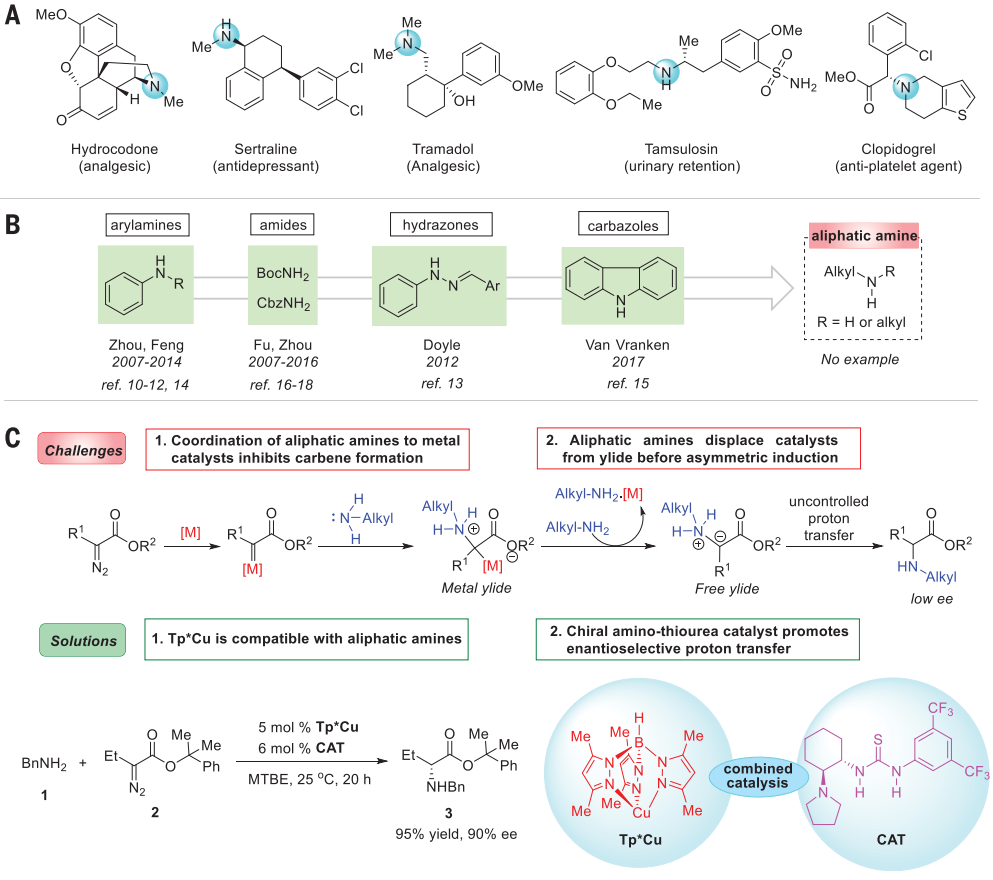


Fig. 1. Strategy for enantiocontrol of N–H insertion reactions of aliphatic amines with carbenes. (**A**) Representative drugs demonstrating the ubiquity of chiral aliphatic amines in bioactive molecules. (**B**) Amine sources reported for enantioselective N–H insertion reactions. (**C**) Enantioselective transition-metal–catalyzed N–H insertion reactions with aliphatic amines: challenges and solutions. Optimal reaction conditions: The reaction of **1** (0.2 mmol), **2** (0.22 mmol), Tp*Cu (5 mole %), and CAT (6 mol %) was carried out in 3 ml of methyl *tert*-butyl ether (MTBE) at 25°C for 20 hours. $BnNH_2$, benzylamine; $BocNH_2$, *tert*-butyl carbamate; $CbzNH_2$, benzyl carbamate; Me, methyl; Et, ethyl; Ph, phenyl; M, metal; ref., reference.

Downloaded from http://science.sciencemag.org/ on November 29, 2019

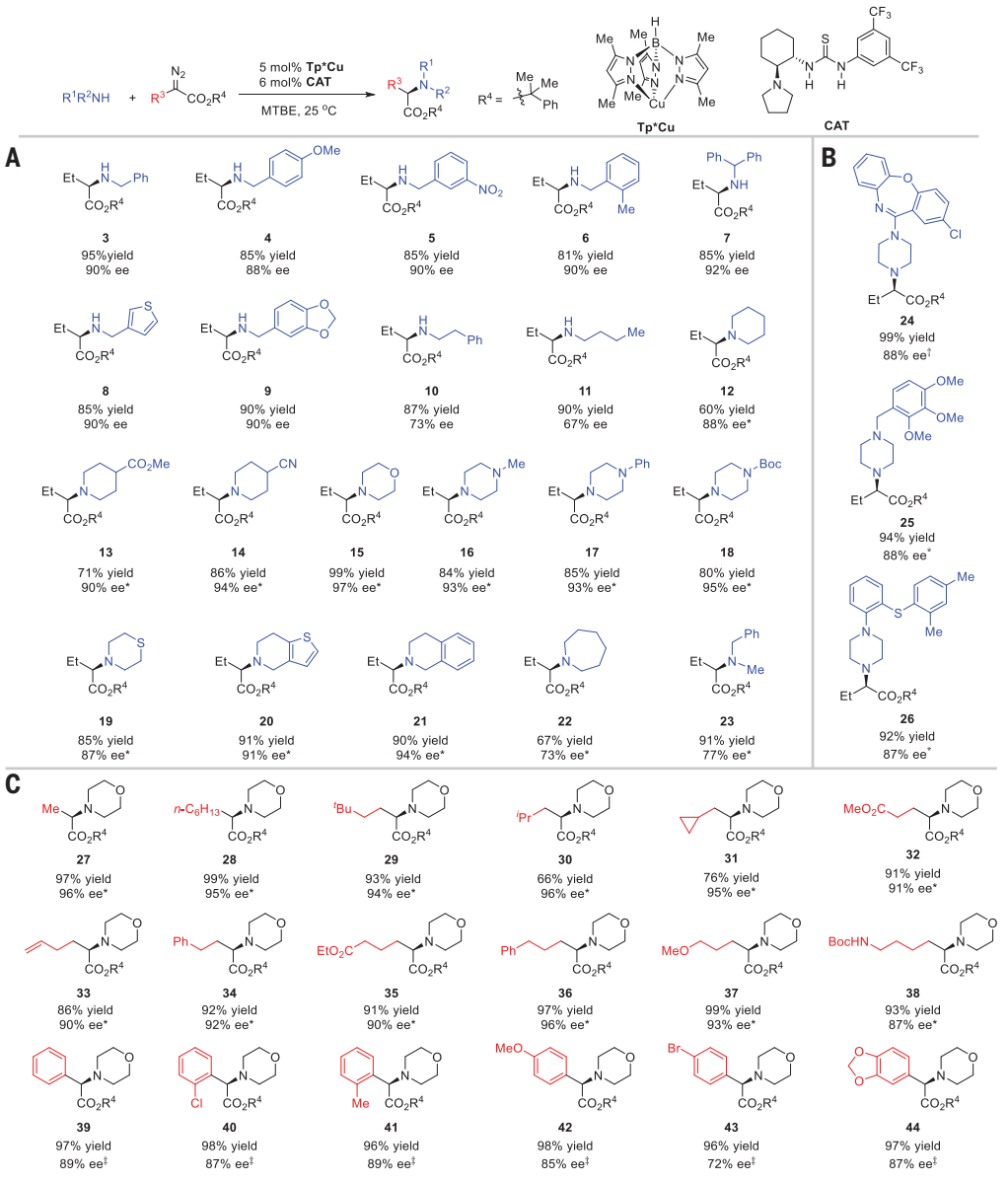


Fig. 2. Scope of aliphatic amines and α-diazo esters in the enantioselective N–H insertion reaction. Reaction conditions: amines (0.2 mmol), α-diazo esters (0.22 mmol), Tp*Cu (5 mol %), CAT (6 mol %), 3 ml MTBE, 25°C, 20 hours. Isolated yields are given. The ee values were determined by high-performance liquid chromatography. (**A**) Scope of aliphatic amines. (**B**) Application to enantioselective late-stage functionalization of pharmaceuticals. (**C**) Scope of α-diazo esters. *Diazo esters (0.3 mmol), 36 hours. †Diazo esters (0.3 mmol), MTBE:CH_2Cl_2 = 10:1, 36 hours. ‡Diazo esters (0.3 mmol), 40°C, 20 hours. ^{t}Bu, *tert*-butyl; ^{i}Pr, *iso*-propyl.

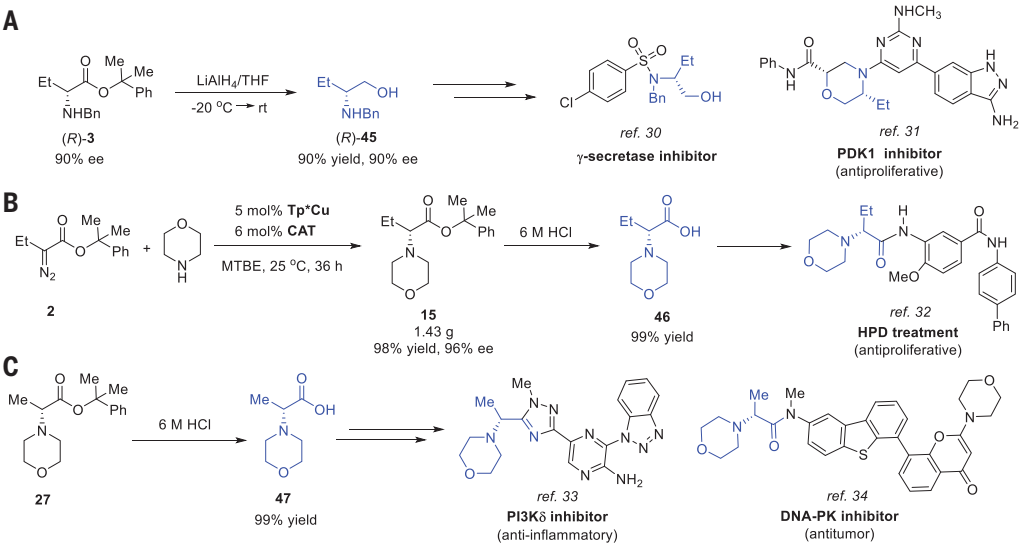
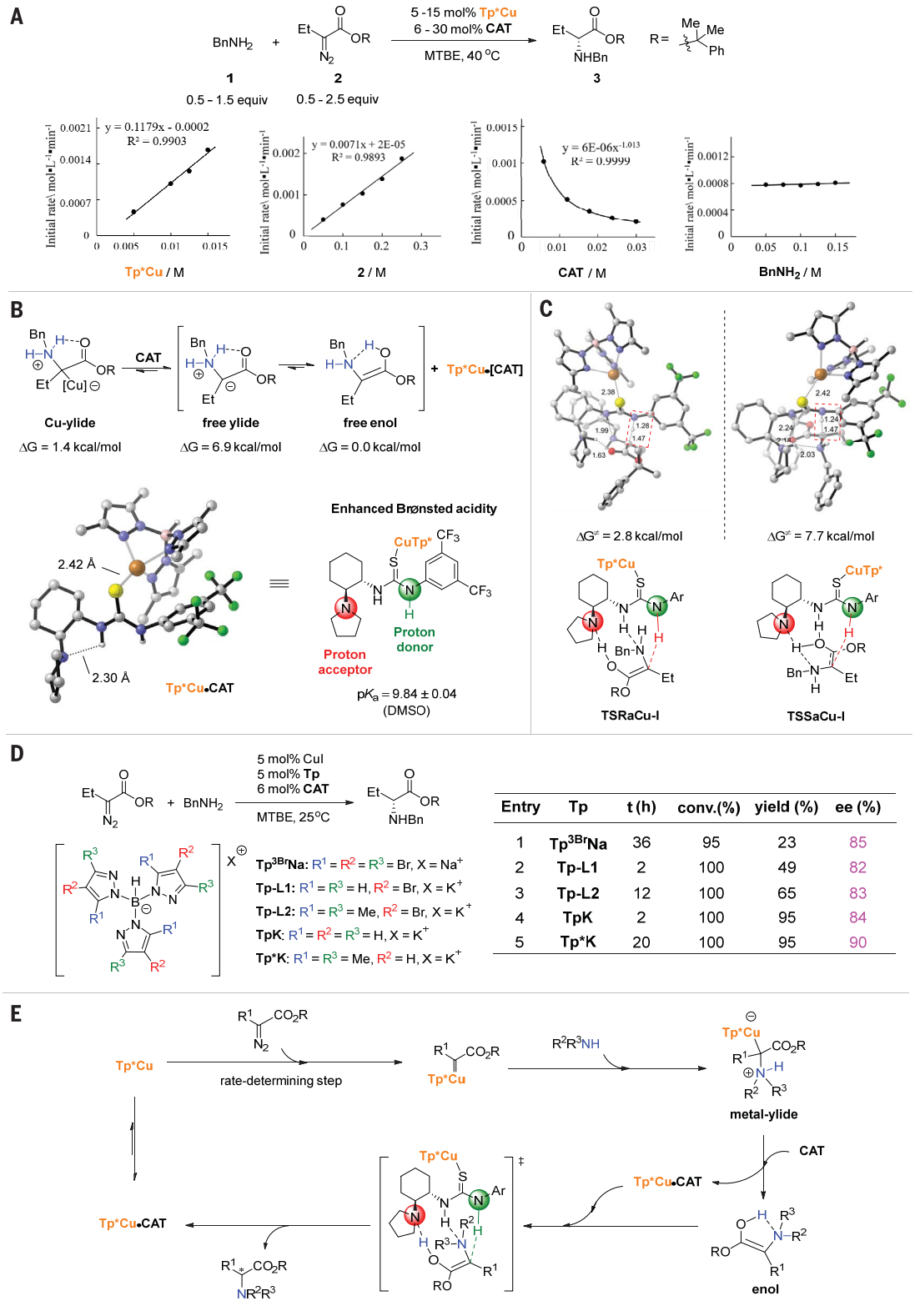


Fig. 3. Synthetic transformations of the N–H insertion products. (**A**) Transformation of **3** to (*R*)-2-benzylamino-butanol [(*R*)-**45**], a key intermediate for the synthesis of bioactive molecules. THF, tetrahydrofuran; rt, room temperature. (**B**) Formal synthesis of HPD treatment agents with the N–H insertion as key step. (**C**) Transformation of **27** to **47**, a key intermediate for the synthesis of bioactive molecules.

Downloaded from http://science.sciencemag.org/ on November 29, 2019



Entry	Tp	t (h)	conv.(%)	yield (%)	ee (%)
1	$Tp^{3Br}Na$	36	95	23	85
2	Tp-L1	2	100	49	82
3	Tp-L2	12	100	65	83
4	TpK	2	100	95	84
5	Tp*K	20	100	95	90

Fig. 4. Mechanistic studies. (**A**) Kinetic profiles of Cu-catalyzed N–H insertion reaction of **2** and $BnNH_2$. (**B**) Calculated Gibbs free energy (ΔG) of Cu-ylide, free ylide, and free enol. Lowest-energy ground-state structure of the Tp*Cu•CAT complex. Structures of alternative higher-energy complexes are provided in fig. S20. (**C**) DFT-optimized lowest-energy transition structures for *R* and *S* products. Calculations were performed at the m062x-D3/def2tzvpp//m062x-D3/def2svp level. Structures of alternative higher-energy complexes are provided in figs. S21 to S25. (**D**) Influence of different Tp ligands. (**E**) Proposed catalytic cycle for the enantioselective carbene insertion into N–H bonds of aliphatic amines. conv., conversion; DMSO, dimethyl sulfoxide; equiv, equivalents.

generates a metal ylide. The catalyst CAT displaces the ylide from the metal-ylide intermediate to generate free enol and the Tp*Cu•CAT complex. The Tp*Cu•CAT complex then promotes proton transfer in the free enol through a push-pull mechanism: The amino moiety accepts a proton from the hydroxy group of the enol while the thiourea moiety donates a proton to the β-carbon of the enol.

The success of the overall transformation relies on the combined properties of the achiral copper catalyst and chiral organocatalyst. This study not only solves a long-standing challenge in enantioselective carbene insertion reactions

Downloaded from http://science.sciencemag.org/ on November 29, 2019

but also provides a potentially general strategy for transition-metal–catalyzed asymmetric transformations involving strongly coordinating substrates.

REFERENCES AND NOTES

1. N. A. McGrath, M. Brichacek, J. T. Njardarson, *J. Chem. Educ.* **87**, 1348–1349 (2010).
2. These pharmaceutical posters are freely accessible to anyone as PDF files at https://njardarson.lab.arizona.edu.
3. T. C. Nugent, *Chiral Amine Synthesis: Methods, Developments and Applications* (Wiley VCH, 2010).
4. Y. Yang, S.-L. Shi, D. Niu, P. Liu, S. L. Buchwald, *Science* **349**, 62–66 (2015).
5. S.-L. Shi, Z. L. Wong, S. L. Buchwald, *Nature* **532**, 353–356 (2016).
6. M. P. Doyle, M. A. McKervey, T. Ye, *Modern Catalytic Methods for Organic Synthesis with Diazo Compounds* (Wiley, 1998).
7. A. Ford *et al*., *Chem. Rev.* **115**, 9981–10080 (2015).
8. S.-F. Zhu, Q.-L. Zhou, *Acc. Chem. Res.* **45**, 1365–1377 (2012).
9. D. Gillingham, N. Fei, *Chem. Soc. Rev.* **42**, 4918–4931 (2013).
10. B. Liu, S.-F. Zhu, W. Zhang, C. Chen, Q.-L. Zhou, *J. Am. Chem. Soc.* **129**, 5834–5835 (2007).
11. Z. Hou *et al*., *Angew. Chem. Int. Ed.* **49**, 4763–4766 (2010).
12. S.-F. Zhu, B. Xu, G.-P. Wang, Q.-L. Zhou, *J. Am. Chem. Soc.* **134**, 436–442 (2012).
13. X. Xu, P. Y. Zavalij, M. P. Doyle, *Angew. Chem. Int. Ed.* **51**, 9829–9833 (2012).
14. Y. Zhu *et al*., *Angew. Chem. Int. Ed.* **53**, 1636–1640 (2014).
15. V. Arredondo, S. C. Hiew, E. S. Gutman, I. D. U. A. Premachandra, D. L. Van Vranken, *Angew. Chem. Int. Ed.* **56**, 4156–4159 (2017).
16. E. C. Lee, G. C. Fu, *J. Am. Chem. Soc.* **129**, 12066–12067 (2007).
17. B. Xu, S.-F. Zhu, X.-L. Xie, J.-J. Shen, Q.-L. Zhou, *Angew. Chem. Int. Ed.* **50**, 11483–11486 (2011).
18. Y.-Y. Ren, S.-F. Zhu, Q.-L. Zhou, *Org. Biomol. Chem.* **16**, 3087–3094 (2018).
19. S. R. Hansen, J. E. Spangler, J. H. Hansen, H. M. L. Davies, *Org. Lett.* **14**, 4626–4629 (2012).
20. I. D. Jurberg, H. M. L. Davies, *Chem. Sci.* **9**, 5112–5118 (2018).
21. A. E. Allen, D. W. C. Macmillan, *Chem. Sci.* **2012**, 633–658 (2012).
22. D.-F. Chen, Z.-Y. Han, X.-L. Zhou, L.-Z. Gong, *Acc. Chem. Res.* **47**, 2365–2377 (2014).
23. S. E. Bromberg *et al*., *Science* **278**, 260–263 (1997).
24. S. Trofimenko, *Scorpionates: Polypyrazolylborate Ligands and Their Coordination Chemistry* (Imperial College Press, 1999).
25. M. E. Morilla *et al*., *Chem. Commun.* **2002**, 2998–2999 (2002).
26. M. S. Sigman, E. N. Jacobsen, *J. Am. Chem. Soc.* **120**, 4901–4902 (1998).
27. T. Okino, Y. Hoashi, Y. Takemoto, *J. Am. Chem. Soc.* **125**, 12672–12673 (2003).
28. S.-Z. Nie *et al*., *Tetrahedron Asymmetry* **21**, 2055–2059 (2010).
29. A. G. Doyle, E. N. Jacobsen, *Chem. Rev.* **107**, 5713–5743 (2007).
30. M. Neitzel, J. Marugg, WO patent 2005090296 (2005).
31. J. R. Medina *et al*., *J. Med. Chem.* **54**, 1871–1895 (2011).
32. K. Thede, W. J. Scott, E. Bender, S. Golz, A. Haegebarth, P. Lienau, F. Puehler, D. Basting, D. Schneider, M. Moewes, WO patent 2014147182 (2014).
33. I. Terstiege *et al*., *Bioorg. Med. Chem. Lett.* **27**, 679–687 (2017).
34. M. Frigerio, M. G. Hummersone, K. A. Menear, M. R. V. Finlay, E. J. Griffen, L. L. Ruston, J. J. Morris, A. K. T. Ting, B. T. Golding, R. J. Griffin, I. R. Hardcastle, S. Rodriguez-Aristegui, WO patent 2010136778 (2010).
35. Y. Liang, H. Zhou, Z.-X. Yu, *J. Am. Chem. Soc.* **131**, 17783–17785 (2009).

ACKNOWLEDGMENTS

We thank X.-S. Xue for discussions of the DFT calculations and M. P. Doyle for comments and suggestions on the preparation of the manuscript. **Funding:** We thank the National Natural Science Foundation of China (21625204, 21790332, 21532003), the "111" project (B06005) of the Ministry of Education of China, and the National Program for Special Support of Eminent Professionals for financial support. **Author contributions:** Q.-L.Z. and S.-F.Z. conceived the study; M.-L.L., S.-F.Z., and Q.-L.Z. designed the experiments and analyzed the data; M.-L.L. performed the reactions and the mechanistic and DFT studies; J.-H.Y. and Y.-H.L. made some of the diazo substrates; and M.-L.L., S.-F.Z. and Q.-L.Z. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Additional optimization and mechanistic data are provided in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/990/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S25
Tables S1 to S9
Spectral Data
Calculation Data
References (*36*–*69*)

20 February 2019; accepted 17 September 2019
10.1126/science.aaw9939

Downloaded from http://science.sciencemag.org/ on November 29, 2019

INFECTION

Host resistance factor SLC11A1 restricts *Salmonella* growth through magnesium deprivation

Olivier Cunrath* and Dirk Bumann†

The pleiotropic host resistance factor SLC11A1 (NRAMP1) defends against diverse intracellular pathogens in mammals by yet-unknown mechanisms. We compared *Salmonella* infection of coisogenic mice with different *SLC11A1* alleles. SLC11A1 reduced *Salmonella* replication and triggered up-regulation of uptake systems for divalent metal cations but no other stress responses. SLC11A1 modestly diminished iron availability and acutely restricted *Salmonella* access to magnesium. Growth of *Salmonella* cells in the presence of SLC11A1 was highly heterogeneous and inversely correlated with expression of the crucial magnesium transporter gene *mgtB*. We observed superimposable single-cell patterns in mice lacking SLC11A1 when we restricted *Salmonella* access to magnesium by impairing its uptake. Together, these findings identify deprivation of the main group metal magnesium as the main resistance mechanism of SLC11A1 against *Salmonella*.

Solute carrier family 11, member 1 [SLC11A1; also called natural resistance-associated macrophage protein 1 (NRAMP1)] is a host resistance factor that controls susceptibility to the intracellular pathogens *Salmonella*, *Mycobacteria*, and *Leishmania* (*1*). SLC11A1 transports Fe^{2+}, Mn^{2+}, and Co^{2+} out of phagosomes and may deprive vacuolar pathogens of these essential micronutrients. In addition, SLC11A1 modulates phagosome maturation, proinflammatory cytokines and activation of innate lymphocytes, and generation of nitric oxide, reactive oxygen species, and lipocalin 2, as well as mammalian iron homeostasis (fig. S1). The main mechanisms of SLC11A1-mediated resistance are still unclear, and metal starvation in the phagosome might be secondary to the effects of inflammatory responses and altered host iron homeostasis (*1*).

We generated coisogenic mice carrying the D169 (aspartic acid at residue 169) null allele (*SLC11A1*^{s}, s) or the functional G169 (Gly^{169}) allele (*SLC11A1*^{r}, r) in the C57BL/6 background. After infection with *Salmonella*, heterozygous (r/s) and particularly homozygous (r/r) mice showed delayed symptoms and reduced *Salmonella* colonization of spleen compared to (s/s) mice (Fig. 1A; fig. S2, a and b). (r/s) and (r/r) mice still succumbed to the disease from day 7 after infection, in contrast to 129S2/SvPasCrl mice, which survived *Salmonella* infection, indicating contributions of additional resistance loci (*2*). *SLC11A1*^{r} reduced *Salmonella* replication as measured with the growth reporter TIMER^{bac} (*3*), from 0.14 ± 0.02 hour^{-1} (mean ± SD) in (s/s) mice to 0.08 ± 0.02 hour^{-1} in (r/s) mice and 0.065 ± 0.01 hour^{-1} in (r/r) mice, thus approximating division rates in highly resistant 129S2/SvPasCrl mice (Fig. 1, B and C; fig. S2, c and d). Slow replication was maintained through day 6 after infection and fully explained the reduced *Salmonella* loads in (r/s) and (r/r) mice, but not in 129S2/SvPasCrl mice (Fig. 1, A, C, and D). SLC11A1 thus regulated *Salmonella* replication (*4*), and additional loci in 129S2/SvPasCrl mice enhanced killing of *Salmonella*.

We sorted *Salmonella* from spleens and determined the protein profiles. We analyzed (s/s) mice at day 4, because they succumbed to infection starting on day 5. For (r/s) and (r/r) mice, we had to wait until day 6, or use higher inocula, to obtain sufficient *Salmonella* material. SLC11A1 had a similar impact on *Salmonella* growth at days 4 and 6 after infection (Fig. 1C) and after low- and high-dose infection (fig. S1e). Comparison of 1833 *Salmonella* proteins revealed increased abundance of uptake systems for metal cations Mg^{2+}, Zn^{2+}, $Fe^{2+/3+}$, and Mn^{2+} in (r/s) and particularly (r/r) mice (Fig. 2A, colored circles). Ribosome subunits were less abundant in (r/s) and (r/r) mice (Fig. 2A, black circles), consistent with slower *Salmonella* replication (Fig. 1C). Otherwise, *Salmonella* proteomes from the different mouse genotypes were rather similar (fig. S3a). We did not observe signs for differential oxidative or nitrosative stress, in agreement with the activities of respective *Salmonella* reporter strains (Fig. 2B; fig. 3b) (*5*) and genetic evidence against a role of these stresses in SLC11A1-mediated resistance (*6*). Major virulence systems and metabolic and cell envelope stress regulons were also unaltered (fig. S3a). Finally, comparison to proteomes of slow- and fast-growing *Salmonella* subsets from BALB/c (s/s) mice (*3*) (fig. S3c) indicated fundamentally different mechanisms of *Salmonella* growth control in the presence or absence of SLC11A1. Taken together, these results indicate that SLC11A1 specifically induces *Salmonella* metal uptake systems but no other stress responses.

Biozentrum, University of Basel, CH-4056 Basel, Switzerland
*Present address: University of Oxford, Department of Zoology, Oxford OX1 3QU, UK.
†Corresponding author. Email: dirk.bumann@unibas.ch

Downloaded from http://science.sciencemag.org/ on November 29, 2019

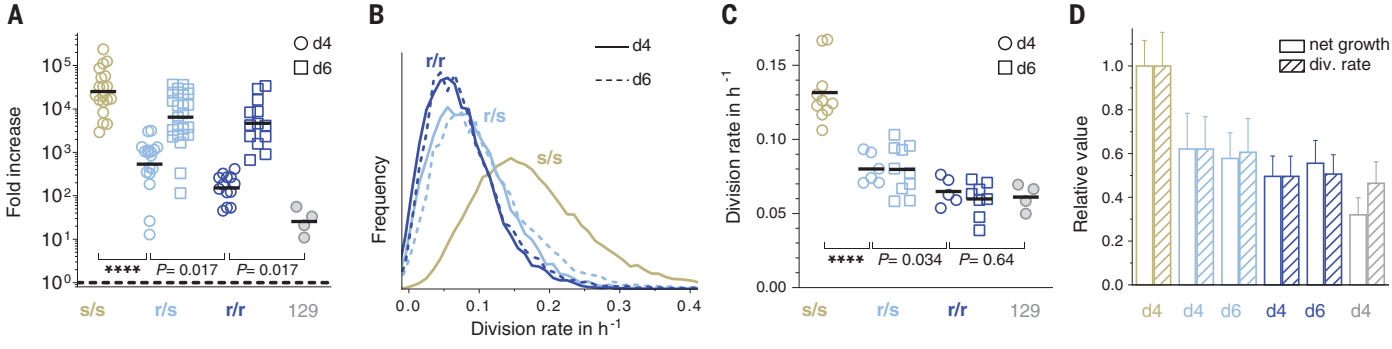


Fig. 1. SLC11A1 restricts *Salmonella* proliferation in mouse spleen. (**A**) Growth of *Salmonella* in spleens of *SLC11A1*$^{s/s}$ (s/s), *SLC11A1*$^{r/s}$ (r/s), *SLC11A1*$^{r/r}$ (r/r), or 129S2/SvPasCrl (129) mice [these data were already reported in (*3*)]. Each symbol represents a single mouse. d4 and d6, data from days 4 and 6, respectively. *****P* < 0.0001, one-way analysis of variance (ANOVA) of log-transformed values with Holm-Šídák correction for multiple comparisons. (**B**) *Salmonella* in vivo division rates as calculated from TIMER^{bac} data shown in fig. S2D. (**C**) Median *Salmonella* division rates in different mouse genotypes. Each symbol represents a single mouse. *****P* < 0.0001, one-way ANOVA of log-transformed values with Holm-Šídák correction. (**D**) Comparison of *Salmonella* net growth based on spleen loads shown in (A) and *Salmonella* division rates shown in (C).

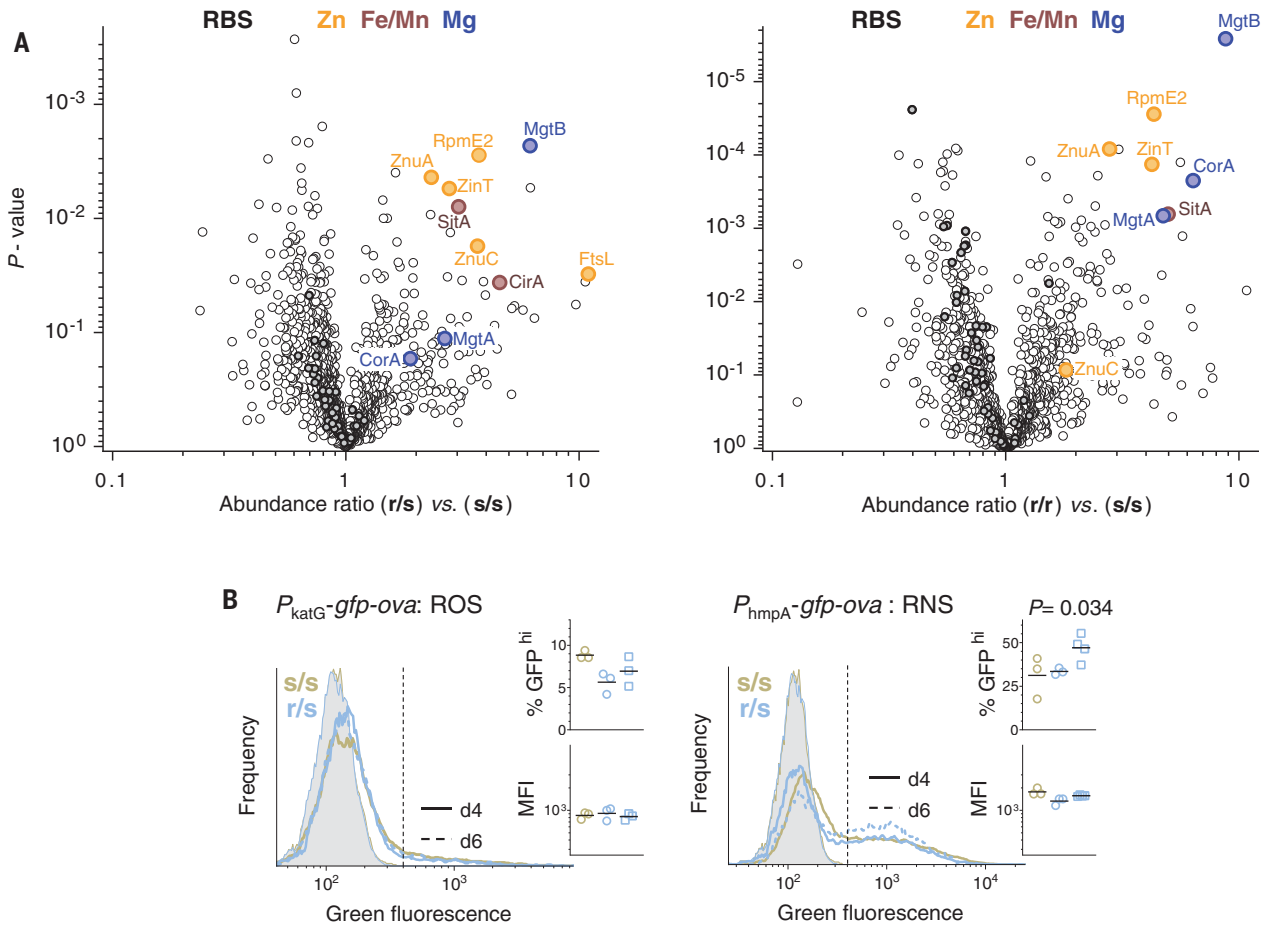


Fig. 2. SLC11A1 triggers *Salmonella* metal uptake. (**A**) Proteome comparisons of *Salmonella* cells sorted from spleens of mice with different *SLC11A1* genotypes [(s/s), day 4 after low-dose infection; (r/s), day 6 after low-dose infection; (r/r), day 4 after high-dose infection]. Each circle represents a protein. Metal uptake systems are shown in colors. RpmE2 increases the cytosolic zinc concentration by replacing the zinc-binding ribosomal subunit L31 (*30*). Ribosomal proteins (RBS) are shown in black. (**B**) Fluorescence of *Salmonella* reporter strains for reactive oxygen species (ROS) or reactive nitrogen species (RNS). Gray areas, autofluorescence; dashed lines, threshold for defining GFP^{hi} subsets; MFI, median green fluorescent protein (GFP) fluorescence intensities of GFP^{hi} subsets. Each circle represents a single mouse. More *Salmonella* cells had high P_{hmpA} activity at day 6 in (r/s) mice, consistent with increasing nitric oxide exposure during disease progression (*5*). In vitro induction data for both promoters are shown in fig. S3b.

Salmonella encodes transporters with different affinities for divalent metal cations (Fig. 3A; table S1). We compared the fitness of *Salmonella* strains with single and multiple gene deletions (detailed description in the supplementary text) in littermates from (r/s) × (s/s) crosses. We analyzed (s/s) mice after 4 days and (r/s) mice after 6 days, to match the number of *Salmonella* in vivo divisions for the two groups (Fig. 3B and fig. S4). For relevant mutants, we additionally analyzed (r/r) mice at days 4 and 6 (Fig. 3C and fig. S4).

The quadruple iron-uptake mutant c had compromised fitness in (s/s) mice, a stronger defect in (r/s) mice, and almost no fitness in (r/r) mice. Available iron concentrations in vivo seemed to be generally insufficient to meet the high requirements of this mutant (30 μM iron in vitro) (Fig. 3D), and SLC11A1 further restricted iron access (*7*). By contrast, the double iron-uptake mutant b retained full fitness in all three mouse genotypes, indicating sufficient iron supply to meet the lower requirements of this mutant (3 μM iron in vitro) (Fig. 3D). Wild-type *Salmonella* required even less iron and should thus not be limited by iron starvation either. SLC11A1 thus reduced iron availability, but this was insufficient for limiting replication of wild-type *Salmonella*.

Mutant e, lacking the high-affinity zinc ABC transporter, had poor fitness in both (s/s) and (r/s) mice, indicating limited zinc supply (*8*). By contrast, mutant d, with a partial zinc uptake defect (*9*), had a slight fitness defect in (s/s) and (r/s) but not (r/r) mice, indicating nearly sufficient access to zinc in all three mouse genotypes. Wild-type *Salmonella* with full zinc uptake capabilities would thus not be limited by zinc limitation either. Alternatively, a low zinc concentration might activate host endosomal hydrolases, such as cathepsins, that could damage *Salmonella* cells (*10*). However, cathepsin activities did not colocalize with *Salmonella* in (s/s) and (r/s) mice (fig. S5) (*11*). Mutants lacking uptake systems for manganese (mutant f) or cobalt (mutant h) or exporters for manganese (mutant g) or copper (mutant i) had no differential fitness defects, arguing against contributions of these metals to SLC11A1-mediated resistance. Although transition metals might become more important at later disease stages with distinct host control mechanisms (*12*) and decreasing amounts of nonheme iron in spleen (*13*, *14*), particularly in genetic backgrounds with lower serum iron concentrations (*15*), our data indicate that transition metal starvation is not required for SLC11A1-mediated resistance.

We also analyzed *Salmonella* mutants with defects in uptake systems for the main group metal magnesium, because its transporter MgtB was highly up-regulated in the presence of SLC11A1 (Fig. 2A). The double mutant l, with

Downloaded from http://science.sciencemag.org/ on November 29, 2019

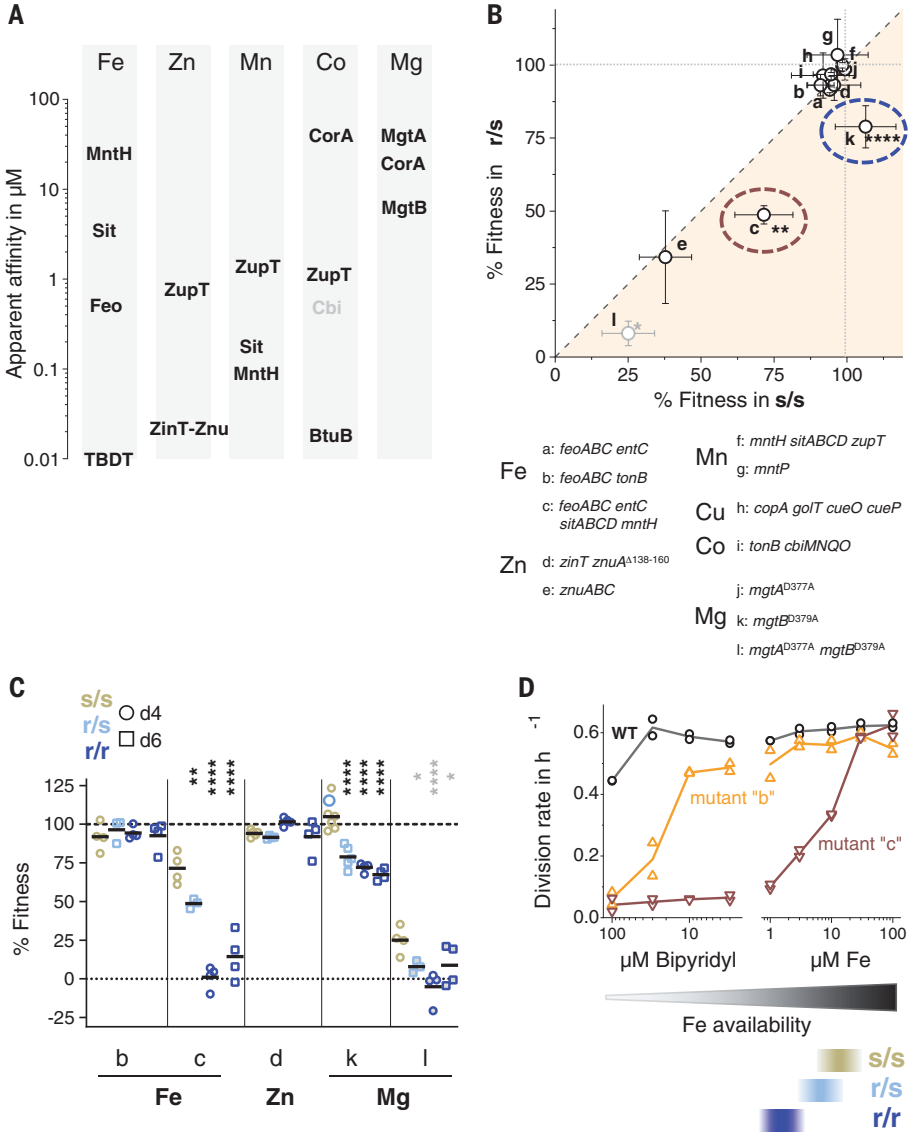


Fig. 3. SLC11A1 causes a specific requirement for the *Salmonella* magnesium transporter MgtB. (**A**) Apparent affinities of *Salmonella* uptake systems for various metals (see table S1 for numerical values). TBDT, TonB-dependent transporter. The CbiMNQ system transports Co^{2+}, but affinities have not yet been reported. (**B**) Fitness of *Salmonella* mutants relative to wild-type in mice with different *SLC11A1* genotypes. Data points in the shaded area indicate stronger fitness defects in the presence of SLC11A1 (dashed ellipses). Averages and standard deviations for three to five mice are shown. ****$P < 0.0001$; **$P < 0.01$; *$P < 0.05$ (two-tailed t test with Holm-Šídák correction). Mutant l is shown in gray to indicate that some data points were below the limit of detection. Corresponding competitive indices (CI) are shown in fig. S4. (**C**) Fitness of selected *Salmonella* mutants. Each symbol represents an individual mouse. Statistical significance of differences to fitness in (s/s) mice was tested using a one-way ANOVA with Holm-Šídák correction. ****$P < 0.0001$; ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. (**D**) Growth of wild-type and mutant *Salmonella* in vitro. Minimal medium was supplemented with 1 µM $FeSO_4$ and variable amounts of 2,2′-bipyridyl (left side), or variable amounts of $FeSO_4$ and no 2,2′-bipyridyl (right side). Individual data and means of two biological replicates are shown. Coarse estimates for in vivo iron availabilities are shown below.

inactivated MgtA and MgtB, had severe fitness defects in all three mouse genotypes (*16*), suggesting that (i) the third magnesium transporter CorA did not contribute to magnesium uptake in vivo (*16*) and (ii) magnesium availability in vivo was insufficient to meet the requirements of this mutant. Mutant l might also suffer from dose-dependent iron accumulation (*17*), although this should cause the most severe fitness defects in (s/s) mice with the highest iron supply (see above), and this was inconsistent with our observations. Mutant j with inactivated MgtA retained full fitness in all mouse genotypes (*18*), whereas mutant k, with inactivated MgtB, had full fitness in (s/s) mice but compromised fitness in (r/s) and (r/r) mice (Fig. 3, B and D; fig. S3), suggesting that there is sufficient magnesium availability only in (s/s) mice. An independently constructed Δ*mgtB* mutant had similar differential fitness phenotypes. SLC11A1 thus caused a specific requirement for MgtB [apparent Michaelis constant (K_m) of 6 µM for Mg^{2+} (*19*)], whereas MgtA, with a slightly lower affinity [apparent K_m of 29 µM (*19*)], was dispensable in the presence of MgtB and compensated only partially for its absence. Attenuation of *Salmonella* Δ*mgtB* in genetically resistant C3H/HeN mice (*20*), but not in susceptible BALB/c mice (*18*), has previously been reported. Our data link this phenotype to SLC11A1 instead of the many other allelic differences between C3H/HeN and BALB/c mice and suggest that SLC11A1 critically limits magnesium availability for *Salmonella*.

To determine how *mgtB* expression relates to the markedly heterogeneous replication of *Salmonella* single cells (Fig. 1B), we combined a *timer*^{bac} expression cassette reporting single-cell division rates (*3*) with a transcriptional fusion of the $P_{mgtCBRcigR}$ promoter driving *mgtB* expression to *bfp* (fig. S6, a to c). *Salmonella* *mgtB* expression and its division rate were inversely correlated in (r/s) and (r/r) mice, but not in (s/s) mice (Fig. 4, A and B). In the presence of SLC11A1, poorly replicating *Salmonella* cells thus maximized Mg^{2+} uptake through the crucial transporter MgtB, possibly to compensate for magnesium limitation. To test this hypothesis, we deprived *Salmonella* of magnesium by an independent mechanism and determined the effects in susceptible (s/s) mice that normally provide sufficient magnesium for *Salmonella* growth (see above). Consistent with our fitness data (Fig. 3B), a *Salmonella* *mgtB* Asp^{379}→Arg (D379A) mutant had single-cell division rates that were similar to that of the wild type (Fig. 4C). Additional inactivation of MgtA abolished replication and triggered exceedingly high *mgtB* expression, suggesting that *Salmonella* cells tried to maximize their Mg^{2+} uptake capabilities (Fig. 4, A and C). For less severe growth restriction, we preserved

Downloaded from http://science.sciencemag.org/ on November 29, 2019

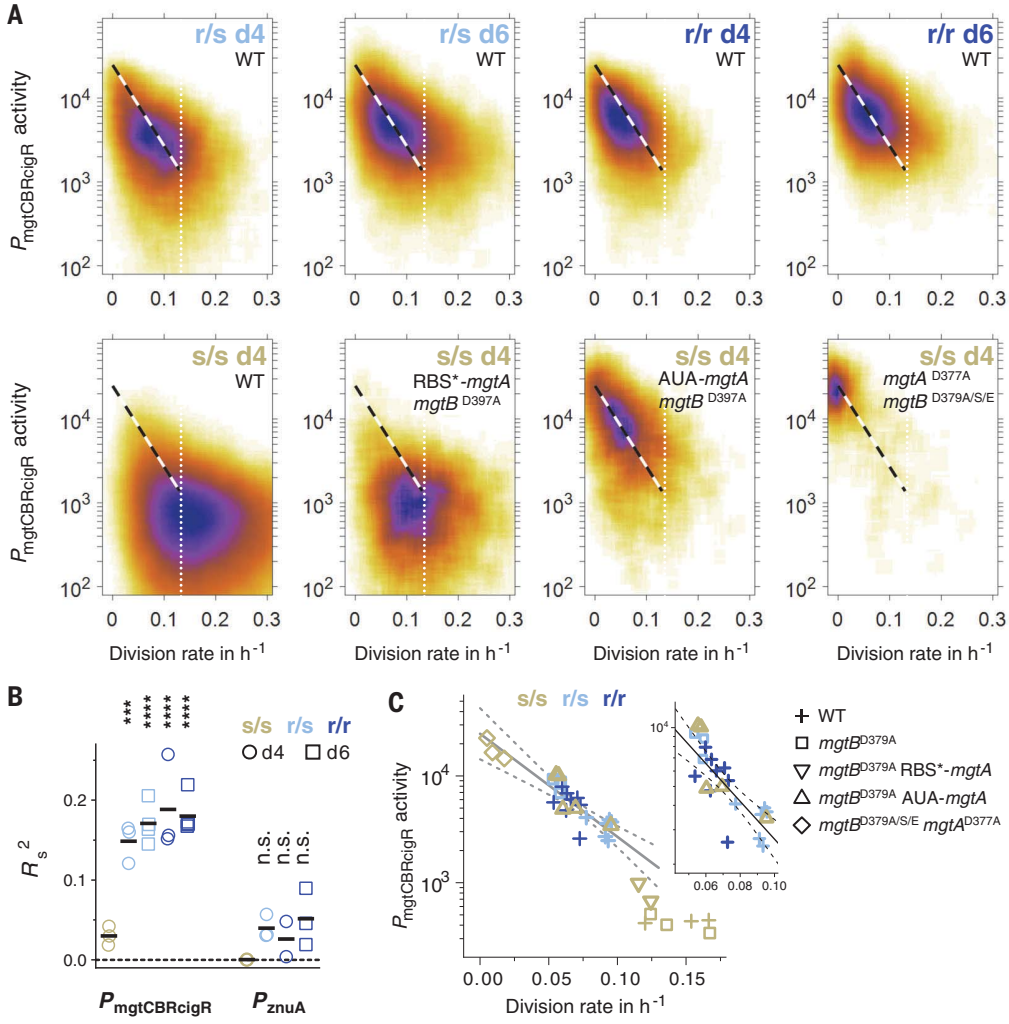


Fig. 4. Equivalent *Salmonella* single-cell responses to SLC11A1 and magnesium deprivation. (**A**) Single-cell density maps of *Salmonella* division rates and $P_{mgtCBRcigR}$ activities (pooled data from three to four mice). The dashed lines represent a fit to median values of (r/s) and (r/r) mice as shown in (C). The dotted vertical lines show the median *Salmonella* division rate in (s/s) mice. (**B**) Squared values (goodness of fit) for Spearman's rank correlation between single-cell *Salmonella* division rates and promoter activities. Each symbol represents a single mouse. ****$P < 0.0001$; ***$P < 0.001$ (statistical significance of difference to values for (s/s) mice based on a one-way ANOVA with Holm-Šídák correction). (**C**) Median division rates and $P_{mgtCBRcigR}$ activities for *Salmonella* and mouse genotypes shown in (A). Each symbol represents a single mouse. The black line represents a linear regression of data for (r/s) and (r/r) mice with 90% confidence bands shown as dashed lines. The inset is an enlargement of the most informative data range.

functional MgtA but reduced its abundance with an inefficient ribosomal binding site or an unconventional initiation codon. Whereas the *Salmonella mgtB* D379A RBS*-*mgtA* mutant had slightly reduced division rates, *Salmonella mgtB* D379A AUA-*mgtA* had substantially reduced division rates in (s/s) mice that approximated those of wild-type *Salmonella* in (r/s) and (r/r) mice (Fig. 4, A and C). By impairing magnesium uptake, we could thus achieve similar levels of growth restriction as imposed by SLC11A1. This was also associated with superimposable single-cell *mgtB* expression (Fig. 4, A and C). Individual *Salmonella* cells thus showed qualitatively and quantitatively equivalent responses to magnesium deprivation because of defective uptake as observed for growth limitation by SLC11A1. By contrast, expression of the critical zinc transporter *znuA* was poorly correlated with *Salmonella* division rate in all three mouse genotypes (Fig. 4B) and showed discordant effects of limiting zinc uptake (fig. S7, a and b), consistent with the data on mutant fitness and arguing against relevant zinc deprivation (Fig. 3, A and B). Together, these data identify magnesium deprivation as the main resistance mechanism of SLC11A1.

These findings were initially surprising, as SLC11A1 has been shown to transport Fe^{2+}, Mn^{2+}, and Co^{2+} (*1*). However, the apparent affinities for these metals are in the low micromolar range (*21*), and it is unclear how SLC11A1 outcompetes corresponding *Salmonella* transporters with 10 to 1000 times higher affinities (table S1). On the other hand, Mg^{2+} is the most abundant divalent cation in mammalian cells, with free concentrations around 0.5 to 1 mM in the cytosol, as well as in *Salmonella*-containing vacuoles early after phagocytosis (*22*–*24*). Bacterial cells require around 20 mM Mg^{2+} in their cytosol (*25*), and proliferating *Salmonella* may quickly exhaust available Mg^{2+} in the vacuole, as leukocyte membranes represent tight barriers for Mg^{2+} (*22*, *23*). Indeed, Mg^{2+} transport becomes relevant for *Salmonella* growth several hours after phagocytosis (*26*), but vacuolar Mg^{2+} concentrations at such late time points are currently inaccessible (*24*). SLC11A1 could restrict *Salmonella* access to Mg^{2+} by modulating fusion of *Salmonella*-containing vacuoles with Mg^{2+}-containing vesicles (*27*), or SLC11A1 might directly transport Mg^{2+}, similar to some bacterial SLC11A1 orthologs (*28*, *29*). Addressing these issues will require overcoming challenges in purifying active SLC11A1 and establishment of Mg^{2+} transport assays without the unavailable radioisotope ^{28}Mg. More generally, this study demonstrates how integrated single-cell analysis of pathogens can elucidate relevant mechanisms of pleiotropic host resistance factors. This strategy may be applicable to unravel mechanisms that control microbial growth in diverse environments.

REFERENCES AND NOTES

1. M. Wessling-Resnick, *J. Biol. Chem.* **290**, 18984–18990 (2015).
2. W. P. Loomis *et al.*, *PLOS ONE* **9**, e111763 (2014).
3. B. Claudi *et al.*, *Cell* **158**, 722–733 (2014).
4. W. H. Benjamin Jr., P. Hall, S. J. Roberts, D. E. Briles, *J. Immunol.* **144**, 3143–3151 (1990).
5. N. A. Burton *et al.*, *Cell Host Microbe* **15**, 72–83 (2014).
6. J. K. White, P. Mastroeni, J. F. Popoff, C. A. Evans, J. M. Blackwell, *J. Leukoc. Biol.* **77**, 311–320 (2005).
7. M. Nairz *et al.*, *Cell. Microbiol.* **11**, 1365–1381 (2009).

Downloaded from http://science.sciencemag.org/ on November 29, 2019

8. S. Campoy *et al*., *Infect. Immun.* **70**, 4721–4725 (2002).
9. P. Petrarca, S. Ammendola, P. Pasquali, A. Battistoni, *J. Bacteriol.* **192**, 1553–1564 (2010).
10. T. D. Lockwood, *Metallomics* **5**, 110–124 (2013).
11. K. McGourty *et al*., *Science* **338**, 963–967 (2012).
12. P. Mastroeni *et al*., *J. Exp. Med.* **192**, 237–248 (2000).
13. D. E. Brown *et al*., *Infect. Immun.* **83**, 2290–2299 (2015).
14. D. E. Brown *et al*., *Vet. Pathol.* **50**, 867–876 (2013).
15. R. E. Fleming *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2707–2711 (2001).
16. K. M. Papp-Wallace *et al*., *J. Bacteriol.* **190**, 6517–6523 (2008).
17. S. Chamnongpol, E. A. Groisman, *Mol. Microbiol.* **44**, 561–571 (2002).
18. A. B. Blanc-Potard, E. A. Groisman, *EMBO J.* **16**, 5376–5385 (1997).
19. M. D. Snavely, S. A. Gravina, T. T. Cheung, C. G. Miller, M. E. Maguire, *J. Biol. Chem.* **266**, 824–829 (1991).
20. E. Choi *et al*., *Sci. Rep.* **7**, 42098 (2017).
21. J. R. Forbes, P. Gros, *Blood* **102**, 1884–1892 (2003).
22. M. E. Maguire, J. A. Cowan, *Biometals* **15**, 203–210 (2002).
23. A. M. Romani, *Arch. Biochem. Biophys.* **512**, 1–23 (2011).
24. N. Martin-Orozco *et al*., *Mol. Biol. Cell* **17**, 498–510 (2006).
25. C. E. Outten, T. V. O'Halloran, *Science* **292**, 2488–2492 (2001).
26. R. L. Smith, M. T. Kaczmarek, L. M. Kucharski, M. E. Maguire, *Microbiology* **144**, 1835–1843 (1998).
27. P. Cuellar-Mata *et al*., *J. Biol. Chem.* **277**, 2258–2265 (2002).
28. J. H. Shin *et al*., *PLOS Genet.* **10**, e1004429 (2014).
29. A. T. Bozzi *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10310–10315 (2016).
30. M. P. Hensley *et al*., *J. Inorg. Biochem.* **111**, 164–172 (2012).

ACKNOWLEDGMENTS

We thank B. Claudi, P. Marini, P. Maturana, and F. R. Cianfanelli for help with cloning and animal experiments, P. Pelczar (University of Basel Center for Transgenic Models) for help with genome editing, S. Stevanova and J. Bögli (Biozentrum FACS Core Facility) for sorting, S. Söderholm and T. Bock (Biozentrum Proteomics Core Facility) for help with proteomics, and S. Thomforde for correcting grammatical errors. **Funding:** This research was supported by the Swiss National Science Foundation (projects 310030_156818 and 310030_182315) to D.B. **Author contributions:** Outlined the study: D.B. and O.C.; performed experiments: O.C.; analyzed data: O.C. and D.B.; interpreted data: O.C. and D.B.; wrote manuscript: D.B. with major contributions from O.C. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. Requests for materials should be addressed to D.B.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/995/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Table S1
References (*31*–*81*)

23 April 2019; resubmitted 12 August 2019
Accepted 25 October 2019
10.1126/science.aax7898

Downloaded from http://science.sciencemag.org/ on November 29, 2019

COMPUTER SCIENCE

Preventing undesirable behavior of intelligent machines

Philip S. Thomas[1]*, Bruno Castro da Silva[2], Andrew G. Barto[1], Stephen Giguere[1], Yuriy Brun[1], Emma Brunskill[3]

Intelligent machines using machine learning algorithms are ubiquitous, ranging from simple data analysis and pattern recognition tools to complex systems that achieve superhuman performance on various tasks. Ensuring that they do not exhibit undesirable behavior—that they do not, for example, cause harm to humans—is therefore a pressing problem. We propose a general and flexible framework for designing machine learning algorithms. This framework simplifies the problem of specifying and regulating undesirable behavior. To show the viability of this framework, we used it to create machine learning algorithms that precluded the dangerous behavior caused by standard machine learning algorithms in our experiments. Our framework for designing machine learning algorithms simplifies the safe and responsible application of machine learning.

Machine learning (ML) algorithms are having an increasing impact on modern society. They are used by geologists to predict landslides (*1*) and by biologists working to create a vaccine for HIV (*2*); they also influence criminal sentencing (*3*), control autonomous vehicles (*4*), and enable medical advances (*5*). The potential for ML algorithms to cause harm—including catastrophic harm—is therefore a pressing concern (*6*). Despite the importance of this problem, current ML algorithms do not provide their users with an effective means for precluding undesirable behavior, which makes the safe and responsible use of ML algorithms difficult. We introduce a framework for designing ML algorithms that allow their users to easily define and regulate undesirable behavior. This framework does not address the problem of imbuing intelligent machines with a notion of morality or human-like values (*7*), nor the problem of avoiding undesirable behavior that the user never considered (*8*). Rather, it provides a remedy for the problem of ML algorithms that exhibit undesirable behavior because their users did not have an effective way to specify and constrain such behavior.

The first step of the current standard approach for designing ML algorithms, which we refer to as the standard ML approach, is to define mathematically what the algorithm should do. At an abstract level, this definition is the same across all branches of ML: Find a solution θ^*, within a feasible set Θ, that maximizes an objective function $f\colon \Theta \to \mathbb{R}$. That is, the goal of the algorithm is to find a solution in

$$\arg\max_{\theta \in \Theta} f(\theta) \tag{1}$$

Note that the algorithm does not know $f(\theta)$ for any $\theta \in \Theta$ (e.g., the true mean squared error); it can only reason about it from data (e.g., by using the sample mean squared error).

One problem with the standard ML approach is that the user of an ML algorithm must encode constraints on the algorithm's behavior in the feasible set or the objective function. Encoding constraints in the objective function [e.g., using soft constraints (*9*) or robust and risk-sensitive approaches (*10*)] requires extensive domain knowledge or additional data analysis to properly balance the relative importance of the primary objective function and the constraints. Similarly, encoding constraints in the feasible set [e.g., using hard constraints (*9*), chance constraints (*11*), or robust optimization approaches (*12*)] requires knowledge of the probability distribution from which the available data are sampled, which is often not available.

Our framework for designing ML algorithms allows the user to constrain the behavior of the algorithm more easily, without requiring extensive domain knowledge or additional data analysis. This is achieved by shifting the burden of ensuring that the algorithm is well-behaved from the user of the algorithm to the designer of the algorithm. This is important because ML algorithms are used for critical applications by people who are experts in their fields, but who may not be experts in ML and statistics.

We now define our framework. Let D, called the data, be the input to the ML algorithm. For example, in the classification setting, D is not a single labeled training example but rather all of the available labeled training examples. D is a random variable and the source of randomness in our subsequent statements regarding probability. An ML algorithm is a function a, where $a(D)$ is the solution output by the algorithm when trained on data D. Let Θ be the set of all possible solutions that an ML algorithm could output. Our framework mathematically defines what an algorithm should do in a way that allows the user to directly place probabilistic constraints on the solution, $a(D)$, returned by the algorithm. This differs from the standard ML approach wherein the user can only indirectly constrain $a(D)$ by restricting or modifying the feasible set Θ or objective function f. Concretely, algorithms constructed using our framework are designed to satisfy constraints of the form $\Pr(g(a(D)) \leq 0) \geq 1 - \delta$, where $g\colon \Theta \to \mathbb{R}$ defines a measure of undesirable behavior (as illustrated later by example) and $\delta \in [0, 1]$ limits the admissible probability of undesirable behavior.

Note that in these constraints, D is the only source of randomness; we denote random variables by capital noncalligraphic letters to make clear which terms are random in statements of probability and expectation. Because these constraints define which algorithms a are acceptable (rather than which solutions θ are acceptable), they must be satisfied during the design of the algorithm rather than when the algorithm is applied. This shifts the burden of ensuring that the algorithm is well-behaved from the user to the designer.

Using our framework for designing ML algorithms involves three steps:

1) Define the goal for the algorithm design process. The designer of the algorithm writes a mathematical expression that expresses a goal—in particular, the properties that the designer wants the resulting algorithm a to have. This expression has the following form, which we call a Seldonian optimization problem after a fictional character (*13*):

$$\arg\max_{a \in \mathcal{A}} f(a)$$
$$\text{s.t. } \forall i \in \{1, \ldots, n\}, \Pr\Big(g_i(a(D)) \leq 0\Big) \geq 1 - \delta_i \tag{2}$$

where $\mathcal{A}$ is the set of all algorithms that will be considered by the designer, $f\colon \mathcal{A} \to \mathbb{R}$ is now an objective function that quantifies the utility of an algorithm, and we allow for $n \geq 0$ constraints, each defined by a tuple (g_i, δ_i), where $i \in \{1, \ldots, n\}$. Note that this is in contrast to the standard ML approach: In the standard ML approach, Eq. 1 defines the goal of the algorithm, which is to produce a solution with a given set of properties, whereas in our framework, Eq. 2 defines the goal of the designer, which is to produce an algorithm with a given set of properties.

2) Define the interface that the user will use. The user should have the freedom to specify one or more g_i that capture the user's own definition of undesirable behavior. This requires the algorithm a to be compatible with many different definitions of g_i. The designer should therefore specify the class of possible definitions of g_i with which the algorithm will be

[1]University of Massachusetts, Amherst, MA, USA. [2]Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil. [3]Stanford University, Stanford, CA, USA.
*Corresponding author. Email: pthomas@cs.umass.edu

Downloaded from http://science.sciencemag.org/ on November 28, 2019

compatible, and should provide a means for the user to tell the algorithm which definition of g_i should be used, without requiring the user to have knowledge of the distribution of D or even the value $g_i(\theta)$ for any $\theta \in \Theta$. Below, we provide examples of how this can be achieved.

3) Create the algorithm. The designer creates an algorithm a, which is a (possibly approximate) solution to Eq. 2 from step 1 and which allows for the class of g_i chosen in step 2. In practice, designers rarely produce algorithms that cannot be improved upon, which implies that they may only find approximate solutions to Eq. 2. Our framework allows for this by requiring a to satisfy only the probabilistic constraints while attempting to optimize f; we call such algorithms Seldonian. We call an algorithm quasi-Seldonian if it relies on reasonable but false assumptions, such as appeals to the central limit theorem. See (*14*) for further discussion regarding the benefits and limitations of quasi-Seldonian algorithms.

Once a Seldonian algorithm has been designed, a user can apply it by specifying one or more g_i (belonging to the class of g_i chosen in step 2 above) to capture the user's desired definition of undesirable behavior, and specifying δ_i, the maximum admissible probability of the undesirable behavior characterized by g_i.

To show the viability of our framework, we used it to design regression, classification, and reinforcement learning algorithms. Constraining the behavior of regression and classification algorithms is important because, for example, they have been used for medical applications where undesirable behavior could delay cancer diagnoses (*15*), and because they have been shown to sometimes cause racist, sexist, and other discriminatory behavior (*3*, *16*). Similarly, reinforcement learning algorithms have been proposed for applications where undesirable behavior can cause financial losses (*17*), environmental damage (*18*), and even death (*19*). The Seldonian algorithms and applications we present below are illustrations to show that it is possible and tractable to design Seldonian algorithms that can tackle important problems of interest. Note that these are intended only as proof of principle; the primary contribution of this work is the framework itself rather than any specific algorithm or application. Like the common application of classification algorithms (*20*) to the Wisconsin breast cancer dataset (*21*), the applications validate our ML algorithms as tools that researchers with medical expertise and domain knowledge could apply (*22*), but do not imply that our learned solutions (classifiers or policies) should be deployed as-is to any particular real-world problem.

The regression algorithm that we designed attempts to minimize the mean squared error of its predictions while ensuring that, with high probability, a statistic of interest, $g(\theta)$, of the returned solution, $\theta = a(D)$, is bounded. The definition of this statistic can be chosen by the user to capture a particular definition of undesirable behavior (e.g., the expected financial loss that results from using a given solution θ). The user may not know the value of this statistic for even a single solution. We must therefore provide the user with a way to tell our algorithm the statistic to be bounded, without requiring the user to provide the value, $g(\theta)$, of the statistic for different solutions θ (see step 2 above). To achieve this (*14*), we allow the user to specify a sample statistic $\hat{g}(\theta, D)$, and we define $g(\theta)$ to be the expected value of this sample statistic: $g(\theta) = \mathbf{E}[\hat{g}(\theta, D)]$, where $\mathbf{E}$ denotes expected value.

The creation of a regression algorithm (step 3) with the properties specified during steps 1 and 2 is challenging. This is to be expected given the shifted burden discussed previously; see (*14*) for a detailed description of how we performed step 3 when designing all of the Seldonian algorithms that we present. Figure 1 overviews our regression algorithms.

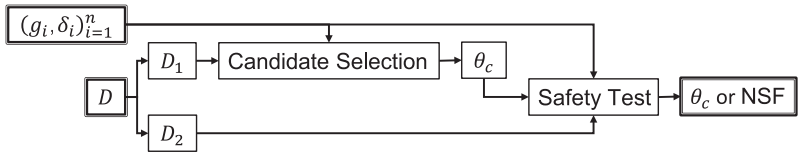


Fig. 1. Overview of Seldonian regression algorithms. The algorithm takes the behavioral constraints $(g_i, \delta_i)_{i=1}^n$ and training data D as input and outputs either a solution θ_c or NSF (no solution found). First, the data are partitioned into two sets, D_1 and D_2. Next, a routine called Candidate Selection uses D_1 to select a single solution, the candidate solution θ_c, which it predicts will perform well under the primary objective f while also being likely to pass the subsequent safety test based on knowledge of the specific form of the test. The Safety Test mechanism checks whether the algorithm has sufficient confidence that $g_i(\theta_c) \leq 0$ for each constraint $i \in \{1, ..., n\}$. If so, it returns the candidate solution θ_c, otherwise it returns NSF. The Safety Test routine uses standard statistical tools such as Student's t test and Hoeffding's inequality to transform sample statistics computed from D_2 into bounds on the probability that $g(a(D)) > 0$ (i.e., bounds on the probability of undesirable behavior).

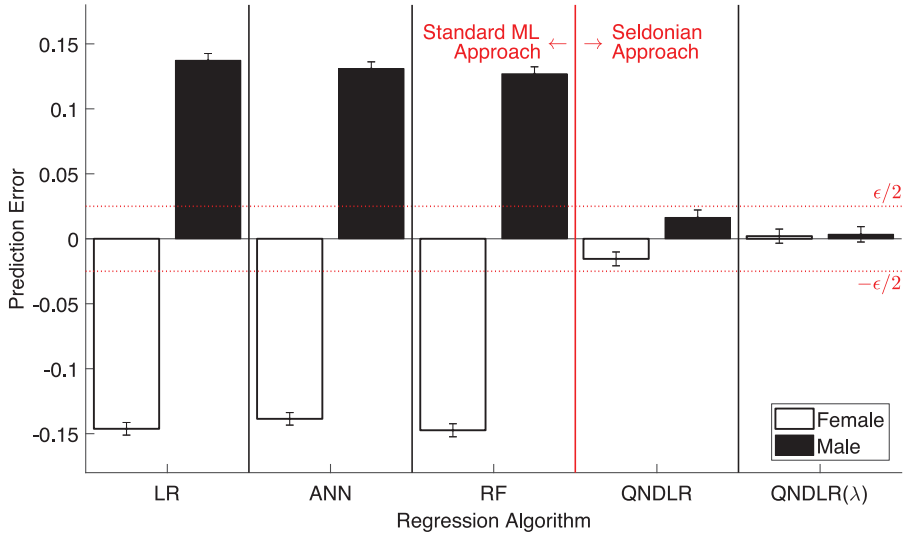


Fig. 2. Seldonian regression algorithm applied to GPA prediction. We used five different regression algorithms to predict students' GPAs during their first three semesters at university based on their scores on nine entrance exams. We used actual data from 43,303 students from Brazil. Here, the user-selected definition of undesirable behavior corresponds to large differences in mean prediction errors (mean predicted GPA minus mean observed GPA) for applicants of different genders. This plot shows the mean prediction errors (±SD) for male and female students when using each regression algorithm. We used three standard ML algorithms—least squares linear regression (LR) (*40*), an artificial neural network (ANN) (*41*), and a random forest (RF) (*42*)—and two variants of our Seldonian algorithm: QNDLR and QNDLR(λ). All shown standard ML methods tend to notably overpredict the performance of male students and underpredict the performance of female students, whereas the two variants of our Seldonian regression algorithm do not. In particular, our algorithms ensure that, with approximately 95% probability, the expected prediction errors for men and women will be within $\varepsilon = 0.05$, and both effectively preclude the sexist behavior that was exhibited by the standard ML algorithms.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Recent methods designed particularly for algorithmic fairness in regression tasks (*23*), developed in parallel to our own, do not give users the freedom to select their own desired definitions of undesirable behavior, nor do they provide guarantees on the avoidance of such behavior.

We applied a variant of our Seldonian regression algorithm to the problem of predicting students' grade point averages (GPAs) during their first three semesters at university on the basis of their scores on nine entrance exams; we used a sample statistic that captures one form of discrimination (sexism). Note that our algorithm is not particular to the chosen measure of discrimination; see (*14*) for a discussion of other definitions of fairness. Figure 2 presents the results of this experiment, showing that commonly used regression algorithms designed using the standard ML approach can discriminate against female

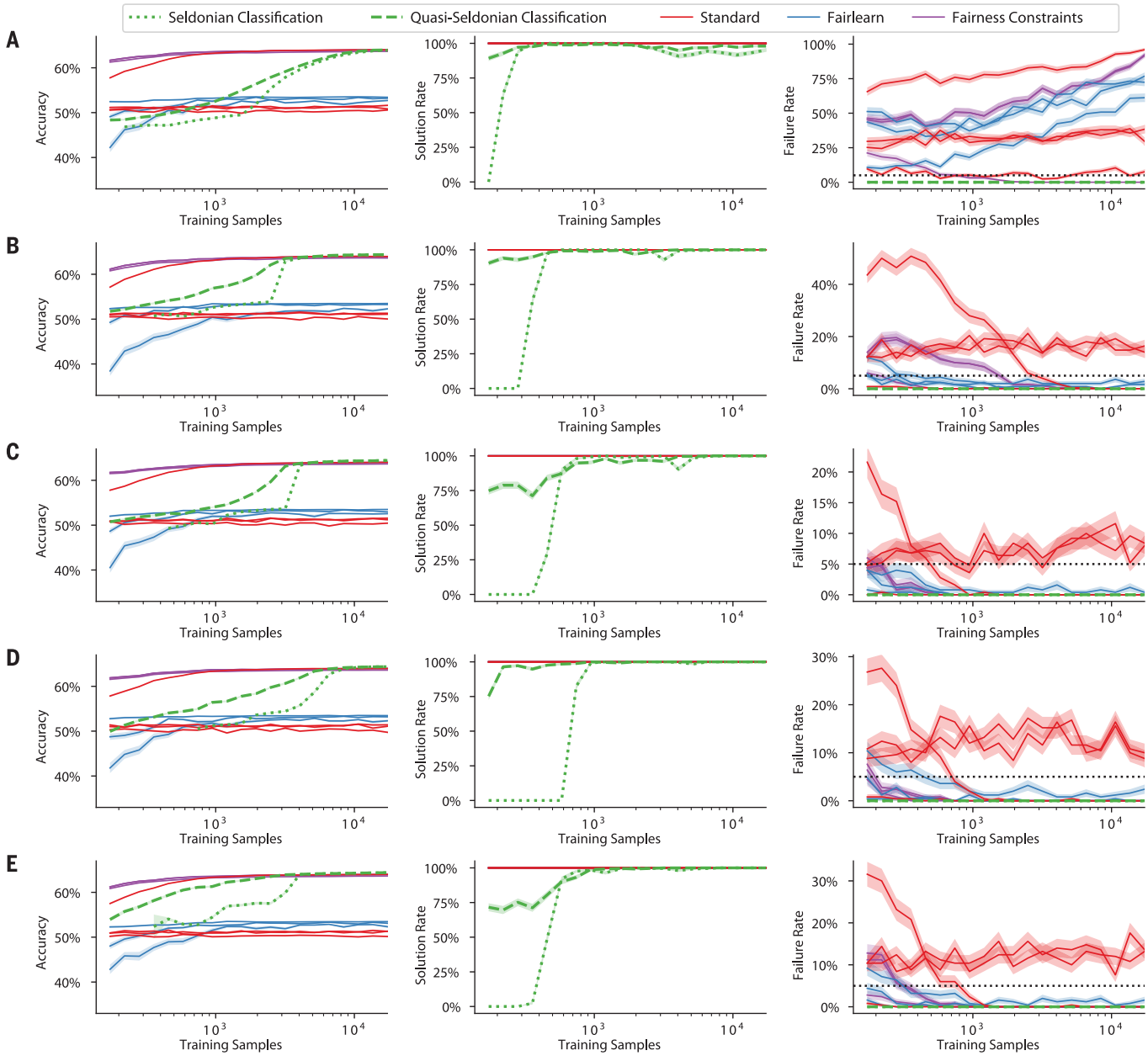


Fig. 3. Seldonian classification algorithm applied to GPA prediction. We applied using classification algorithms to predict whether student GPAs will be above 3.0. Shaded regions represent SE over 250 trials. The curves labeled "Standard" correspond to common classification algorithms designed using the standard ML approach; the multiple curves for Fairlearn and Fairness Constraints correspond to different hyperparameter settings for each algorithm (*14*). Each row corresponds to a different fairness definition: (**A**) disparate impact, (**B**) demographic parity, (**C**) equal opportunity, (**D**) equalized odds, (**E**) predictive equality. The horizontal axes of all plots correspond to the amount of training data and have logarithmic scale. The left column shows the accuracy of the trained classifiers, the center column shows the probability that each algorithm returned a solution (non-Seldonian algorithms always returned solutions), and the right column shows the probability that each classifier violated a behavioral constraint. When showing the failure rate of each algorithm, the horizontal dashed line corresponds to 100δ%, where δ = 0.05. In all cases, the Seldonian and quasi-Seldonian algorithms returned solutions using a reasonable amount of data (center), did so without significant losses to accuracy (left), and were the only algorithms to reliably enforce all five fairness definitions (right).

Downloaded from http://science.sciencemag.org/ on November 28, 2019

students when applied without considerations for fairness. In contrast, the user can easily limit the observed sexist behavior in Fig. 2 using our Seldonian regression algorithm.

To emphasize that Seldonian algorithms are compatible with a variety of definitions of fairness and to better situate our research relative to current state-of-the-art fairness-aware ML algorithms, we present a Seldonian classification algorithm (*14*). This classification algorithm differs from our regression algorithm in its primary objective (classification loss rather than mean squared error) and in its more sophisticated interface, which allows the user to type an expression that defines $g(\theta)$ in terms of common statistics (such as the false negative rate or false positive rate given that the protected attribute, here gender, takes a specific value), constants, operators (such as addition, division, and absolute value), and statistics for which the user can provide unbiased estimates, as in the regression example. We applied our classification algorithm to predicting whether student GPAs will be above 3.0 using the dataset described in Fig. 2, while constraining five popular definitions of fairness for classification (Fig. 3). The Seldonian classification algorithm properly limited the specified form of unfair behavior across all trials. Unlike our approach, fairness-aware classification algorithms designed using the standard ML approach do not provide probabilistic guarantees that the resulting classifier is acceptably fair when applied to unseen data. We observed that two state-of-the-art fairness-aware algorithms that we ran for comparison, Fairlearn (*24*) and Fairness Constraints (*25*), each produced unfair behavior under at least one definition of fairness.

Next, we used our framework to design a general-purpose Seldonian reinforcement learning algorithm: one that, unlike regression and classification algorithms, makes a sequence of dependent decisions. In this context, a solution θ is called a policy; a history H (a random variable) denotes the outcome of using a policy to make a sequence of decisions; and the available data D is a set of histories produced by some initial policy θ_0. Because it is Seldonian, our algorithm searches for an optimal policy while ensuring that $\Pr(g(a(D)) \leq 0) \geq 1 - \delta$. The algorithm we designed is compatible with g of the form $g(\theta) = \mathbf{E}[r'(H)|\theta_0] - \mathbf{E}[r'(H)|\theta]$, where the user selects $-r'(H)$ to measure a

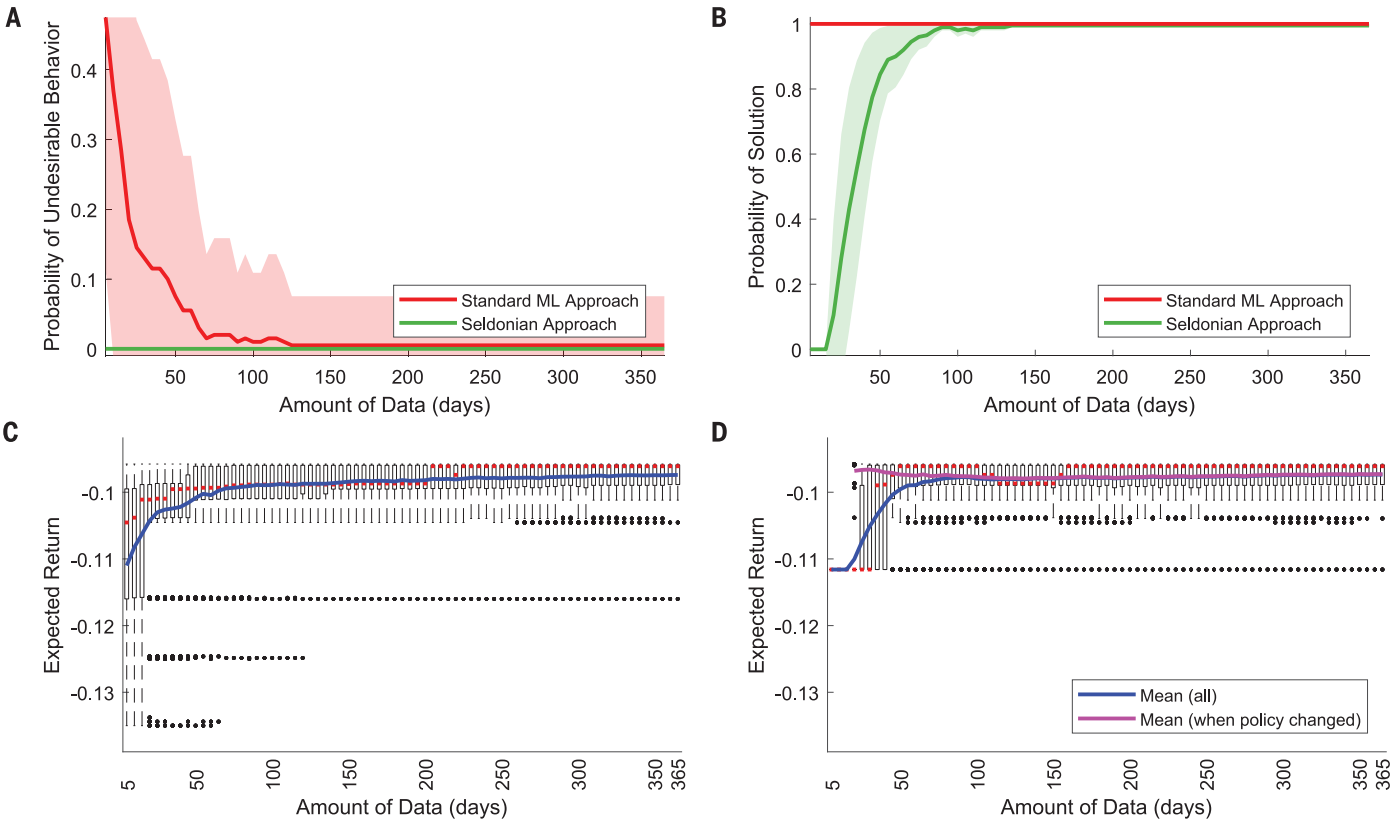


Fig. 4. Seldonian reinforcement learning algorithm for proof-of-principle bolus calculation in type 1 diabetes. Results are averaged over 200 trials; shaded regions denote SE. The Seldonian algorithm is compared to an algorithm built using the standard ML approach that penalizes the prevalence of low blood sugar. (**A**) Probability that each method returns policies (solutions) that increase the prevalence of low blood sugar. The algorithm designed using the standard ML approach often proposed policies that increased the prevalence of low blood sugar, violating the safety constraint, even though it used an objective function (reward function) that penalized instances of hypoglycemia. In contrast, across all trials, our Seldonian algorithm was safe; it never changed the treatment policy in a way that increased the prevalence of low blood sugar. (**B**) Probability that each method returns a policy that differs from the initial policy. Our Seldonian algorithm was able to safely improve upon the initial policy with just 1 to 5 months of data. (**C**) Box plot (with outliers plotted) of the distribution of the expected returns (objective function values) of the treatment policies returned by the standard ML algorithm. The blue line depicts the sample mean; red lines within the boxes mark the medians. All points below −0.1116 [where the blue curve in (D) begins] correspond to cases where the standard ML algorithm both decreased performance and produced undesirable behavior (an increase in the prevalence of low blood sugar). (**D**) Similar to (C), but showing results for the Seldonian algorithm. The magenta line is the average of the performance when the algorithm produced a policy that differed from the initial policy. Notice that all points have values of at least −0.1116, indicating that our algorithm never produced undesirable behavior. When boxes appear to be missing, the boxes have zero width and are obscured by the red line indicating the median of the box.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

particular definition of how undesirable the history H is. That is, with probability at least $1 - \delta$, the algorithm will not output a policy θ that increases the user-specified measure of undesirable behavior. Notice that the user need only be able to recognize undesirable behavior to define r'; the user does not need to know the distributions over histories H that result from applying different policies. For example, the user might define $r'(H) = -1$ if undesirable behavior occurred in H, and $r'(H) = 0$ otherwise.

Some previous reinforcement learning methods are guaranteed to increase the primary objective with high probability (*26*–*28*). These algorithms can be viewed as Seldonian or quasi-Seldonian algorithms that are restricted to only work with one definition of undesirable behavior: a decrease in the primary objective. This restricted definition of undesirable behavior precludes their application to problems where undesirable behavior does not align perfectly with the primary objective (see fig. S31 for an example where the behavioral constraint and primary objective are conflicting). Similarly, data-driven robust optimization (*29*) has also provided high-probability guarantees on constraint satisfaction, but only for convex constraints and a subset of objectives f that do not include the regression, classification, and reinforcement learning examples we consider (*14*).

Of the many high-risk, high-reward applications of reinforcement learning that have been proposed, we selected one to show the feasibility of our approach: automatically adjusting the treatment for a person with type 1 diabetes (*30*, *31*). In this application, a policy θ (as defined above) is a bolus calculator, which determines the amount of insulin that a person should inject prior to ingestion of a carbohydrate-containing meal to avoid high blood sugar levels. To simulate the metabolism of a human, we used a detailed metabolic simulator (*32*). Each history H corresponds to the outcome of 1 day, and we defined $-r'(H)$ to be a measure of the prevalence of low blood sugar (with particularly large penalties for hypoglycemia, i.e., dangerously low blood sugar levels) in the history H. Enforcing high-probability safety constraints on hypoglycemia is important because of the severe health consequences caused by hypoglycemia, including altered mental status, confusion, coma, and even death (*33*–*35*).

Figure 4 shows the result of applying both our Seldonian algorithm and a baseline algorithm designed using the standard ML approach. The baseline algorithm uses a technique called importance sampling (*36*) to estimate the performance of all policies using the data D generated by the initial policy θ_0, and it returns the policy predicted to perform best. This non-Seldonian algorithm (*14*) closely resembles our Seldonian algorithm with the behavioral constraints removed. Neither the Seldonian algorithm nor the corresponding standard ML approach algorithm are meant to be used directly in clinical practice; however, comparing their behavior provides insight into the effects of our Seldonian framework. Note from Fig. 4 that our algorithm does not propose a new policy until it has high confidence that the prevalence of low blood sugar will not increase. Our algorithm is not specific to this particular choice of constraint [see (*14*) for implementation of alternative constraints, such as constraints on the mean time hyperglycemic]. Our approach is complementary to existing work on personalized bolus calculators that do not use reinforcement learning but rely on experts or prior data to set critical parameters (*14*, *37*). These parameters could be adapted for each individual using a reinforcement learning approach, and a Seldonian reinforcement learning algorithm would ensure that it would alter the parameters only when it is highly confident that the change would not cause undesirable behavior (e.g., increase the prevalence of hypoglycemia) for the particular individual. Although any clinical application would leverage a more complicated policy than what we consider here, we use this as an illustration of how a Seldonian algorithm could be used as part of a broader effort to provide personalized policies for high-stakes applications.

Given the recent rise of real-world ML applications and the corresponding surge of potential harm that they could cause, it is imperative that ML algorithms provide their users with an effective means for controlling behavior. To this end, we have proposed a framework for designing ML algorithms and shown how it can be used to construct algorithms that provide their users with the ability to easily (that is, without requiring additional data analysis) place limits on the probability that the algorithm will produce any specified undesirable behavior. Algorithms designed using our framework are not just a replacement for ML algorithms in existing applications; it is our hope that they will pave the way for new applications for which the use of ML was previously deemed to be too risky.

REFERENCES AND NOTES

1. R. W. Jibson, *Eng. Geol.* **91**, 209–218 (2007).
2. M. Bhasin, G. Raghava, *Vaccine* **22**, 3195–3204 (2004).
3. J. Angwin, J. Larson, S. Mattu, L. Kirchner, Machine bias. *ProPublica*, May 2016; www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
4. D. A. Pomerleau, *Adv. Neural Inform. Process. Syst.* **1**, 305–313 (1988).
5. S. Saria, *IEEE Intell. Syst.* **29**, 82–87 (2014).
6. N. Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford Univ. Press, 2014).
7. S. Russell, *Sci. Am.* **314**, 58–59 (June 2016).
8. D. Amodei *et al.*, arXiv 1606.06565 [cs.AI] (25 July 2016).
9. S. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge Univ. Press, 2004).
10. D. Bertsimas, G. J. Lauprete, A. Samarov, *J. Econ. Dyn. Control* **28**, 1353–1381 (2004).
11. A. Charnes, W. W. Cooper, *Manage. Sci.* **6**, 73–79 (1959).
12. A. Ben-Tal, L. El Ghaoui, A. Nemirovski, *Robust Optimization* (Princeton Univ. Press, 2009).
13. I. Asimov, *Foundation* (Gnome, 1951).
14. See supplementary materials.
15. O. L. Mangasarian, W. N. Street, W. H. Wolberg, *Oper. Res.* **43**, 570–577 (1995).
16. L. Weber, "Your résumé vs. oblivion." *Wall Street Journal* (2012); www.wsj.com/articles/SB10001424052970204624204577178941034941330.
17. L. Li, W. Chu, J. Langford, R. E. Schapire, A contextual-bandit approach to personalized news article recommendation. In *International World Wide Web Conference* (2010), pp. 661–670.
18. R. M. Houtman *et al.*, *Int. J. Wildland Fire* **22**, 871–882 (2013).
19. B. Moore, P. Panousis, V. Kulkarni, L. Pyeatt, A. Doufas, Reinforcement learning for closed-loop propofol anesthesia: A human volunteer study. In *Proceedings of the Twenty-Second Innovative Applications of Artificial Intelligence Conference* (2010), pp. 1807–1813; www.aaai.org/ocs/index.php/IAAI/IAAI10/paper/view/1572/2359.
20. K. Grabczewski, W. Duch, Heterogeneous forests of decision trees. In *International Conference on Artificial Neural Networks* (2002), pp. 504–509.
21. D. Dheeru, E. Karra Taniskidou, UCI Machine Learning Repository (2017); http://archive.ics.uci.edu/ml.
22. K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, D. I. Fotiadis, *Comput. Struct. Biotechnol. J.* **13**, 8–17 (2015).
23. J. Komiyama, A. Takeda, J. Honda, H. Shimao, *Proc. Mach. Learn. Res.* **80**, 2737–2746 (2018).
24. A. Agarwal, A. Beygelzimer, M. Dudík, J. Langford, H. Wallach, *Proc. Mach. Learn. Res.* **80**, 60–69 (2018).
25. M. B. Zafar, I. Valera, M. G. Rodriguez, K. P. Gummadi, *Proc. Mach. Learn. Res.* **54**, 962–970 (2017).
26. P. S. Thomas, G. Theocharous, M. Ghavamzadeh, *Proc. Mach. Learn. Res.* **37**, 2380–2388 (2015).
27. M. Ghavamzadeh, M. Petrik, Y. Chow, *Adv. Neural Inform. Process. Syst.* **29**, 2298–2306 (2016).
28. R. Laroche, P. Trichelair, R. T. des Combes, *Proc. Mach. Learn. Res.* **97**, 3652–3661 (2019).
29. D. Bertsimas, V. Gupta, N. Kallus, *Math. Program.* **167**, 235–292 (2018).
30. M. Bastani, thesis, University of Alberta (2014).
31. S. Schmidt, K. Nørgaard, *J. Diabetes Sci. Technol.* **8**, 1035–1041 (2014).
32. C. Dalla Man *et al.*, *J. Diabetes Sci. Technol.* **8**, 26–34 (2014).
33. S. W. Suh, E. T. Gum, A. M. Hamby, P. H. Chan, R. A. Swanson, *J. Clin. Invest.* **117**, 910–918 (2007).
34. A. J. Bree, E. C. Puente, D. Daphna-Iken, S. J. Fisher, *Am. J. Physiol. Endocrinol. Metab.* **297**, E194–E201 (2009).
35. E. C. McNay, V. E. Cotero, *Physiol. Behav.* **100**, 234–238 (2010).
36. D. Precup, R. S. Sutton, S. Dasgupta, Off-policy temporal-difference learning with function approximation. In *Proceedings of the 18th International Conference on Machine Learning* (2001), pp. 417–424; https://dl.acm.org/citation.cfm?id=655817.
37. H. Zisser, L. Jovanovic, F. Doyle III, P. Ospina, C. Owens, *Diabetes Technol. Ther.* **7**, 48–57 (2005).
38. Data related to this publication are available through Harvard Dataverse. DOI: 10.7910/DVN/O35FW8
39. Source code for all experiments is available through Zenodo. DOI: 10.5281/zenodo.3490615
40. T. M. Mitchell, *Machine Learning* (McGraw-Hill, 1997).
41. D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Nature* **323**, 533–536 (1986).
42. A. Liaw, M. Wiener, *R News* **2**, 18–22 (2002).

ACKNOWLEDGMENTS

We thank G. Theocharous and M. Ghavamzadeh for their guidance in the development of the high-confidence policy improvement algorithms that initiated this line of research, and multiple colleagues and reviewers who provided valuable feedback. **Funding:** Supported by a gift from Adobe, NSF

Downloaded from http://science.sciencemag.org/ on November 28, 2019

CAREER awards 1350984 (E.B.) and 1453474 (Y.B.), NSF grant 1763423, and Institute of Educational Science grant R305A130215. The opinions expressed are those of the authors and do not represent views of Adobe, NSF, the Institute, or the U.S. Department of Education. **Author contributions:** P.S.T. conceived the idea with A.G.B. providing early guidance. P.S.T. and E.B. developed the framework formalization. P.S.T., B.C.d.S., S.G., Y.B., and E.B. designed and executed the experiments and analyzed the data, with P.S.T. and B.C.d.S. focusing on the regression experiments, S.G. and Y.B. focusing on the classification experiments, and P.S.T. and E.B. focusing on the reinforcement learning experiments. P.S.T., B.C.d.S., A.G.B., S.G., Y.B., and E.B. wrote and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data discussed in the main text and supplementary materials, as well as source code for reproducing all experiments, are available online (*38*, *39*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/999/suppl/DC1
Supplementary Text
Figs. S1 to S39
References (*43*–*214*)

11 November 2016; resubmitted 31 August 2017
Accepted 25 October 2019
10.1126/science.aag3311

Downloaded from http://science.sciencemag.org/ on November 28, 2019

TRAIT DEVELOPMENT

Beetle horns evolved from wing serial homologs

Yonggang Hu*, David M. Linz, Armin P. Moczek*

Understanding how novel complex traits originate is a foundational challenge in evolutionary biology. We investigated the origin of prothoracic horns in scarabaeine beetles, one of the most pronounced examples of secondary sexual traits in the animal kingdom. We show that prothoracic horns derive from bilateral source tissues; that diverse wing genes are functionally required for instructing this process; and that, in the absence of Hox input, prothoracic horn primordia transform to contribute to ectopic wings. Once induced, however, the transcriptional profile of prothoracic horns diverges markedly from that of wings and other wing serial homologs. Our results substantiate the serial homology between prothoracic horns and insects wings and suggest that other insect innovations may derive similarly from wing serial homologs and the concomitant establishment of structure-specific transcriptional landscapes.

How novel complex traits originate is a fundamental yet largely unresolved question in evolutionary biology (*1*, *2*). The most commonly used definition of novelty entails the absence of homology to ancestral traits (*3*). This definition, however, is increasingly difficult to reconcile with empirical findings across diverse taxa, which emphasize the differential repurposing of conserved developmental modules outside their traditional developmental context as a dominant route to innovation (*4*). How, and to what degree, evolutionary novelties may emerge from the confines of homology thus remains largely unknown. In this work, we investigated the origin of the prothoracic horns of scarabaeine beetles, a classic example of evolutionary innovation, from homologs of tissues and associated gene networks that instruct the formation of insect wings in adjacent segments.

The origin of insect wings has fueled a century-long debate. Insect wings were postulated to have arisen either as extensions of the dorsal plate (tergum, terga) of thoracic segments or, alternatively, from ancient proximal leg segments and their associated branches—structures in existence before the origin of insects and since absorbed into the side wall (pleuron, pleura) of segments (*5*). These competing hypotheses became united in the dual origin hypothesis, which posits that bona fide wings (on the second and third thoracic segments, T2 and T3) are composite structures with contributions from both tergal and pleural sources (*6*–*15*). Work in *Tribolium* and *Tenebrio* beetles has further shown that segmentally reiterated (i.e., serially homologous) tergal and pleural source tissues can also be found as distinct and morphologically diversified structures in nonwinged segments; for example, the tergal serial homolog facilitated the formation of a bilateral, edgelike structure along the prothoracic segment (T1), known as the carinated margin, as well as pupal-specific defensive structures found in abdominal segments, known as gin traps (*7*, *8*, *16*, *17*). These findings support the notion that the presence of two distinct sets of wing serial homologs per segment reflects the ancestral condition of thoracic and abdominal segments and that at least some lineages succeeded in using wing serial homologs outside T2 and T3 to evolve structures other than wings.

To determine whether prothoracic horns of scarabaeine beetles derive from tergal wing serial homologs, we first assessed the function of wing patterning genes during prothoracic horn formation. Specifically, we assessed the wing selector gene *vestigial* (*vg*); two genes critical to the early patterning of wing formation, *apterous* (*ap*) and *nubbin* (*nub*); as well as *cubitus interruptus* (*ci*) and *dishevelled* (*dsh*), key members of the hedgehog and wingless signaling pathways, respectively. Further, we examined the function of *abrupt* (*ab*), which, at least in *Drosophila* and *Tribolium*, is similarly critical for bona fide wing formation (*18*). We executed our approach in three *Onthophagus* species (*O. sagittarius*, *O. taurus*, and *O. binodis*), which reflect much of the diversity of prothoracic horn formation found in scarabaeine beetles (fig. S2) (*19*).

In all three beetle species examined, RNA interference (RNAi)–mediated knockdown of any of the six focal genes yielded the expected reduction of bona fide wings (on T2 and T3; see figs. S1 and S2) and—with only one exception (*nub*)—also profoundly affected the formation of prothoracic horns (Fig. 1, C to L, and figs. S3 to S5). Furthermore, we observed that hypomorphic down-regulation of *vg*, *dsh*, *ci*, *ab*, or *ap* (both paralogs simultaneously, hereafter *apA+B*), resulted in the retention of paired, bilateral vestiges of the prothoracic horn (Fig. 1 and figs. S3 and S4). Notably, bilateral prothoracic horn precursors can also be observed in wild-type individuals late in larval development (Fig. 1, N to P, and figs. S6 and S7). In addition to wings and prothoracic horns, vg^{RNAi} and, to a lesser extent, dsh^{RNAi}, $apA{+}B^{RNAi}$, ab^{RNAi}, or ci^{RNAi} also affected part of the pleural and tergal structures in T1, including the carinated margin and lateral body wall plates (Fig. 1 and fig. S5), consistent with studies in other Coleoptera (*7*, *17*).

To determine whether RNAi-mediated effects on prothoracic horn development may be an indirect consequence of wing-gene function elsewhere in the thorax, we examined the expression pattern of the wing selector gene, *vg*, using in situ hybridization chain reaction on cryosectioned thoracic tissue obtained from prepupae of *O. sagittarius* (Fig. 1M), i.e., at a stage when the larval epidermis transforms to give rise to pupal (and future adult) traits. We found that *vg* is expressed throughout developing wing tissue, carinated margin, and bilateral prothoracic horn precursors (Fig. 1, N to R, and fig. S7), thus closely paralleling the physical locations of vg^{RNAi} phenotypes.

Taken together, our results suggest (i) prothoracic horns derive from tergal bilateral source tissues that fuse to form a single medial outgrowth, (ii) wing genes are functionally required for instructing this process, and (iii) based on the specific identity of these genes, prothoracic horns may constitute partial wing serial homologs. Additional support for these conclusions can be found in previously published expression and functional analyses of four genes also known to be critical for wing formation (*patched*, *pangolin*, *homothorax*, and *decapentaplegic*) (*20*–*23*). In each case, RNAi-mediated depletion reduced or eliminated the formation of both wings and prothoracic horns as well as other wing serial homologs (*20*–*23*).

However, it remains conceivable that prothoracic horns do not derive from wing serial homologs and that, instead, the nine genes functionally implicated thus far were simply independently co-opted to instruct horn development. To begin assessing these alternative explanations, we executed two sets of experiments. First, we knocked down the Hox gene *Sex combs reduced* (*Scr*) to transform the identity of T1 to that of T2, which induces the formation of ectopic T1 wings. Specifically, we reasoned that if prothoracic horns evolved via the independent co-option of components of the wing gene network, rather than through modification of serially homologous tissues, the prothoracic horn primordia should not contribute to ectopic T1 wings. Alternatively, if horns are serially homologous to wings, ectopic wing induction should be paralleled by a commensurate reduction of horns, with the most severe T1 wing induction being paralleled by the most severe reduction of horn growth (*24*).

Department of Biology, Indiana University, Bloomington, IN 47405, USA.
*Corresponding author. Email: yohu@iu.edu (Y.H.); armin@indiana.edu (A.P.M.)

Downloaded from http://science.sciencemag.org/ on November 29, 2019

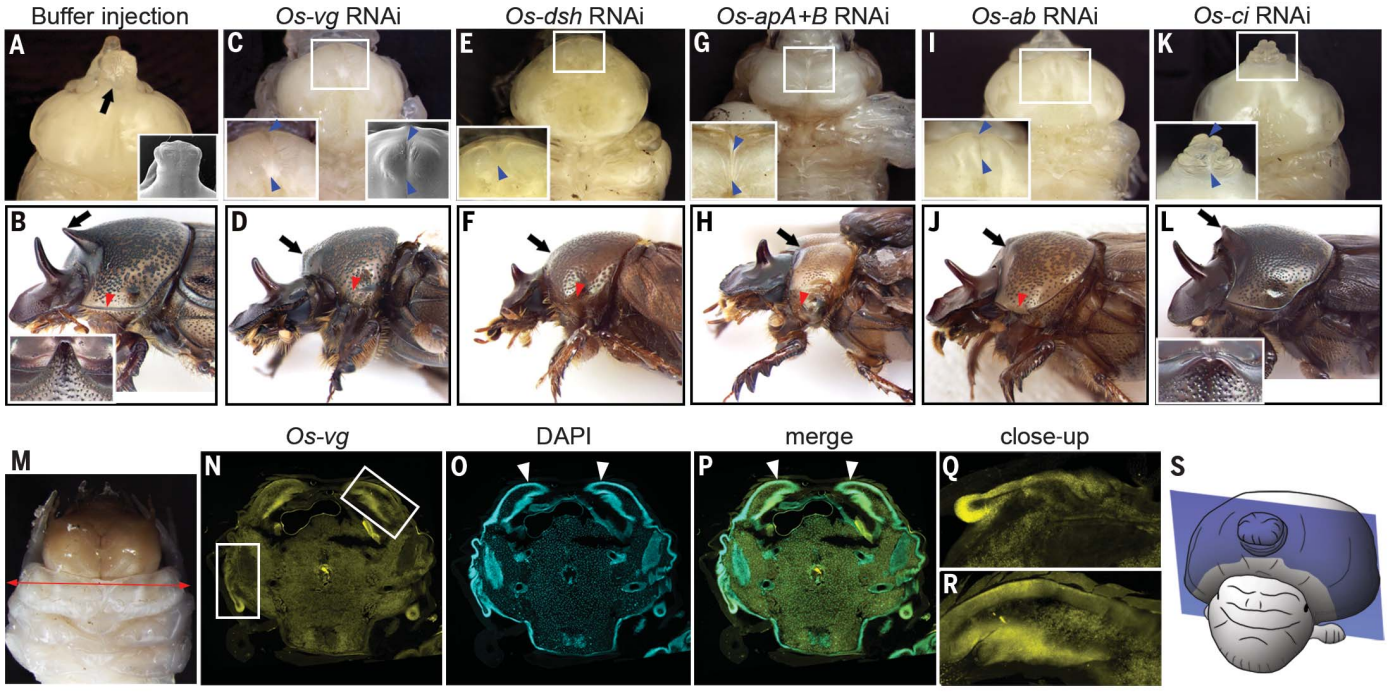


Fig. 1. Wing genes are required for prothoracic horn formation. (**A** and **B**) Buffer-injected control. The prothoracic horn [arrows in (A) and (B)] and carinated margin [red arrowhead in (B)] are indicated. Inset in (B) shows the prothoracic horn. Representative phenotypes obtained in *O. sagittarius* (*Os*) are shown in panels (C) to (L) as follows: (**C** and **D**) vg^{RNAi}; (**E** and **F**) dsh^{RNAi}; (**G** and **H**) $apA{+}B^{RNAi}$; (**I** and **J**) ab^{RNAi}; and (**K** and **L**) ci^{RNAi}. Wing gene RNAi reduces pupal prothoracic horns to paired bilateral vestiges [insets in (C), (E), (G), (I), and (K); the furrow between paired vestiges is indicated by blue arrowheads] and leads to a reduction or elimination of the adult prothoracic horns [arrows in (D), (F), (H), (J), and (L)] and defects in the carinated margin [red arrowheads in (D), (F), (H), and (J)]. Inset in (L) shows partially reduced prothoracic horn. Inset in (A) and right inset in (C) are scanning electron microscopy (SEM) images. Shown are representative phenotypes obtained in *O. sagittarius* (for *O. taurus* and *O. binodis* see fig. S3). (**M**) Prepupa illustrating the section plane in (N) through (R). (**N** to **R**) Cryosections stained with *vg* riboprobes (N) and 4′,6-diamidino-2-phenylindole (DAPI) (O) showing *vg*-positive cells in prothoracic horn primordia and carinated margin. (P) Merge of (N) and (O). Arrowheads in (O) and (P) indicate prothoracic horn primordia. (Q and R) Close-up of insets in (N) showing carinated margin and prothoracic horn primordia, respectively. (**S**) Cartoon of pupa illustrating section plane in (M) through (P).

Scr^{RNAi} resulted in the formation of ectopic T1 wings resembling elytra (Fig. 2, D to F), in line with results from studies in other taxa (*7*, *12*, *17*, *25*). In all three species, this induction was paralleled by a severe reduction of prothoracic horns to minute, and again paired, vestiges (Fig. 2D and figs. S8 and S9). Additionally, analysis of hypomorphic Scr^{RNAi} phenotypes showed the predicted inverse relationship between the degree to which prothoracic horn tissue was retained and the completeness of the ectopic wing induced (figs. S8 and S9).

However, such results may also be expected if *Scr* independently governs the T1-specific repression of wings and induction of thoracic horns. To assess this hypothesis, we investigated T1 wing induction while simultaneously ablating the prothoracic horn via down-regulating *pannier* (*pnr*), which patterns dorsomedial tissue identity in *Drosophila* (*26*). By itself, pnr^{RNAi} removes all dorsomedial projections, including the prothoracic horn, yet it does not affect the formation of wings on T2 and T3 nor the lateral carinated margin in T1 (Fig. 2, G to I). We reasoned that if prothoracic horn primordia contribute to ectopic T1 wings, ectopic wing size should be maximized in *Scr* single-knockdown individuals but diminished when Scr^{RNAi} is executed in a pnr^{RNAi} background that removes prothoracic horn primordia before they can be rerouted toward an ectopic wing fate. To test this hypothesis, we compared ectopic T1 wing formation in Scr^{RNAi} knockdown individuals with that of $pnr{+}Scr^{RNAi}$ double-knockdown individuals in *O. binodis*, which exhibits consistently strong RNAi phenotypes (fig. S13 and table S2). We further standardized our comparisons by controlling developmental timing of knockdowns (table S2) and by using the severity of Scr^{RNAi}-mediated mouthpart transformations and exposure of the coxal segment as a proxy for transformation severity (fig. S9). We found that ectopic wings formed in a pnr^{RNAi} background are markedly smaller in size and lack the recognizable dorsal surface traits observed in *Scr* single-knockdown individuals (Fig. 2, J to L, and figs. S9 and S10). These results provide strong support for the hypothesis that prothoracic horn primordia contribute to ectopic bilateral T1 wings.

Lastly, we sought to assess whether the presumed serial homology between prothoracic horns, wings, and other wing serial homologs, such as gin traps and their equivalents (*8*, *16*, *17*), is also reflected in unbiased, tissue-wide transcriptional profiles. We selected *O. taurus* for this approach, given the availability of a fully sequenced and annotated genome for this species. We also included in this analysis eight distinct epidermal regions, including T2 wings, four distinct T1 regions (fig. S12), and abdominal support structures from the third through sixth abdominal segment (PSS; partial wing serial homologs akin to gin traps or carinated margin, see fig. S11) (*8*, *16*, *17*, *27*). We included the dorsocentral prothorax and dorsal head epidermis, regions whose formation is, at present, considered completely unrelated to wing development. *vg* expression was detectable in all epidermal tissues examined except for the dorsal head and dorsocentral prothorax, consistent with the proposed partial serial homology between

Downloaded from http://science.sciencemag.org/ on November 29, 2019

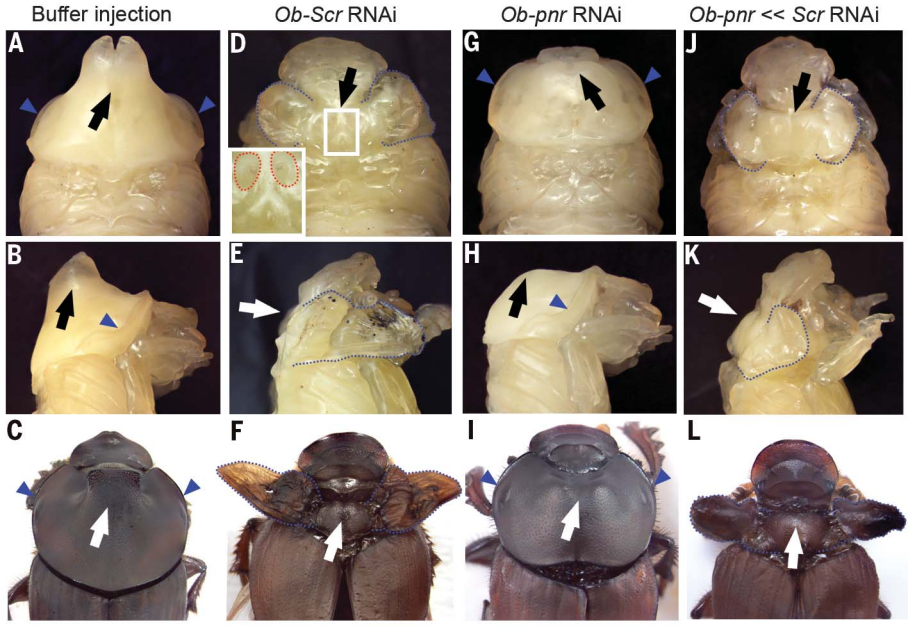


Fig. 2. The prothoracic horn contributes to ectopically induced prothoracic wings. (**A** to **C**) Buffer injected control. The prothoracic horn [arrows in (A) to (C)] and carinated margin [blue arrowheads in (A) to (C)] are indicated. (**D** to **F**) Pupal and adult phenotypes of Scr^{RNAi}. Scr^{RNAi} reduces pupal prothoracic horn [arrows in (D) to (F)] resulting in small, paired horn vestiges [inset in (D)] and also induces large ectopic prothoracic wings [blue outlines in (D) to (F)]. (**G** to **I**) pnr^{RNAi}. pnr^{RNAi} removes prothoracic horns [arrows in (G) to (I)] without affecting the carinated margin [blue arrowheads in (G) to (I)]. (**J** to **L**) pnr^{RNAi} followed by Scr^{RNAi}. Sequential knockdown of *pnr* followed by *Scr* removes prothoracic horns [arrows in (J) to (L)] and results in only partial induction of ectopic prothoracic wings [blue outlines in (J) to (L)]. Across all images, the prothoracic horn and carinated margin are indicated by arrows and blue arrowheads, respectively. The ectopic wings are outlined with dotted lines.

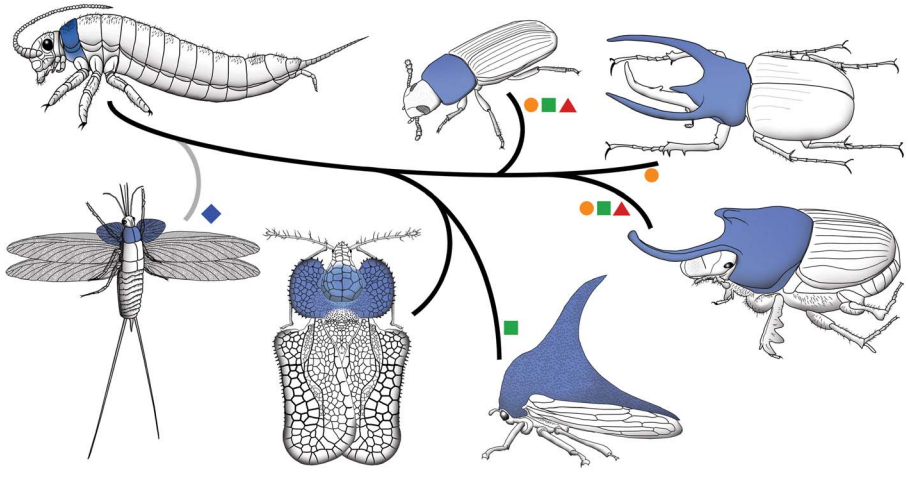


Fig. 3. Diversification of prothoracic morphology among various extant and extinct insect lineages. Shapes indicate lineages where paleontological evidence (blue diamond) (*11*), expression data (green squares) (*7*, *29*), functional data (orange circles) (*7*, *28*), or transformation (via Hox manipulation) data (red triangles) (*7*) are available to support the hypothesis of wing serial homology. Insect lineages shown are (clockwise from top left) a hypothetical apterygote insect ancestor, Tenebrionidae, Dynastinae, Scarabaeinae, Membracidae, Tingidae, and Palaeodictyoptera.

prothoracic horns, PSS, carinated margin, and wings (fig. S12C). Expanding this approach to include 41 genes known to be functionally required for some aspect of wing formation, clusters T2 wings together with the carinated margin and pupal support structures, as expected. However, this aproach reveals significant transcriptional divergence of prothoracic horn tissue from other wing serial homologs (fig. S12D). Lastly, transcriptome-wide clustering of all 4191 differentially expressed transcripts shows no distinct clustering patterns corresponding to wing relatedness among any of the tissues examined. Combining the results of our functional analysis detailed above with these unbiased RNA sequencing (RNA-seq) findings raises the possibility that prothoracic horn formation may indeed be instructed by a core gene network, serially homologous to that also used in bona fide wings. However, on the level of gene networks, such serial homology may be restricted to core regulators and the early stages of tissue specification.

Our results help advance the current debate surrounding the origin of morphological novelty in insects and in animal evolution more broadly. First, our findings may help explain why the prothorax of insects has emerged as a hotspot of evolutionary innovation. Morphological elaborations have evolved independently in diverse insect orders, including the wing-like elaborations of extinct Paleodictyoptera, the lateral, leaflike outgrowths of lace bugs (Tingidae, Hemiptera), the helmets of treehoppers (Membracidae, Hemiptera), or the posteriorly projecting outgrowths of tetrigid grasshoppers (Tetrigidae, Orthoptera). Our analyses of gene function, gene expression, and Hox gene–mediated transformation support the hypothesis that scarabaeine prothoracic horns evolved from bilateral, partial wing serial homologs, whereas our transcriptomic profiling documents the simultaneous establishment of structure-specific transcriptional landscapes. Taken together, our findings raise the possibility that other insect innovations may have similarly evolved by using wing serial homologs as developmental-genetic starting points around which transcriptional repertoires became established, and through these diversified networks are now able to support differentiation events specific to each trait. Paleontological data, gene expression, and functional analyses across diverse taxa are beginning to accumulate support for such a scenario (Fig. 3) (*7*, *11*, *28*, *29*).

More generally, our results contribute to a growing call for reexamining the usefulness of defining morphological novelty through the absence of homology. According to this view

Downloaded from http://science.sciencemag.org/ on November 29, 2019

novelty begins where homology ends, yet exactly where homologous relationships cease has become increasingly difficult to delineate, as findings in evolutionary developmental biology have forced a revision of homology away from a binary designation and toward a more layered understanding of homologous relationships, resulting in the emergence of concepts such as "deep" or "partial homology" (*4*). In contrast, others have advocated disconnecting novelty and homology entirely (*30*). According to this perspective, as evolutionary biology is fundamentally positioned within a framework of descent with modification, everything new must ultimately emerge from the old, and it may therefore be most productive to follow a dichotomy originally proposed by Wilkins (*31*) that divides evolutionary novelties into operational types—those whose evolutionary and developmental origins we can trace and those whose precursors and ancestral affinities we have yet to discover. Here we show that a textbook example of evolutionary novelty, the prothoracic horns of beetles, derives partly from wing serial homologs, whose existence predates the origin of insects. On one side, this may cause us to question whether prothoracic horns should still be considered an evolutionary novelty. Alternatively, our results may serve to illustrate how substantial morphological innovation, rather than somehow emerging in the absence of homology, may instead be initiated through it.

REFERENCES AND NOTES

1. R. A. Raff, *The Shape of Life: Genes, Development, and the Evolution of Animal Form* (Univ. of Chicago Press, 1996).
2. G. P. Wagner, *Homology, Genes, and Evolutionary Innovation* (Princeton Univ. Press, 2014).
3. G. B. Müller, G. P. Wagner, *Annu. Rev. Ecol. Syst.* **22**, 229–256 (1991).
4. N. Shubin, C. Tabin, S. Carroll, *Nature* **457**, 818–823 (2009).
5. C. M. Clark-Hachtel, Y. Tomoyasu, *Curr. Opin. Insect Sci.* **13**, 77–85 (2016).
6. N. Niwa *et al*., *Evol. Dev.* **12**, 168–176 (2010).
7. C. M. Clark-Hachtel, D. M. Linz, Y. Tomoyasu, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16951–16956 (2013).
8. D. M. Linz, Y. Tomoyasu, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E658–E667 (2018).
9. A. P. Rasnitsyn, *J. Morphol.* **168**, 331–338 (1981).
10. G. Crampton, *J. N.Y. Entomol. Soc.* **24**, 1–39 (1916).
11. J. Prokop *et al*., *Curr. Biol.* **27**, 263–269 (2017).
12. V. Medved *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15946–15951 (2015).
13. C. M. Clark-Hachtel, M. R. Moe, Y. Tomoyasu, *Arthropod Struct. Dev.* **47**, 352–361 (2018).
14. M. Elias-Neto, X. Belles, *R. Soc. Open Sci.* **3**, 160347 (2016).
15. D. Requena *et al*., *Curr. Biol.* **27**, 3826–3836.e5 (2017).
16. Y. Hu *et al*., *Proc. Biol. Sci.* **285**, 20181373 (2018).
17. T. Ohde, T. Yaginuma, T. Niimi, *Science* **340**, 495–498 (2013).
18. P. Ravisankar, Y.-T. Lai, N. Sambrani, Y. Tomoyasu, *Dev. Biol.* **409**, 518–529 (2016).
19. A. P. Moczek, *Am. Nat.* **168**, 711–729 (2006).
20. B. R. Wasik, A. P. Moczek, *Dev. Genes Evol.* **221**, 17–27 (2011).
21. B. R. Wasik, A. P. Moczek, *Genesis* **50**, 404–414 (2012).
22. A. P. Moczek, D. J. Rose, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8992–8997 (2009).
23. T. Kijimoto, A. P. Moczek, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5982–5987 (2016).
24. Y. Tomoyasu, T. Ohde, C. Clark-Hachtel, *F1000Res.* **6**, 268 (2017).
25. Y. Tomoyasu, S. R. Wheeler, R. E. Denell, *Nature* **433**, 643–647 (2005).
26. M. Calleja *et al*., *Development* **127**, 3971–3980 (2000).
27. A. P. Moczek, T. E. Cruickshank, A. Shelby, *Evolution* **60**, 2329–2341 (2006).
28. T. Ohde *et al*., *PLOS Genet.* **14**, e1007651 (2018).
29. B. Prud'homme *et al*., *Nature* **473**, 83–86 (2011).
30. D. B. Wake, in *Keywords and Concepts in Evolutionary and Developmental Biology*, B. K. Hall, W. M. Olson, Eds. (Harvard Univ. Press, 2003), pp. 191–200.
31. A. S. Wilkins, *The Evolution of Developmental Pathways* (Sinauer Associates, 2002).

ACKNOWLEDGMENTS

We thank Y. Tomoyasu for comments on the manuscript, B. Stein for help with SEM, E. Ragsdale for providing access to the Zeiss imager Z2 microscope, and J. Liu (Indiana University Center for Genomics and Bioinformatics) for help with RNA-seq. **Funding:** This work was supported by National Science Foundation grants (IOS 1256689 and IOS 1901680) and a grant from the John Templeton Foundation to A.P.M. **Author contributions:** Y.H., D.M.L., and A.P.M. conceived the project and designed the experiments; Y.H. and D.M.L. performed the experiments; and Y.H., D.M.L., and A.P.M. analyzed the data and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The sequences of cloned gene fragments are available at GenBank (NCBI) with accession numbers MK249376 to MK249384 and MN331527 to MN331533. The datasets of RNA-seq are available at Gene Expression Omnibus (NCBI) with accession number GSE137455.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/1004/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 to S3
References (*32*–*43*)

6 December 2018; resubmitted 16 September 2019
Accepted 15 October 2019
10.1126/science.aaw2980

Downloaded from http://science.sciencemag.org/ on November 29, 2019

NEUROSCIENCE

A cortical-brainstem circuit predicts and governs compulsive alcohol drinking

Cody A. Siciliano[1,2,3]*, Habiba Noamany[1], Chia-Jung Chang[1], Alex R. Brown[1,2,3], Xinhong Chen[1], Daniel Leible[1], Jennifer J. Lee[1], Joyce Wang[1], Amanda N. Vernon[1], Caitlin M. Vander Weele[1], Eyal Y. Kimchi[1], Myriam Heiman[1], Kay M. Tye[1,4]*

What individual differences in neural activity predict the future escalation of alcohol drinking from casual to compulsive? The neurobiological mechanisms that gate the transition from moderate to compulsive drinking remain poorly understood. We longitudinally tracked the development of compulsive drinking across a binge-drinking experience in male mice. Binge drinking unmasked individual differences, revealing latent traits in alcohol consumption and compulsive drinking despite equal prior exposure to alcohol. Distinct neural activity signatures of cortical neurons projecting to the brainstem before binge drinking predicted the ultimate emergence of compulsivity. Mimicry of activity patterns that predicted drinking phenotypes was sufficient to bidirectionally modulate drinking. Our results provide a mechanistic explanation for individual variance in vulnerability to compulsive alcohol drinking.

More than 80% of adults are exposed to alcohol during their lifetime (*1*), yet less than 30% will develop an alcohol use disorder (AUD) (*2*). How exposure to alcohol can produce such disparate outcomes between individuals remains poorly understood.

Compulsive alcohol drinking, defined as continued drinking in the face of a negative consequence (*3*, *4*), is a distinguishing feature of AUDs (*5*). The medial prefrontal cortex (mPFC) is critical in mediating pathological drug-seeking behaviors, including compulsion (*6*–*10*). Both preexisting (*11*–*13*) and alcohol-induced changes in PFC function can contribute to maladaptive behaviors including compulsive drinking (*14*–*17*). Although rodent models have advanced our understanding of drinking behavior (*14*, *18*–*21*), neural correlates are typically assessed at a single end point, after long-term exposure to alcohol, thereby occluding individual differences in compulsion vulnerability as well as the longitudinal nature of its development.

We developed a "binge-induced compulsion task" (BICT) to assess how predisposition interacts with experience to produce compulsive drinking (Fig. 1A). Initially, an auditory conditioned stimulus (CS+) predicted delivery of sucrose until animals reliably responded (see the materials and methods). During pre-binge (days 1 to 3), the CS+ predicted delivery of alcohol alone (15%). On days 4 to 5, increasing concentrations of quinine, a bitter tastant used as a punishment (*3*), were added to the alcohol (alcohol+quinine). During binge drinking (days 6 to 19), animals had unlimited access to water and alcohol for 0, 2, or 4 hours

[1]The Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Pharmacology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA. [3]Vanderbilt Center for Addiction Research, Vanderbilt University School of Medicine, Nashville, TN 37232, USA. [4]The Salk Institute for Biological Studies, La Jolla, CA 92037, USA.
*Corresponding author. Email: tye@salk.edu (K.M.T.); cody.siciliano@vanderbilt.edu (C.A.S.)

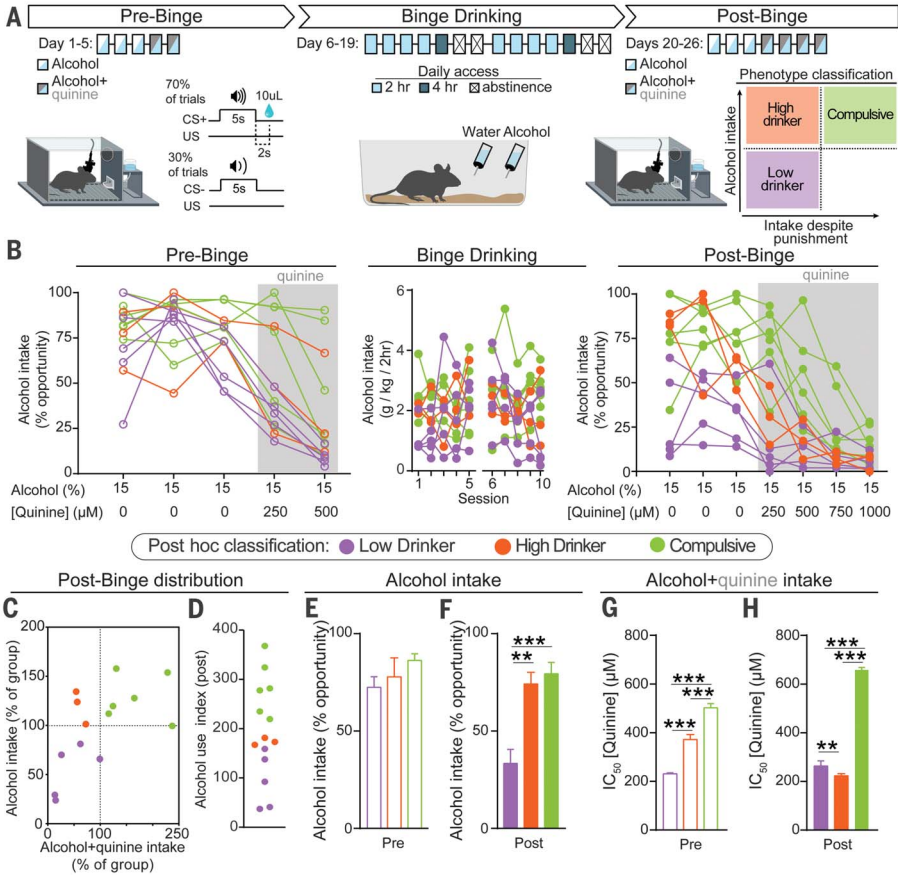


Fig. 1. Binge-induced compulsion task (BICT) for tracking the emergence of individual differences in compulsive alcohol drinking. (**A**) Schematic of the BICT. (**B**) Individual animals' alcohol consumption. (**C**) Normalized distributions of alcohol and alcohol+quinine consumption during the post-binge conditioning phase plotted to classify each animal's alcohol-drinking phenotype, which was applied post hoc to the dataset. (**D**) Alcohol use index. (**E**) Average performance across the pre-binge alcohol-only sessions (days 1 to 3) did not differ between groups [one-way analysis of variance (ANOVA), $F_{(2,11)} = 1.922$, $p = 0.19$]. (**F**) During post-binge conditioning sessions, high drinkers and compulsive animals consumed more alcohol (one-way ANOVA, $F_{(2,11)} = 15.41$, $p = 0.0006$). (**G**) Concentration required to produce an IC_{50} calculated from pre-binge conditioning data (one-way ANOVA, $F_{(2,11)} = 430.3$, $p < 0.0001$). (**H**) After binge drinking, compulsive animals exhibited robust punishment-resistant alcohol intake compared with both low and high drinkers (one-way ANOVA, $F_{(2,11)} = 1298.0$, $p < 0.0001$). All post hoc comparisons used Tukey's test: $^{**}p < 0.01$; $^{***}p < 0.001$. Error bars indicate ± SEM.

Downloaded from http://science.sciencemag.org/ on November 29, 2019

per day, producing high, "binge-like" levels of alcohol intake (*22*). During post-binge (days 20 to 26), animals returned to the pre-binge conditioning context, in which alcohol alone was presented for 3 days, followed by alcohol +quinine for the next 4 days (Fig. 1A). Intake volumes correlated with blood alcohol content (fig. S1).

The BICT allows for longitudinal assessment of two behavioral outcomes associated with diagnostic criteria for AUDs (*5*): alcohol consumption and continued consumption despite negative outcomes. After binge drinking, there were wide individual differences in drinking, both in the absence and presence of quinine (Fig. 1B). Three phenotypic classifications were made based on post-binge behavior: mice that displayed low alcohol intake with and without punishment were termed "low drinkers"; mice that showed high levels of alcohol drinking but were sensitive to punishment were termed "high drinkers"; mice with high levels of drinking that persisted with punishment were termed "compulsive" (Fig. 1C). Values from the alcohol-only and alcohol +quinine distributions were summed to create an "alcohol use index" for each animal (Fig. 1D).

Each animal was classified on the basis of its behavior during post-binge and designation was retroactively applied. Mice that were eventually divided into the three subgroups showed no detectable differences in unadulterated alcohol intake before binge drinking (Fig. 1E). After binge drinking, low drinkers' intake decreased, even in the absence of punishment (Fig. 1F and fig. S2D). Before binge drinking, compulsive animals showed greater resistance to punishment than both low and high drinkers, as measured by the concentration of quinine required to produce a half-maximal effect on alcohol consumption [half-maximal inhibitory concentration (IC_{50}); Fig. 1G and fig. S2E]. After binge drinking, this phenotype was exacerbated as compulsive animals showed a robust insensitivity to punishment (Fig. 1H and fig. S2F). Longitudinal examination highlighted a substantial divergence among groups when punishment was present: high drinkers showed increased sensitivity to quinine's effects on alcohol intake after binge drinking, whereas compulsive animals showed decreased sensitivity (fig. S2G). There were no group differences in alcohol consumption during binge drinking (fig. S2H).

To determine whether phenotypic differences in drinking reflected responses to punishment in general and were not driven by quinine-specific effects, we punished alcohol consumption with foot shock (fig. S3A). Phenotypic drinking behavior was

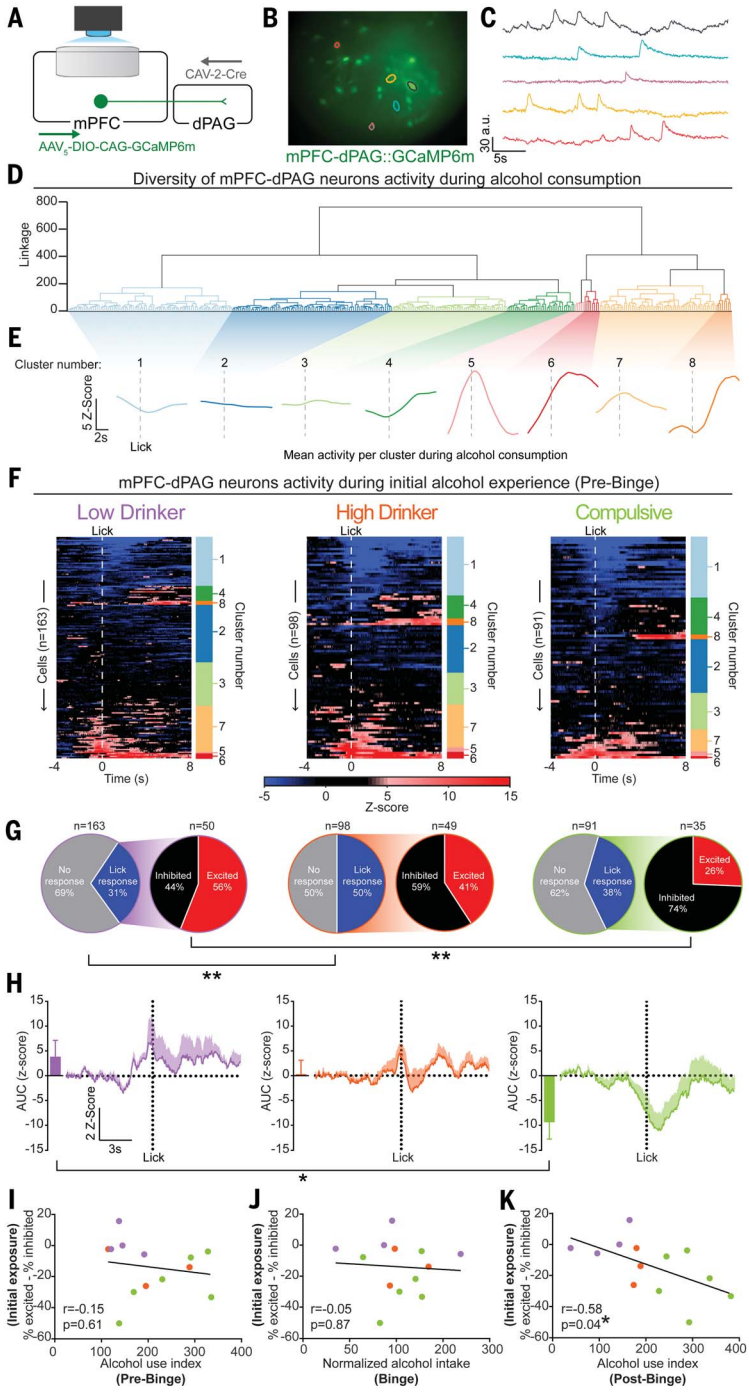


Fig. 2. Activity in mPFC-dPAG neurons during initial experience with alcohol is a vulnerability marker for future alcohol abuse-like behaviors. (**A**) Monitoring mPFC-dPAG activity using in vivo calcium imaging. (**B** and **C**) Field of view (B) and activity traces (C) from example cells. (**D**) Agglomerative hierarchical clustering of calcium activity traces during the first session of pre-binge (n = 13 animals, 352 cells). (**E**) Smoothed and averaged peristimulus time histograms per cluster. (**F**) Cluster designations are to the right of each neuron's heatmap of z-scored trial-averaged activity. (**G**) Differences in distributions of activity during alcohol consumption (Fisher's exact test: $^{**}p < 0.01$). (**H**) Population activity from lick-responsive neurons. Inset: Area under the curve (AUC) for each trace (one-way ANOVA, $F_{(2,10)} = 4.531$, $^{*}p = 0.039$; Tukey's post hoc test, $^{*}p < 0.05$). (**I** to **K**) Balance of excitatory–inhibitory activity during alcohol consumption in the first pre-binge session plotted against each animal's alcohol use index. No correlation between the excitation–inhibition balance during initial exposure and alcohol use index from pre-binge data (I) or from binge drinking (J). (K) Increased inhibitory activity during alcohol consumption during initial exposure predicted heightened pathological-like drinking behaviors during post-binge. Error bars indicate ± SEM.

Downloaded from http://science.sciencemag.org/ on November 29, 2019

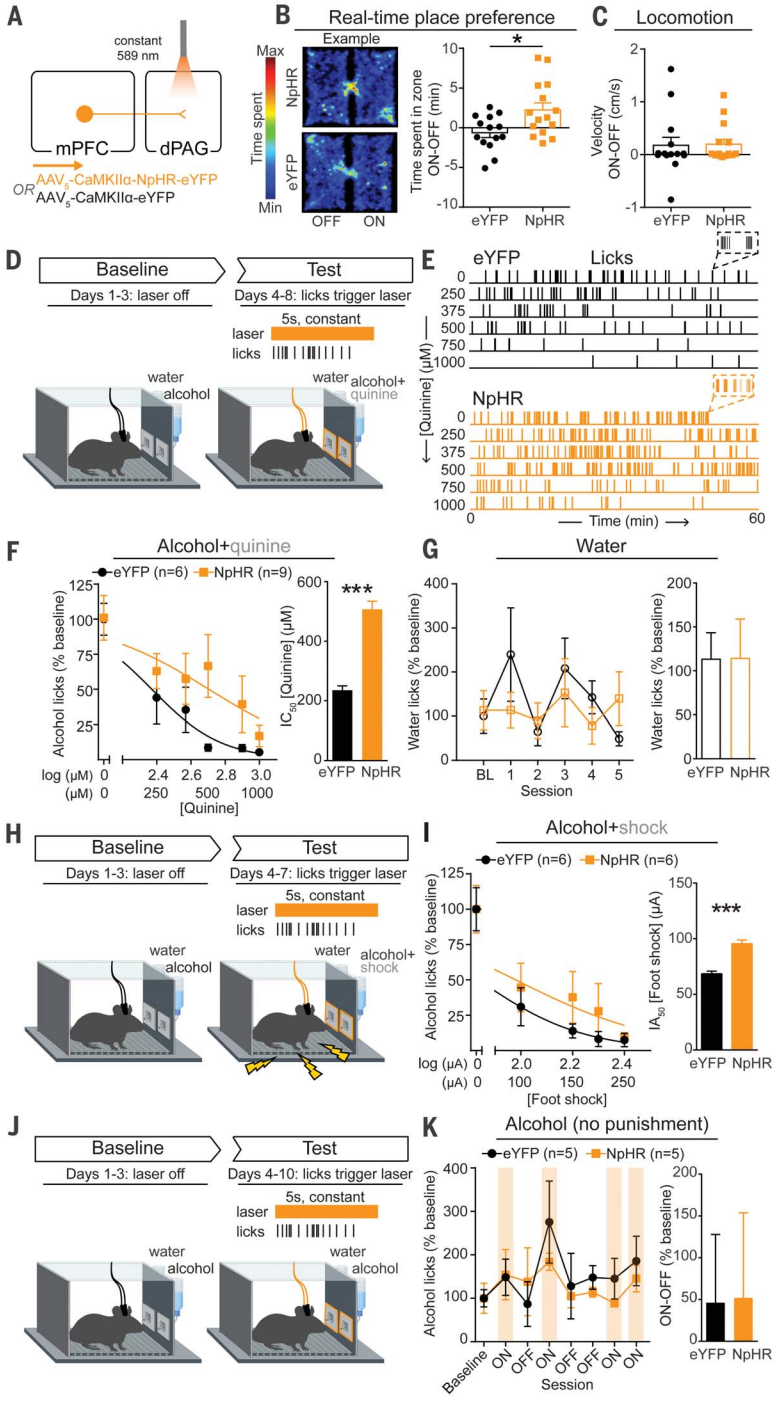


Fig. 3. Inhibition of mPFC-dPAG neurons drives compulsive drinking but does not alter drinking in the absence of punishment. (**A**) Strategy to inhibit mPFC terminals in the dPAG. (**B**) Inhibition of mPFC terminals in the dPAG was preferred in a real-time place preference task (unpaired *t* test, $t_{(27)}$ = 2.647, **p* = 0.013). (**C**) Photoinhibition did not alter locomotion (unpaired *t* test, $t_{(27)}$ = 0.1191, *p* = 0.91). (**D**) On test days, water or alcohol spout contacts triggered a photoinhibition period. During the test, the quinine concentration was increased across days (alcohol bottle only). (**E**) Example alcohol lick event records. (**F**) The concentration of quinine required to decrease alcohol spout licking to 50% of baseline (IC_{50}) was greater in NpHR animals (unpaired *t* test, $t_{(13)}$ =22.05, ****p* < 0.0001). (**G**) No difference in licking for water between groups (unpaired *t* test, $t_{(13)}$ = 0.016, *p* = 0.99). (**H**) Alcohol drinking punished with foot shock. (**I**) Foot-shock amplitude required to attenuate alcohol spout licks by 50% of baseline [half-maximal inhibitory amplitude (IA_{50})] was increased in NpHR animals (unpaired *t* test, $t_{(10)}$ = 6.498, ****p* < 0.0001). (**J**) Alcohol drinking in the absence of punishment. (**K**) Photoinhibition did not alter licking for alcohol in the absence of punishment (unpaired *t* test, $t_{(8)}$ = 0.045, *p* = 0.97). Error bars indicate ± SEM.

retained, demonstrating that these behavioral traits are reproducible and generalizable (fig S3, B and C).

We reasoned that mPFC circuits involved in "top-down" control of avoidance behavior may underlie susceptibility to developing compulsive drinking behaviors. The periaqueductal gray (PAG) is involved in responding to aversive events (*23*–*27*), as well as negative affective states and hyperalgesia during alcohol withdrawal (*28*, *29*). mPFC neurons projecting to dorsal PAG (mPFC-dPAG) encode aversive events (*24*). We hypothesized that functional deficits in mPFC-dPAG neurons could disrupt aversive processing to drive compulsive drinking.

We used cellular-resolution calcium imaging (*30*) to visualize the activity of mPFC-dPAG neurons during the BICT (Fig. 2). An anterogradely traveling virus allowing for cre-dependent expression of GCaMP6m was injected in the mPFC and a retrogradely traveling virus carrying cre-recombinase was injected into the dPAG (Fig. 2A). A gradient-refractive index lens (fig. S4) and a head-mounted microendoscope allowed observation of calcium dynamics. An example field of view illustrates neurons imaged during the BICT (Fig. 2B) and extracted activity traces (Fig. 2C). Hierarchical clustering performed on activity from 352 neurons aligned around initiation of alcohol consumption during the first pre-binge session revealed eight distinct clusters (Fig. 2, D and E).

Although there was no difference between groups in alcohol intake during the "alcohol-only" sessions throughout pre-binge, dynamics of mPFC-dPAG neurons during alcohol consumption differed between the phenotypic groups (Fig. 2F). During the initial alcohol experience (day 1), more mPFC-dPAG neurons exhibited inhibitory responses for compulsive animals than for low drinkers (Fig. 2G). mPFC-dPAG neurons displayed more excitatory activity in low drinkers than in compulsive animals during alcohol consumption (Fig. 2H). A small proportion of mPFC-dPAG neurons displayed responses to the alcohol-predictive cue (CS+) (fig. S5).

Although there were no detectable differences among groups in *behavioral* performance during initial alcohol exposure (Fig. 1, B and E), the *neural* response during initial exposure predicted the future development of compulsive drinking (Fig. 2, F to H). The proportion of excitatory to inhibitory responses of individual mPFC-dPAG neurons for each animal did not correlate with behavior during pre-binge (Fig. 2I) or binge drinking (Fig. 2J), but *did* correlate with post-binge behavior >2 weeks after the neural recordings during initial exposure were collected (Fig. 2K). A support-vector machine

Downloaded from http://science.sciencemag.org/ on November 29, 2019

decoded future behavioral selection of drinking (go) versus not drinking (no go) based on the activity of mPFC-dPAG neurons during consumption of alcohol on the previous trial (fig. S6). This supports the notion that this circuit plays a key role in triggering the transition from moderate to compulsive drinking.

To test whether mimicking endogenous activity could alter behavior, we bilaterally expressed halorhodopsin (NpHR) in mPFC neurons and implanted bilateral optic fibers over the dPAG (Fig. 3A and fig. S7). In a real-time place preference assay, NpHR mice displayed modest preference for the photoinhibition-paired side of the chamber compared with fluorophore [enhanced yellow fluorescent protein (eYFP)]-expressing control mice (Fig. 3B). Photoinhibition did not produce any detectable changes in locomotion (Fig. 3C) or anxiety-related behavior (fig. S8, A to F).

Animals were given concurrent access to alcohol and water for three sessions to establish a baseline level of alcohol intake (Fig. 3D). On day 4, quinine was added to the alcohol bottle only, and the quinine concentration was increased across sessions to assess alcohol intake in the face of punishment (Fig. 3E). During quinine sessions, contacts on water or alcohol lickometers triggered photoinhibition, intended to mimic the inhibitory activity observed in compulsive animals during alcohol licking (Fig. 2H). Photoinhibition concomitant with licking for the alcohol+quinine solution was sufficient to induce a rightward shift in the quinine concentration response curve, resulting in a greater than twofold increase in the IC_{50} of quinine compared with eYFP controls (Fig. 3F) without affecting water consumption (Fig. 3G). When alcohol spout contacts were punished with foot shock (Fig. 3H), photoinhibition again promoted compulsive drinking (Fig. 3I and fig. S9, A and B).

To determine whether photoinhibition of mPFC-dPAG activity drives compulsive drinking by increased reinforcing effects of alcohol or decreased sensitivity to punishment, we measured each component in isolation. Photoinhibition did not change alcohol consumption in the absence of punishment (Fig. 3, J and K, and fig. S9, C and D). To determine whether photoinhibition altered responses to noxious stimuli in the absence of reward, we photoinhibited mPFC-dPAG synapses while animals' tails were immersed in 50°C water and found that photoinhibition slowed latency to withdraw (fig. S8, J to L). Photoinhibition did not support intracranial self-stimulation (fig. S8, G to I) and did not alter extinction of operant alcohol self-administration (fig. S10).

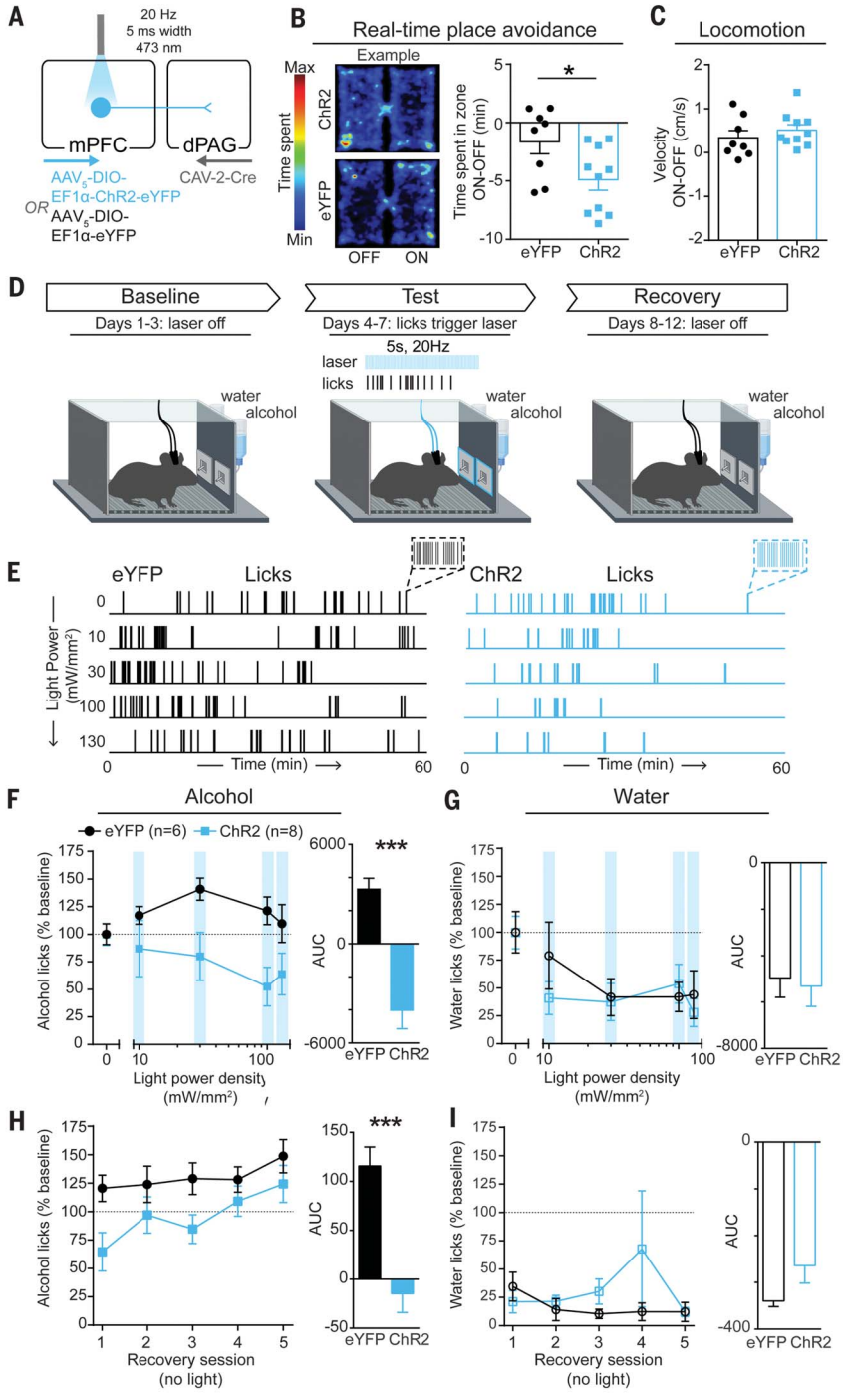


Fig. 4. Activation of mPFC-dPAG neurons mimics the effects of punishment on alcohol consumption. (**A**) Strategy to activate mPFC-dPAG neurons. (**B**) A 20-Hz photostimulation of mPFC-dPAG neurons was avoided in a real-time place avoidance task (unpaired *t* test, $t_{(16)} = 2.356$, **p* = 0.032). (**C**) Photostimulation did not alter locomotion (unpaired *t* test, $t_{(16)} = 0.884$, $p = 0.39$). (**D**) During test days, water or alcohol spout contacts triggered photostimulation delivered at increasing intensities over days (10 to 130 mW/mm^2). During recovery sessions, no light was delivered. (**E**) Example alcohol lick event records. (**F**) Area under the light power density curve was lower in ChR2 animals than in eYFP controls (unpaired *t* test, $t_{(12)} = 5.811$, ****p* = 0.0002). (**G**) Area under the light power density curve did not differ between ChR2 animals and eYFP controls (unpaired *t* test, $t_{(12)} = 0.2834$, $p = 0.78$). (**H**) AUC for alcohol licks during recovery sessions was decreased in ChR2 animals compared with eYFP controls (unpaired *t* test, $t_{(12)} = 4.677$, ****p* = 0.0005). (**I**) AUC for licks on the water spout during recovery sessions did not differ between ChR2 animals and eYFP controls (unpaired *t* test, $t_{(12)} = 1.682$, $p = 0.1184$). Error bars indicate ± SEM.

Downloaded from http://science.sciencemag.org/ on November 29, 2019

We posit that photoinhibition drives compulsive drinking by disrupting the transmittance of a punishment signal from the mPFC to the dPAG. Whereas this circuit encodes the aversive aspects of stimuli (*24*), it does not appear to be specific to pain, given that quinine functions as a punishment but is not a nociceptive stimulus (Fig. 3).

To determine the behavioral impact of driving excitatory activity in this circuit, we bilaterally expressed channelrhodopsin-2 (ChR2) in mPFC-dPAG neurons and implanted optic fibers over the mPFC (Fig. 4A and fig. S11). The photostimulation-paired side was avoided in a real-time place aversion assay (Fig. 4B) without affecting locomotor activity (Fig. 4C) or anxiety-related behavior (fig. S12).

To test the effects of mPFC-dPAG activation on drinking, we again used a two-bottle choice task in which contacts on either the water or alcohol lickometer triggered photostimulation (Fig. 4D). Alcohol and water remained unadulterated throughout the experiment, and the light power delivered to drive photoexcitation was increased across sessions (10 to 130 mW/mm^2), followed by recovery sessions without photostimulation (Fig. 4E). Photostimulation was sufficient to act as a punishment, producing light-power-dependent decreases in licking for alcohol (Fig. 4F) but not water (Fig. 4G), with lasting decreases in alcohol consumption during recovery (Fig. 4, H and I). Microstructural analysis of licking behavior revealed photostimulation-induced changes in bout structure and timing (fig. S13). Photostimulation produced robust and long-lasting decreases in front-loading behavior (drinking a disproportionate amount of alcohol during the initial portion of the access period), a hallmark measure of addiction-like behaviors (fig. S13, L and N).

In conclusion, we established a behavioral model for multidimensional analysis of drinking behaviors and their evolution across time and with experience. We identified a cortical-brainstem circuit that serves as both a biomarker and a circuit-specific cellular substrate for the development of compulsive drinking.

REFERENCES AND NOTES

1. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality, *Results from the 2017 National Survey on Drug Use and Health: Detailed Tables* (2017); www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.pdf.
2. B. F. Grant, R. B. Goldstein, T. D. Saha, S. P. Chou, J. Jung, H. Zhang, R. P. Pickering, W. J. Ruan, S. M. Smith, B. Huang, D. S. Hasin, *JAMA Psychiatry* **72**, 757–766 (2015).
3. F. W. Hopf, H. M. B. Lesscher, *Alcohol* **48**, 253–264 (2014).
4. L. J. M. J. Vanderschuren, B. J. Everitt, *Science* **305**, 1017–1019 (2004).
5. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5®*) (American Psychiatric Association, 2013).
6. K. McFarland, C. C. Lapish, P. W. Kalivas, *J. Neurosci.* **23**, 3531–3537 (2003).
7. B. T. Chen, H.-J. Yau, C. Hatch, I. Kusumoto-Yoshida, S. L. Cho, F. W. Hopf, A. Bonci, *Nature* **496**, 359–362 (2013).
8. G. F. Koob, N. D. Volkow, *Neuropsychopharmacology* **35**, 217–238 (2010).
9. P. W. Kalivas, N. Volkow, J. Seamans, *Neuron* **45**, 647–650 (2005).
10. J. W. Dalley, B. J. Everitt, T. W. Robbins, *Neuron* **69**, 680–694 (2011).
11. M. M. Silveri, J. Rogowska, A. McCaffrey, D. A. Yurgelun-Todd, *Alcohol. Clin. Exp. Res.* **35**, 218–228 (2011).
12. A. D. Schweinsburg, M. P. Paulus, V. C. Barlett, L. A. Killeen, L. C. Caldwell, C. Pulido, S. A. Brown, S. F. Tapert, *Ann. N. Y. Acad. Sci.* **1021**, 391–394 (2004).
13. O. M. Mahmood, D. Goldenberg, R. Thayer, R. Migliorini, A. N. Simmons, S. F. Tapert, *Addict. Behav.* **38**, 1435–1441 (2013).
14. T. Seif, S.-J. Chang, J. A. Simms, S. L. Gibb, J. Dadgar, B. T. Chen, B. K. Harvey, D. Ron, R. O. Messing, A. Bonci, F. W. Hopf, *Nat. Neurosci.* **16**, 1094–1100 (2013).
15. O. George, C. Sanders, J. Freiling, E. Grigoryan, S. Vu, C. D. Allen, E. Crawford, C. D. Mandyam, G. F. Koob, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18156–18161 (2012).
16. A. Holmes, P. J. Fitzgerald, K. P. MacPherson, L. DeBrouse, G. Colacicco, S. M. Flynn, S. Masneuf, K. E. Pleil, C. Li, C. A. Marcinkiewcz, T. L. Kash, O. Gunduz-Cinar, M. Camp, *Nat. Neurosci.* **15**, 1359–1361 (2012).
17. K. E. Pleil, E. G. Lowery-Gionta, N. A. Crowley, C. Li, C. A. Marcinkiewcz, J. H. Rose, N. M. McCall, A. M. Maldonado-Devincci, A. L. Morrow, S. R. Jones, T. L. Kash, *Neuropharmacology* **99**, 735–749 (2015).
18. D. M. Lovinger, J. C. Crabbe, *Nat. Neurosci.* **8**, 1471–1480 (2005).
19. C. Giuliano, Y. Peña-Oliver, C. R. Goodlett, R. N. Cardinal, T. W. Robbins, E. T. Bullmore, D. Belin, B. J. Everitt, *Neuropsychopharmacology* **43**, 728–738 (2018).
20. M. Roberto, M. T. Cruz, N. W. Gilpin, V. Sabino, P. Schweitzer, M. Bajo, P. Cottone, S. G. Madamba, D. G. Stouffer, E. P. Zorrilla, G. F. Koob, G. R. Siggins, L. H. Parsons, *Biol. Psychiatry* **67**, 831–839 (2010).
21. K. Goltseker, F. W. Hopf, S. Barak, *Alcohol* **74**, 73–82 (2019).
22. T. E. Thiele, M. Navarro, *Alcohol* **48**, 235–241 (2014).
23. T. B. Franklin, B. A. Silva, Z. Perova, L. Marrone, M. E. Masferrer, Y. Zhan, A. Kaplan, L. Greetham, V. Verrechia, A. Halman, S. Pagella, A. L. Vyssotski, A. Illarionova, V. Grinevich, T. Branco, C. T. Gross, *Nat. Neurosci.* **20**, 260–270 (2017).
24. C. M. Vander Weele, C. A. Siciliano, G. A. Matthews, P. Namburi, E. M. Izadmehr, I. C. Espinel, E. H. Nieh, E. H. S. Schut, N. Padilla-Coreano, A. Burgos-Robles, C.-J. Chang, E. Y. Kimchi, A. Beyeler, R. Wichmann, C. P. Wildes, K. M. Tye, *Nature* **563**, 397–401 (2018).
25. Y. Li, J. Zeng, J. Zhang, C. Yue, W. Zhong, Z. Liu, Q. Feng, M. Luo, *Neuron* **97**, 911–924.e5 (2018).
26. D. A. Evans, A. V. Stempel, R. Vale, S. Ruehle, Y. Lefler, T. Branco, *Nature* **558**, 590–594 (2018).
27. R. R. Rozeske, D. Jercog, N. Karalis, F. Chaudun, S. Khoder, D. Girard, N. Winke, C. Herry, *Neuron* **97**, 898–910.e6 (2018).
28. A. Cabral, N. Isoardi, C. Salum, C. E. Macedo, M. J. Nobre, V. A. Molina, M. L. Brandão, *Exp. Neurol.* **200**, 200–208 (2006).
29. E. M. Avegno, T. D. Lobell, C. A. Itoga, B. B. Baynes, A. M. Whitaker, M. M. Weera, S. Edwards, J. W. Middleton, N. W. Gilpin, *J. Neurosci.* **38**, 7761–7773 (2018).
30. C. A. Siciliano, K. M. Tye, *Alcohol* **74**, 47–63 (2019).
31. E. J. Kremer, S. Boutin, M. Chillon, O. Danos, *J. Virol.* **74**, 505–512 (2000).
32. J. S. Rhodes, K. Best, J. K. Belknap, D. A. Finn, J. C. Crabbe, *Physiol. Behav.* **84**, 53–63 (2005).
33. J. S. Rhodes, M. M. Ford, C.-H. Yu, L. L. Brown, D. A. Finn, T. Garland Jr., J. C. Crabbe, *Genes Brain Behav.* **6**, 1–18 (2007).
34. M. V. Wilcox, V. C. Cuzon Carlson, N. Sherazee, G. M. Sprow, R. Bock, T. E. Thiele, D. M. Lovinger, V. A. Alvarez, *Neuropsychopharmacology* **39**, 579–594 (2014).
35. Y. Ziv, L. D. Burns, E. D. Cocker, E. O. Hamel, K. K. Ghosh, L. J. Kitch, A. El Gamal, M. J. Schnitzer, *Nat. Neurosci.* **16**, 264–266 (2013).
36. K. K. Ghosh, L. D. Burns, E. D. Cocker, A. Nimmerjahn, Y. Ziv, A. E. Gamal, M. J. Schnitzer, *Nat. Methods* **8**, 871–878 (2011).
37. P. Zhou, S. L. Resendez, J. Rodriguez-Romaguera, J. C. Jimenez, S. Q. Neufeld, A. Giovannucci, J. Friedrich, E. A. Pnevmatikakis, G. D. Stuber, R. Hen, M. A. Kheirbek, B. L. Sabatini, R. E. Kass, L. Paninski, *eLife* **7**, e28728 (2018).
38. M. Murugan, H. J. Jang, M. Park, E. M. Miller, J. Cox, J. P. Taliaferro, N. F. Parker, V. Bhave, H. Hur, Y. Liang, A. R. Nectow, J. W. Pillow, I. B. Witten, *Cell* **171**, 1663–1677.e16 (2017).
39. J. M. Otis, V. M. K. Namboodiri, A. M. Matan, E. S. Voets, E. P. Mohorn, O. Kosyk, J. A. McHenry, J. E. Robinson, S. L. Resendez, M. A. Rossi, G. D. Stuber, *Nature* **543**, 103–107 (2017).
40. C.-J. Chang, *Technical Report: Building a Neural Ensemble Decoder by Extracting Features Shared Across Multiple Populations* (MIT, 2019); https://dspace.mit.edu/bitstream/handle/1721.1/122041/Chang-Technical_Report092019.pdf?sequence=1&isAllowed=y.
41. A. M. Stamatakis, G. D. Stuber, *Nat. Neurosci.* **15**, 1105–1107 (2012).
42. A. V. Kravitz, L. D. Tye, A. C. Kreitzer, *Nat. Neurosci.* **15**, 816–818 (2012).
43. J. Olds, P. Milner, *J. Comp. Physiol. Psychol.* **47**, 419–427 (1954).
44. G. D. Stuber, D. R. Sparta, A. M. Stamatakis, W. A. van Leeuwen, J. E. Hardjoprajitno, S. Cho, K. M. Tye, K. A. Kempadoo, F. Zhang, K. Deisseroth, A. Bonci, *Nature* **475**, 377–380 (2011).
45. S. Pellow, P. Chopin, S. E. File, M. Briley, *J. Neurosci. Methods* **14**, 149–167 (1985).

ACKNOWLEDGMENTS

We thank Y.-N. Leow for experimental assistance and C. Wildes and R. Wichmann for technical assistance. **Funding:** C.A.S. was supported by NIH grants F32 MH111216 (NIMH) and K99 DA045103 (NIDA) and a NARSAD Young Investigator Award from the Brain and Behavior Research Foundation. K.M.T. is a New York Stem Cell Foundation–Robertson Investigator and McKnight Scholar and was supported by funding from the JPB Foundation, New York Stem Cell Foundation, R01-MH102441 (NIMH), the NIH Director's New Innovator Award DP2-DK102256 (NIDDK), and Pioneer Award DP1-AT009925 (NCCIH). M.H. was supported by the JPB Foundation. D.L. is enrolled as a student at the Medical Faculty Heidelberg, Heidelberg University, Germany, and was supported by a fellowship from the Boehringer Ingelheim Fonds. **Author contributions:** C.A.S. and K.M.T. jointly conceived of the project and designed the experiments. C.A.S., X.C., and A.R.B. performed stereotaxic surgeries. C.A.S. performed calcium-imaging experiments. C.A.S. and E.Y.K. analyzed the data. C.J.C. performed clustering and machine-learning analyses. C.A.S., H.N., D.L., and X.C. performed and analyzed optogenetic experiments. C.A.S., D.L., C.M.V.W., A.R.B., and X.C. performed histological verifications. C.J.C., H.N., J.J.L., and E.Y.K. provided MATLAB scripts and advice for data analysis. C.M.V.W. and J.W. provided technical training. C.A.S., K.M.T., E.Y.K., C.J.C., H.N., M.H., A.N.V., A.R.B., and C.M.V.W. made additional significant intellectual contributions. C.A.S. and K.M.T. wrote the paper. All authors contributed to the editing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All experimental data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/1008/suppl/DC1
Materials and Methods
Figs. S1 to S13
References (*31*–*50*)

View/request a protocol for this paper from *Bio-protocol*.

21 May 2019; accepted 4 October 2019
10.1126/science.aay1186

Downloaded from http://science.sciencemag.org/ on November 29, 2019

IMMUNOMETABOLISM

Glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion

Robert D. Leone[1], Liang Zhao[1], Judson M. Englert[1], Im-Meng Sun[1], Min-Hee Oh[1], Im-Hong Sun[1], Matthew L. Arwood[1], Ian A. Bettencourt[1], Chirag H. Patel[1], Jiayu Wen[1], Ada Tam[1], Richard L. Blosser[1], Eva Prchalova[2], Jesse Alt[2], Rana Rais[2], Barbara S. Slusher[2], Jonathan D. Powell[1]*

The metabolic characteristics of tumors present considerable hurdles to immune cell function and cancer immunotherapy. Using a glutamine antagonist, we metabolically dismantled the immunosuppressive microenvironment of tumors. We demonstrate that glutamine blockade in tumor-bearing mice suppresses oxidative and glycolytic metabolism of cancer cells, leading to decreased hypoxia, acidosis, and nutrient depletion. By contrast, effector T cells responded to glutamine antagonism by markedly up-regulating oxidative metabolism and adopting a long-lived, highly activated phenotype. These divergent changes in cellular metabolism and programming form the basis for potent antitumor responses. Glutamine antagonism therefore exposes a previously undefined difference in metabolic plasticity between cancer cells and effector T cells that can be exploited as a "metabolic checkpoint" for tumor immunotherapy.

Cancer cells have prodigious anabolic and energetic requirements and use distinct metabolic pathways for growth and survival (*1*, *2*). Termed the Warburg effect, cancer cells are highly glycolytic, metabolizing glucose to lactic acid to produce adenosine triphosphate (ATP) and regenerate nicotinamide adenine dinucleotide (NAD^+) (*1*, *3*). In addition to glucose, glutamine plays an important role in providing both carbon and nitrogen necessary for anabolic metabolism (*3*, *4*). As glucose is metabolized through glycolytic pathways to lactic acid, glutamine can fuel the tricarboxylic acid (TCA) cycle, generating metabolic intermediates to serve as building blocks for lipids, proteins, and nucleic acids that are crucial to anabolic growth and proliferation. Notably, these same metabolic programs have also been attributed to facilitating the tremendous growth associated with T cell activation and proliferation (*5*, *6*). An important consequence of the specialized metabolism of cancer cells is the creation of a hypoxic, acidic, nutrient-depleted tumor microenvironment (TME) that is hostile to antitumor immune responses (*7*–*14*). In other words, the specialized metabolic programming of cancer cells not only promotes tumor growth but can also block the generation of an effective antitumor immune response.

We hypothesized that blocking glutamine metabolism would not only inhibit tumor growth but also relieve a master immunometabolic checkpoint and enable restoration of antitumor immunity. Glutaminase facilitates the conversion of glutamine to glutamate, and the subsequent conversion of glutamate to α-ketoglutarate represents a major pathway for glutamine carbons to enter the TCA cycle (*1*, *3*). However, the therapeutic response of tumors in vivo to targeted blockade of glutaminase is generally limited (*15*–*17*). Instead of inhibiting a single enzyme, we chose to comprehensively inhibit glutamine metabolism by using drugs based on the glutamine antagonist 6-diazo-5-oxo-L-norleucine (DON). In addition to glutaminase, DON inhibits a broad range of glutamine-requiring enzymes (*18*). Consequently, DON exposure potently inhibits tumor cell viability, proliferation, and cell cycle progression (fig. S1, A to E). Indeed, DON has previously been evaluated in a number of clinical trials for cancer, but these pursuits were eventually abandoned secondary to unacceptable toxicity, particularly to the gastrointestinal tract (*19*–*24*). Given the extraordinary potential of DON to not only inhibit tumor growth but also condition the TME and enhance immunotherapy, we have developed a series of prodrugs of DON that are designed to circulate intact and inert but are activated in the TME upon enzymatic cleavage of the promoieties, thus mitigating the previously reported toxicity of glutamine antagonists on susceptible tissues such as the gut (fig. S2, A and B) (*19*, *25*). Of note, although we do observe an enrichment of the active, parent compound, DON, in tumor over plasma in our mouse model (fig. S2C), these prodrugs are much more rapidly activated in rodents compared with larger animals (*25*–*27*). Nonetheless, we established a robust therapeutic window for in vivo mouse studies using the compound JHU083 to explore whether tumor glutamine metabolism can be targeted as an immunologic checkpoint (fig. S2, A to C) (*19*, *25*, *28*–*30*).

To examine the ability of JHU083 to inhibit tumor growth, we performed a series of experiments using syngeneic tumor models. Mice were injected subcutaneously with MC38 colon cancer, EL-4 lymphoma, CT26 colon cancer, or B16 melanoma. For each tumor type tested, treatment with JHU083 led to a marked decrease in tumor growth and improved survival (Fig. 1A and fig. S2, D to F). In the case of MC38, JHU083 monotherapy for 14 days led to a substantial proportion of durable responses. Given the ability of JHU083 to cause tumor regression, we used the MC38 in vivo model to examine the specific effects on tumor metabolism. Using ^{13}C-glucose tracing with mass spectrometry, we observed differential enrichment of glucose metabolites in metabolic pathways. Blocking glutamine metabolism inhibited glucose metabolism through the TCA cycle, as well as through glycolysis and related pathways, thus disabling Warburg physiology (Fig. 1B and fig. S2G). We also noted a significant decline in glucose contribution to purine synthesis, the hexosamine pathway, and the pentose phosphate pathway (Fig. 1B). This broad suppression of metabolic activity was accompanied by a significant increase in glutamine and glucose content within tumors (Fig. 1C), as well as a decrease in tumor hypoxia (Fig. 1D). Such findings suggest that blocking glutamine metabolism can critically disrupt the metabolism of the tumor as a whole, with marked effects on the nutrient milieu within the TME.

Given the pronounced changes observed in the TME, we asked whether targeting glutamine metabolism would enhance the efficacy of immunotherapy. We were initially concerned that blocking glutamine metabolism would severely impair lymphocyte function, thus metabolic treatment was administered before initiating adoptive cell transfer (ACT) immunotherapy. Mice harboring ovalbumin (OVA)–expressing B16 melanoma (B16OVA) were pretreated with JHU083 for 3 days before adoptive transfer of activated OVA-specific $CD8^+$ T cells (OT1 cells). Treated mice showed improvement in tumor control and survival (Fig. 1E), suggesting that glutamine blockade can condition the TME to enhance ACT. Next, we combined glutamine targeting with programmed cell death protein 1 (PD-1) checkpoint immunotherapy. Because it is possible that glutamine antagonism could suppress antitumor immune cells, we designed our initial treatment regimens to use sequential dosing of JHU083 for a defined period of several days, followed by anti–PD-1 treatment. However, we found that the most efficacious regimens for inducing antitumor response required concurrent, not sequential, dosing of JHU083 and anti–PD-1. To this end, we found that tumor-bearing mice treated with concurrent dosing of the glutamine antagonist and anti–PD-1 showed dramatically improved antitumor effects compared with anti–PD-1 therapy alone, with complete response rates approaching 100%

[1]The Bloomberg-Kimmel Institute for Cancer Immunotherapy at Johns Hopkins, Baltimore, MD 21287, USA. [2]Johns Hopkins Drug Discovery, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA.
*Corresponding author. Email: poweljo@jhmi.edu

Downloaded from http://science.sciencemag.org/ on December 2, 2019

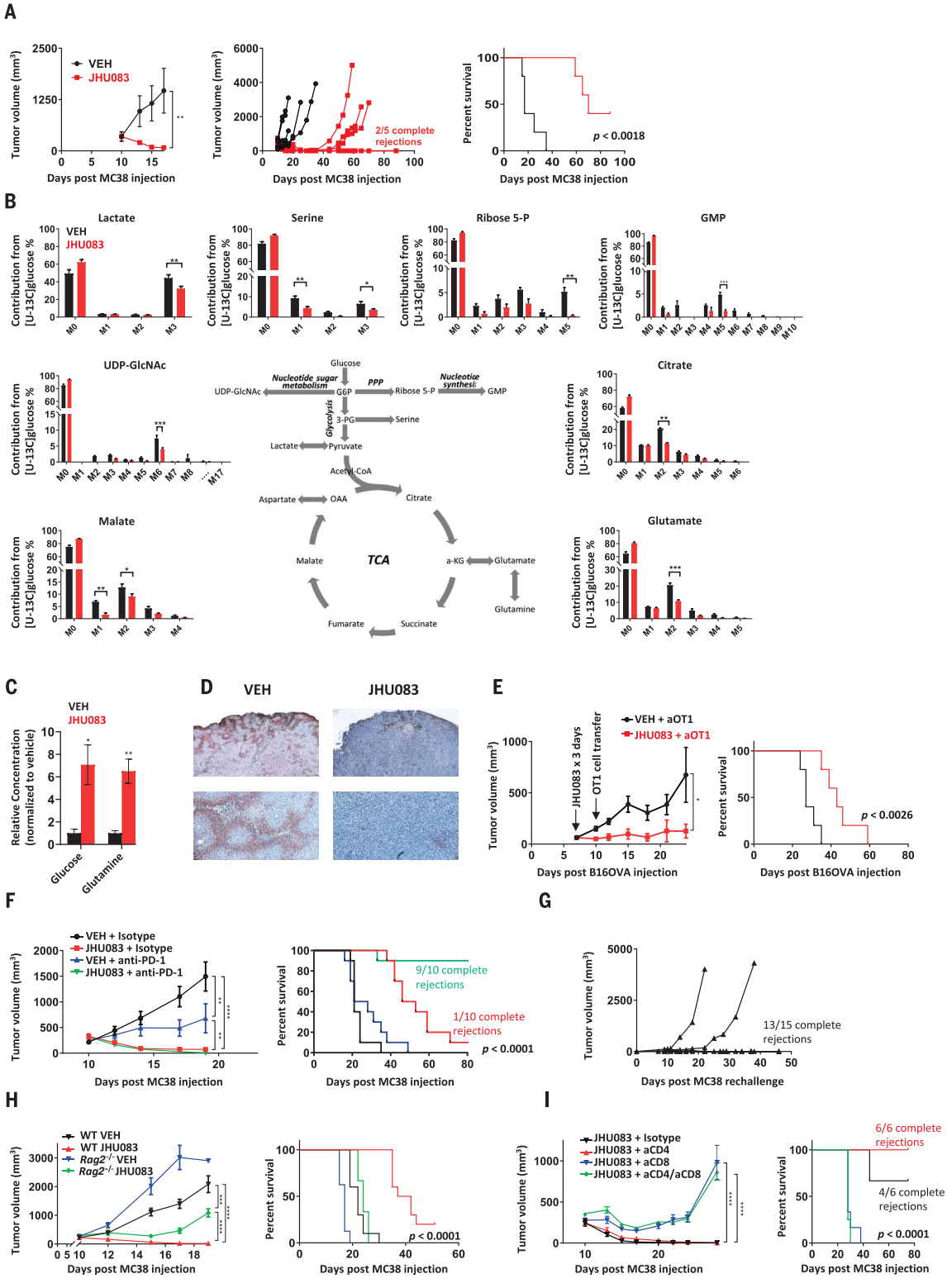


Fig. 1. Glutamine blockade suppresses cancer cell metabolic programs and enhances antitumor immune response. (**A**) Average tumor growth curve (left), spider plots (center), and survival curve (right) from vehicle (VEH)– and JHU083-treated MC38-bearing mice. (**B**) In vivo ^{13}C-glucose tracing experiment in MC38 tumor–bearing mice. M, unlabeled mass of isotope; M+*n*, native metabolite mass (M) plus number of isotopically labeled carbons (*n*); UDP-GlcNAc, uridine diphosphate N-acetylglucosamine; G6P, glucose-6-phosphate; GMP, guanosine monophosphate; 3-PG, D-glycerate 3-phosphate; OAA, oxaloacetate; α-KG, α-ketoglutarate. (**C**) Relative mass spectrometric quantification of glucose and glutamine in MC38 tumors from vehicle- and JHU083-treated mice (per milligram tumor tissue, normalized to vehicle group). (**D**) Pimonidazole immunohistochemistry staining for hypoxia in tumor sections from vehicle- and JHU083-treated mice. (**E**) B16OVA-bearing C57BL/6 mice treated with JHU083 or vehicle on days 7 to 9 after tumor inoculation received 1.5 × 10^6 activated OTI T cells on day 10. Tumor growth curve (left) and survival curve (right) are shown. (**F**) MC38-bearing C57BL/6 mice treated with vehicle, JHU083, anti–PD-1, or combination JHU083 and anti–PD-1 beginning on day 10 after tumor inoculation. Tumor growth curve (left) and survival curve (right) are shown. (**G**) Mice initially cured with 14 days of JHU083 treatment were rechallenged ≥30 days after last dose of JHU083; spider plots of tumor volume are shown. (**H**) MC38-bearing C57BL/6 wild-type (WT) and *Rag2*$^{-/-}$ mice treated with 14 days of vehicle or JHU083. Average tumor volume (until first sacrifice in WT VEH group) (left) and survival curve (right) are shown. (**I**) MC38-bearing C57BL/6 mice treated with JHU083 after depletion of CD8 cells, CD4 cells, or both compared with isotype control. Tumor growth curve (left) and survival curve (right) are shown. Error bars represent SEM. Data are representative of one (B), three [(A), (E) and (F)], or five (D) independent experiments with $n = 3$ to 10 mice per group. Tumor growth curves were assessed by two-way analysis of variance (ANOVA). Log-rank (Mantel-Cox) tests were performed for survival data. Metabolite data assessed with two-tailed Student's *t* test for multiple comparisons. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$ using two-tailed Student's *t* test.

Downloaded from http://science.sciencemag.org/ on December 2, 2019

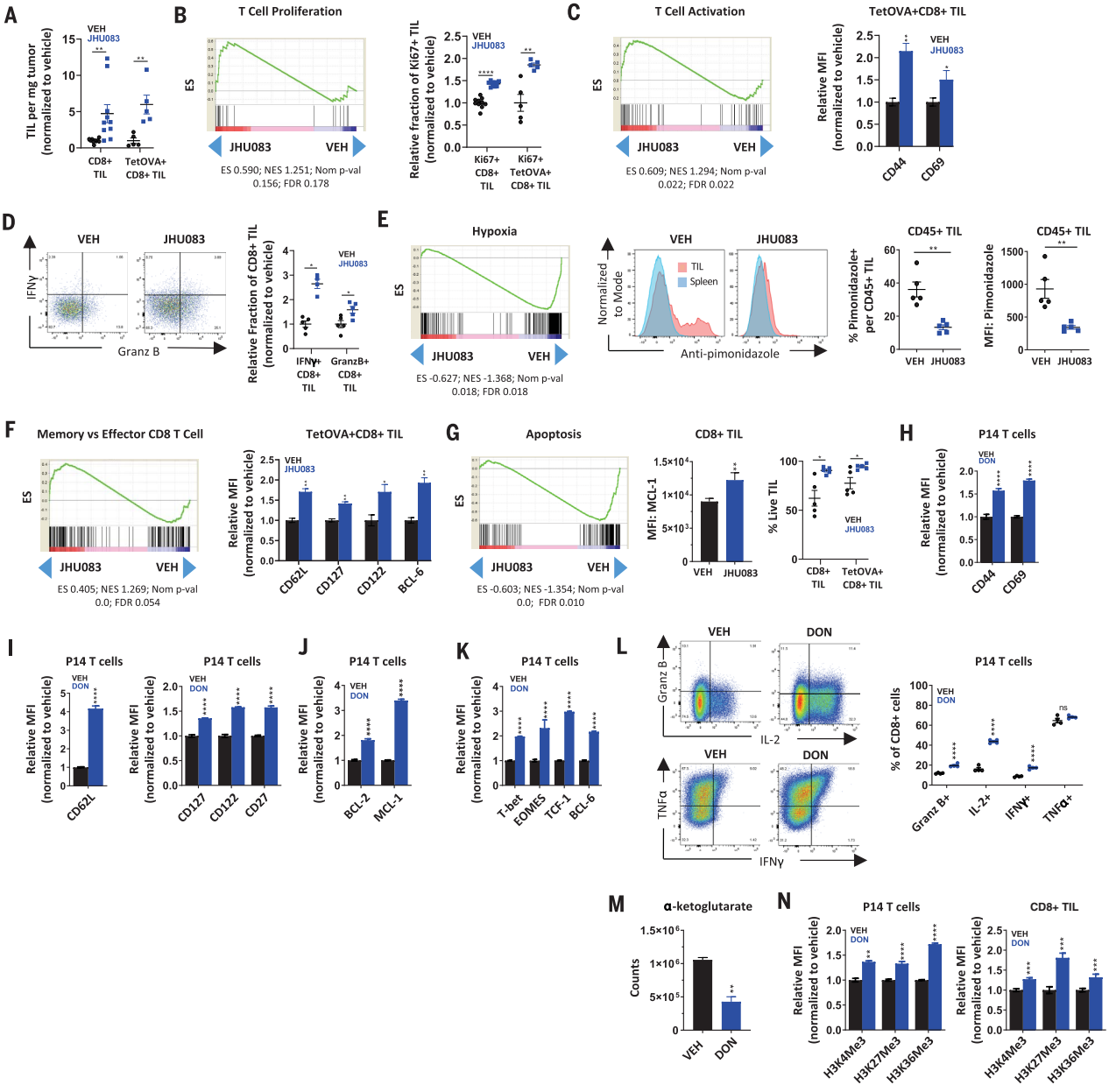


Fig. 2. Glutamine blockade conditions CD8$^+$ TILs toward a highly proliferative, activated, and long-lived phenotype. C57BL/6 mice with MC38 or MC38OVA tumors were treated on days 14 to 18 with vehicle or JHU083 (0.3 mg per kilogram of body weight per day) and sacrificed on day 18 for analysis. (**A**) Percentage of live CD8$^+$ or TetOVA$^+$CD8$^+$ TILs per tumor weight for MC38 and MC38OVA models, respectively. TetOVA, tetramer-OVA. (**B**) GSEA tracing for positive regulation of αβ T cell proliferation (left) and percentage of Ki67$^+$ cells per CD8$^+$ in MC38 model or TetOVA$^+$CD8$^+$ in MC38OVA model (right). ES, enrichment score; NES, normalized enrichment score; FDR, false discovery ratio. (**C**) GSEA tracing for αβ T cell activation (left) and mean fluorescence intensity (MFI) of CD44 and CD69 of TetOVA$^+$CD8$^+$ TILs in MC38OVA model (right). (**D**) Relative fraction (normalized to vehicle) of IFN-γ$^+$ and granzyme B$^+$ per CD8$^+$ TILs from the MC38OVA model after ex vivo stimulation with SIINFEKL peptide for 4 hours. (**E**) GSEA tracing for hypoxic exposure (left) and FACS plots (center) and data summary (right) showing percentage of pimonidazole positive and pimonidazole MFI of CD45$^+$ TILs in the MC38 model. (**F**) GSEA tracings for memory versus KLRG1^{high} effector CD8$^+$ T cells (left) and relative MFI of CD62L, CD127, CD122, and BCL-6 of TetOVA$^+$CD8$^+$ TILs (right). (**G**) GSEA tracing for apoptosis (left), MFI of MCL-1 on CD8$^+$ in MC38 model (center), and live-cell percentage of CD8$^+$ TILs in MC38 model and TetOVA$^+$CD8$^+$ in MC38OVA model (right). (**H** to **K**) Naïve P14 T cells activated in the presence of vehicle or DON (1 μM) for 2 days, rested in IL-2 for two additional days with vehicle or DON, and analyzed by FACS for activation markers (H), memory markers (I), survival markers (J), and transcription factors (K). (**L**) P14 T cells activated for 2 days as described and rested for 4 days in the presence of vehicle or DON before restimulation (with no drug present) for 4 hours. Flow plots (left) and data summary (right) for intracellular cytokines are shown. (**M**) Relative α-ketoglutarate levels at 36 hours post–P14 activation. (**N**) H3 histone trimethylation levels in P14 T cells 3 days post-activation (left) and CD8$^+$ TILs (right). H3K4Me3, trimethylated histone H3 lysine 4; H3K27Me3, trimethylated histone H3 lysine 27; H3K36Me3, trimethylated histone H3 lysine 36. Error bars represent SEM. For MC38OVA experiments, FACS data are representative of two or three independent experiments with $n = 5$ mice per group. For MC38 experiments, FACS summary plots are combination data from two independent experiments with $n = 5$ mice per group. For RNA-seq, data are from treated and untreated groups of five mice in each group. In vitro experiments are representative of three to five independent experiments with $n = 3$ to 6. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, and ns is not significant using two-tailed Student's t test.

Downloaded from http://science.sciencemag.org/ on December 2, 2019

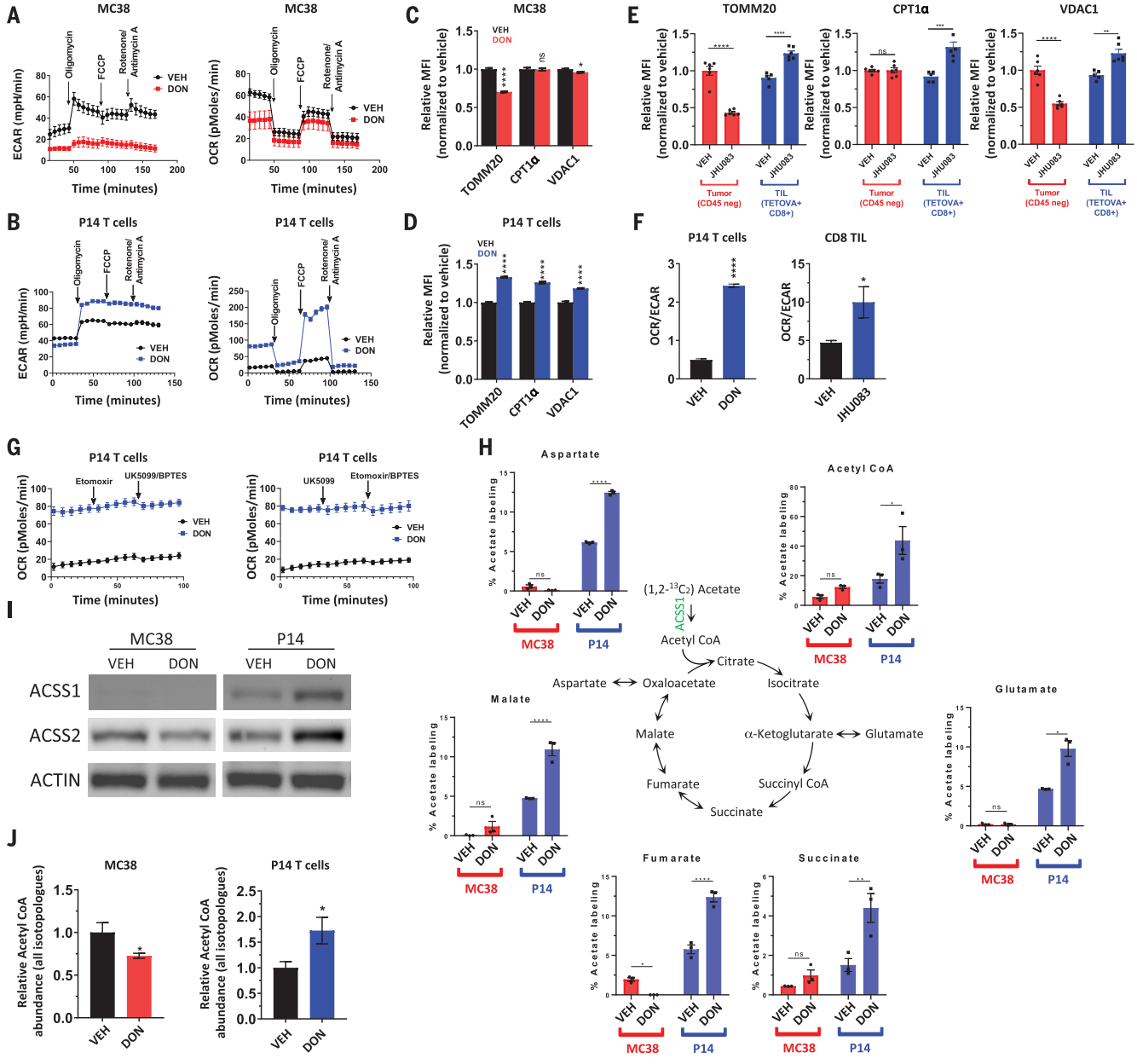
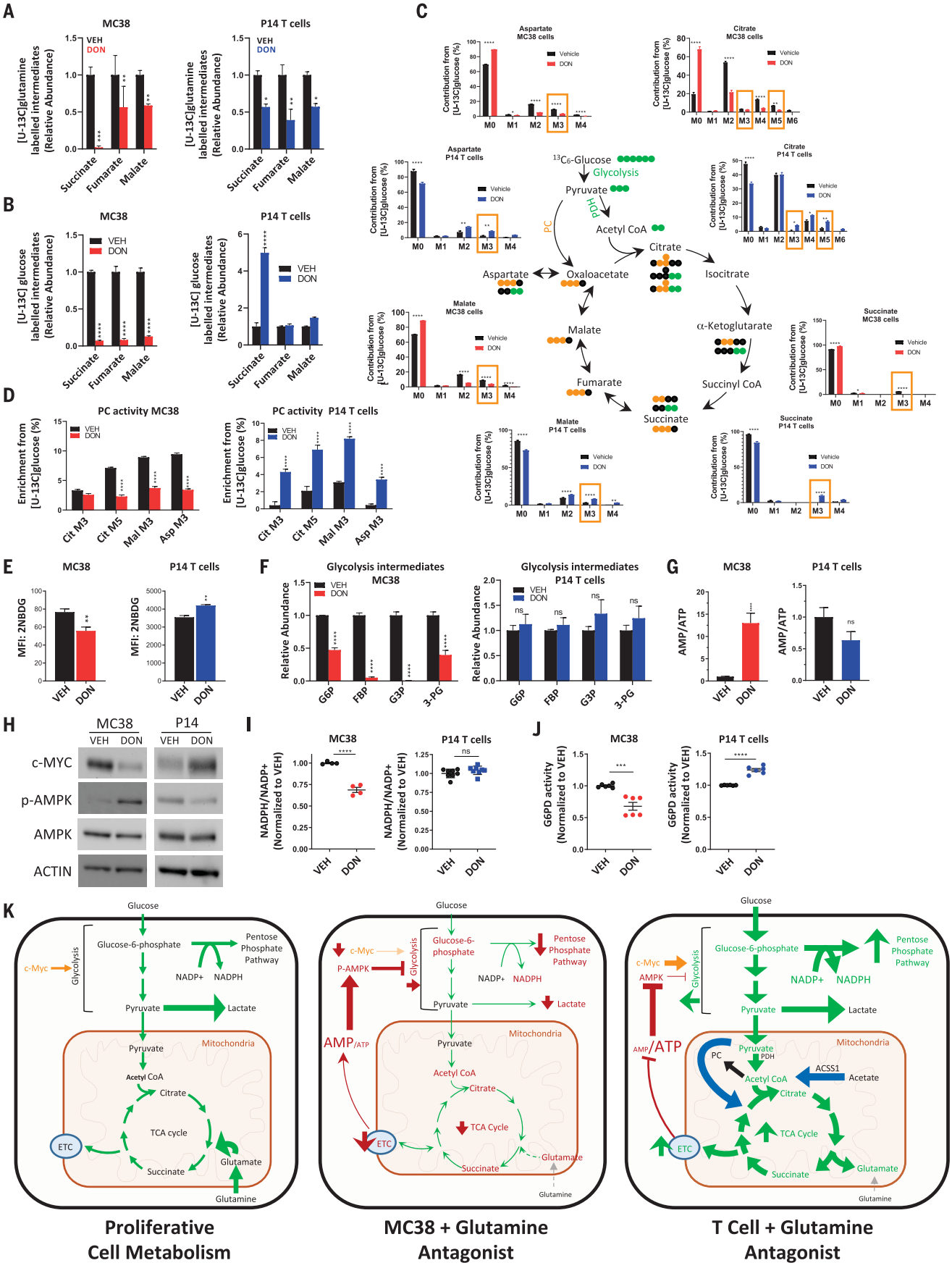


Fig. 3. Activated CD8$^+$ T cells and MC38 tumor cells enact distinct metabolic programs in response to glutamine antagonist treatment. (**A** to **D**) Differential metabolic characteristics of vehicle- versus DON-treated MC38 cells and vehicle- versus DON-treated activated P14 CD8$^+$ T cells in vitro. ECAR and OCR from metabolic flux analyses [(A) and (B)]. FCCP, carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone. Relative expression of mitochondrial proteins [(C) and (D)]. (**E**) Differential expression of mitochondrial proteins in CD8$^+$ TILs and CD45-negative tumor cells from explanted MC38 tumors. (**F**) OCR/ECAR ratio in vehicle- and DON-treated activated P14 CD8$^+$ T cells in vitro (left) and in CD8$^+$ TILs harvested from MC38 murine tumors after vehicle or JHU083 treatment (right). (**G**) OCR response to etomoxir, UK5099, and BPTES in vehicle- versus DON-treated activated P14 CD8$^+$ T cells in vitro. (**H** to **J**) Vehicle-versus DON-treated MC38 cells and vehicle- and DON-treated activated T CD8$^+$ T cells in vitro. Liquid chromatography–mass spectrometry (LC-MS) analysis of TCA intermediates after stable isotope tracing with [1,2-^{13}C] acetate (H). Western blot analysis of ACSS1 and ACSS2 expression. ACTIN is used as a loading control (I). LC-MS analysis of relative acetyl-CoA abundance (J). Error bars represent SEM. Data are representative of three [(A), (C), (D), (E), (G), and (I)] or six [(B) and (F)] independent experiments with $n = 3$ to 8 mice per group. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, and ns is not significant using two-tailed Student's t test.

in the MC38 model (versus no complete response in mice receiving anti–PD-1 monotherapy) (Fig. 1F and fig. S2H). Importantly, these findings indicate that blockade of glutamine metabolism, despite the established role of glutamine for activated lymphocyte proliferation and function, does not metabolically disable antitumor immune cells, but to the contrary, appears to robustly enhance their function and effectiveness.

These findings, in addition to the observed ability of glutamine antagonism to promote durable tumor remissions as a single agent (Fig. 1A), led us to investigate the intriguing hypothesis that blocking glutamine metabolism alone could enhance endogenous antitumor responses, even in the absence of additional immunotherapy. To test this, we

Downloaded from http://science.sciencemag.org/ on December 2, 2019



Downloaded from http://science.sciencemag.org/ on December 2, 2019

rechallenged mice that had been cured with JHU083 monotherapy with an equal burden of the same tumor injected on the opposite flank. Notably, almost all of the animals cured by single-agent JHU083 completely rejected the tumor upon rechallenge (Fig. 1G), suggesting the establishment of immunologic memory. To confirm the immunologic basis of this phenomenon, we compared JHU083 treatment of MC38-bearing *Rag2*$^{-/-}$ mice and wild-type mice. RAG2-deficient mice lack the V(D)J recombinase machinery necessary for rearranging antigen receptor genes and, as such, do not produce mature B or T lymphocytes and are incapable of mounting an adaptive immune response. Although glutamine blockade with JHU083 had some initial effect on tumor growth in the *Rag2*$^{-/-}$ mice, tumor growth rates recovered after several days and the animals succumbed to the disease in a similar time frame to untreated wild-type mice (Fig. 1H). Using T cell depleting antibodies (*31*), we were able to demonstrate that the efficacy of glutamine antagonist therapy was completely dependent upon the activity of the CD8$^+$ T cell subset of the adaptive immune response (Fig. 1I). These observations demonstrate that glutamine blockade, without additional immunotherapy, markedly enhances endogenous antitumor immunity and, more generally, that tumor immune rejection and adaptive immune memory can be triggered solely through this metabolic intervention.

To further investigate the mechanism of this enhanced antitumor response to JHU083 monotherapy, we examined tumor-infiltrating lymphocytes (TILs) from vehicle- and JHU083-treated mice bearing either MC38 or OVA-expressing MC38 tumors. OVA-expressing MC38 tumors (M38OVA) allow identification and study of antigen specific CD8$^+$ T cells by tetramer staining. Consistent with our depletion studies, we observed a marked increase in infiltrating CD8$^+$ T cells with JHU083 treatment, and a greater percentage of these TILs were antigen specific (Fig. 2A). To explore the mechanism of the enhanced endogenous antitumor immunity, we performed RNA sequencing (RNA-seq) on CD8$^+$ TILs sorted by fluorescence-activated cell sorting (FACS). RNA-seq revealed distinct differences between TILs from treated and untreated mice (fig. S3, A and B), with 4313 genes differentially expressed in a statistically significant manner. Gene set enrichment analyses (GSEA) showed that CD8$^+$ TILs from treated mice had transcriptional programs consistent with a highly proliferative (Fig. 2B), robustly activated (Fig. 2C), less exhausted and less anergic phenotype (fig. S3C). These transcriptional changes were also reflected at the protein level. FACS analyses of CD8$^+$ TILs from MC38 tumors and of tetramer-positive CD8$^+$ TILs from OVA-expressing MC38 tumors showed increased proliferation and activation markers (Fig. 2, B and C, and fig. S3D), a decreased percentage of PD-1$^+$LAG-3$^+$ double-positive TILs (fig. S3E), and enhanced interferon-γ (IFN-γ) (Fig. 2D), granzyme B (Fig. 2D), and interleukin-2 (IL-2) (fig. S3F) production upon ex vivo stimulation. As was anticipated from our preliminary TME analysis, GSEA also suggested that CD8$^+$ TILs in treated mice experience less hypoxia and oxidative stress (Fig. 2E and fig. S3G), which was corroborated by flow cytometry showing decreased pimonidazole (a marker of exposure to low oxygen tension) staining on CD45$^+$ TILs (Fig. 2E). Furthermore, we observed a relative up-regulation of transcripts associated with long-lived memory T cells (Fig. 2F) and a marked down-regulation of apoptotic transcriptional programs in TILs from mice treated with glutamine blockade (Fig. 2G and fig. S3, H and I). Indeed, increased expression of memory markers (CD62L, CD127, CD122, and BCL-6) (Fig. 2F and fig. S3, J and K) and increased expression of anti-apoptotic protein MCL-1 (Fig. 2G) were also observed by flow cytometry. We also confirmed by reverse transcription polymerase chain reaction (RT-PCR) that apoptotic genes uncovered through leading edge GSEA analysis (fig. S3, H and I) were indeed down-regulated in CD8$^+$ TILs from JHU083-treated mice (fig. S3L). This down-regulation has a tangible effect, as demonstrated by the fact that CD8$^+$ TILs, or tetramer-OVA$^+$CD8$^+$ TILs, were less susceptible to intratumoral cell death in JHU083-treated mice (Fig. 2G, right). In accord with recent reports, these findings suggest that an important mechanism of CD8$^+$ T cell dysfunction and tumor immune evasion may be apoptosis of tumor-specific T cells within the TME (*32*). Blocking glutamine metabolism led to similar changes when combined with anti–PD-1 treatment (fig. S3, D, E, J, and K). Taken together, our results demonstrate that glutamine blockade conditions TILs toward a long-lived, memory-like phenotype that is highly proliferative, markedly activated, and capable of enhanced effector function.

Our initial studies were motivated by the hypothesis that targeting tumor metabolism would alter the TME to enhance antitumor immune responses. Indeed, the decreased exposure of TILs to hypoxia in treated mice directly supports this mechanism (Figs. 1D and 2E). However, the notable changes observed in the activation and differentiation states of the TILs from treated and untreated mice led us to question whether glutamine antagonism was also directly affecting T cell activation and function. To assess the cell-intrinsic nature of these findings, we activated and expanded naïve CD8$^+$ T cells with anti-CD3, anti-CD28, and IL-2 in the presence or absence of DON. (Because JHU083 requires in vivo enzymatic cleavage of promoieties, DON is used for all in vitro studies.) Much like our in vivo TIL analysis, in vitro activated T cells, in the presence of DON, showed robust up-regulation of activation markers (CD44 and CD69), memory markers (CD62L, CD127, CD122, and CD27), anti-apoptotic proteins (BCL-2 and MCL-1), and transcription factors (T-bet, EOMES, TCF-1,

Downloaded from http://science.sciencemag.org/ on December 2, 2019

Fig. 4. In response to glutamine blockade, activated CD8$^+$ T cells, but not MC38 tumor cells, replenish TCA intermediates by up-regulating glucose anaplerosis. (**A** to **J**) Metabolic characteristics of vehicle- versus DON-treated MC38 cells (red) and vehicle- versus DON-treated activated P14 CD8$^+$ T cells (blue) in vitro. Relative abundance of labeled intermediates of the TCA cycle during [U-^{13}C] glutamine (A) and [U-^{13}C] glucose (B) tracing experiments. LC-MS analysis of TCA intermediates with [U-^{13}C] glucose tracing (C). For (C), in the bar graphs, M3 and M5 values enclosed in yellow rectangles correspond to isotopologs indicative of PC activity, and in the TCA cycle diagram, PDH is pyruvate dehydrogenase; green and yellow circles indicate carbon atoms derived from PDH and PC activity, respectively; and black circles indicate unlabeled carbon atoms. Relative labeling of TCA isotopologues characteristic of PC activity (D). 2-NBDG uptake by flow cytometry analysis (E). Relative abundance of glycotic intermediates in MC38 (left) and P14 T cells (right) (F). G6P, glucose-6-phosphate; FDP, fructose-1,6-bis-phosphate; G3P, D-glyceraldehyde 3-phosphate; 3-PG, D-glycerate 3-phosphate. Relative AMP/ATP ratio (G). Western blot of c-MYC, phospho-AMPK, and total AMPK expression (H). ACTIN is used as a loading control. Relative NADP$^+$/NADPH ratio (I). Relative G6PD activity (J). Error bars represent SEM. Data are representative of three or four independent experiments [(E), (H), (I), and (J)] with $n = 3$ or 4 [(E), (I), and (J)]. (G) is abundance data compiled from three independent tracing studies. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, and ns is not significant using two-tailed Student's t test. (**K**) Model depicting relative activity of fundamental metabolic pathways in highly proliferative cells (left), glutamine-inhibited MC38 cancer cells (center), and glutamine-inhibited effector CD8$^+$ T cells (right). Highly proliferative cells in nutrient-rich microenvironments engage high levels of aerobic glycolysis (Warburg physiology), PPP activity, and glutaminolysis to maintain energy, redox, and metabolite homeostasis. Disruption of glutamine metabolism in MC38 cells leads to increased AMP/ATP ratio and decreased c-MYC, such that proximal glycolytic metabolism is suppressed and cells can no longer rely on Warburg physiology, the PPP, or TCA cycle activity. By contrast, activated T cells adapt to glutamine blockade and maintain redox and energy homeostasis by up-regulating OXPHOS through acetate catabolism, generating high levels of acetyl-CoA as a two-carbon source for the TCA cycle, and up-regulating PC activity for glucose anaplerosis. ETC, electron transport chain.

and BCL-6) (Fig. 2, H to K). Moreover, restimulation of DON-treated $CD8^+$ T cells reveals an increased capacity to produce cytokines and granzyme B (Fig. 2L). These phenotypic changes are also observed when glutamine metabolism is inhibited in the context of acute infection with lymphocytic choriomeningitis virus (LCMV) Armstrong (fig. S3, M to O). Transcriptional changes associated with a memory T cell phenotype have previously been reported in response to modulation of α-ketoglutarate–dependent demethylases, including critical epigenetic remodeling enzymes (*33*, *34*). α-Ketoglutarate is a proximal metabolite of glutamine metabolism. We observed significant attenuation of α-ketoglutarate levels in DON-treated CD8 T cells 36 hours post-activation (Fig. 2M), with increased methylation (corresponding to decreased activity of histone demethylase enzymes) on a broad range of histone methylation sites (Fig. 2N, left). Importantly, identical changes in histone methylation are evident on TILs from explanted tumors (Fig. 2N, right). These findings demonstrate that glutamine blockade conditions, even in the absence of tumor, activated T cells in a cell-intrinsic manner toward a highly activated, long-lived, memory-like differentiation state.

Given the differential ability of glutamine metabolic inhibition to disable tumor cells while enhancing T cell function, we were interested in defining the cell-specific metabolic adaptations of cancer cells and $CD8^+$ effector T cells in response to glutamine blockade. Metabolic flux analyses revealed that glutamine blockade by DON suppresses aerobic glycolysis (i.e., Warburg physiology) in cancer cells and activated $CD8^+$ T cells (Fig. 3, A and B, and fig. S4, A and B) as measured by extracellular acidification rate (ECAR). However, whereas DON profoundly suppresses oxidative phosphorylation (OXPHOS) in cancer cells, DON exposure during $CD8^+$ T cell activation triggers a dramatic up-regulation of oxidative metabolism and spare respiratory capacity (SRC), defined as the ability of cells to increase OXPHOS above baseline (Fig. 3, A and B, and fig. S4C). In DON-treated T cells, but not in MC38 cancer cells, these changes in metabolic programs are associated with up-regulation of mitochondrial proteins TOMM20, CPT1α, and VDAC1 (Fig. 3, C and D). This is correlated with in vivo changes showing differential changes in expression of the same mitochondrial proteins in $CD8^+$ TILs versus CD45-negative tumor cells in the TME (Fig. 3E). Functional changes favoring OXPHOS (with increased SRC) over Warburg physiology in T cells are consistent with metabolic characteristics of a T cell memory phenotype and can be approximated by the oxygen consumption rate (OCR)/ECAR ratio (*35*). In accord with memory-like phenotypic changes we observed in T cells exposed to JHU083 (Fig. 2, F, I, J, and K), we observed a markedly increased OCR/ECAR ratio in both in vitro T cell experiments as well as in $CD8^+$ TILs sorted from treated and untreated MC38-bearing mice (Fig. 3F). To further explore the different metabolic responses in T cells and cancer cells, we were interested in determining what nutrient or nutrients were fueling increased OXPHOS in T cells undergoing glutamine blockade. Interestingly, we found that DON-treated T cells maintain dramatically increased levels of OXPHOS despite pharmacologic inhibition of mitochondrial transport of long-chain fatty acids, pyruvate, and glutamine with etomoxir, UK0599, and BPTES, respectively (Fig. 3G). Although metabolism of branched-chain amino acids does not appear to be a substantial source of fuel for DON-treated T cells (fig. S4D), ^{13}C-tracing experiments show a dramatic ability of T cells to utilize acetate as a carbon source for the TCA cycle in the setting of glutamine blockade (Fig. 3H). The distinct ability of DON-treated T cells to metabolize acetate as a TCA fuel is supported by the up-regulation of acyl–coenzyme A (CoA) synthetase short-chain family member 1 (ACSS1) and ACSS2 in vitro and in vivo (Fig. 3I and fig. S4E). These findings are in accord with recent reports of the importance of acetate in $CD8^+$ T cell function (*36*–*38*). By contrast, DON-treated MC38 cells down-regulate acetate metabolizing enzymes and show minimal incorporation of ^{13}C-labeled acetate in TCA intermediates (Fig. 3, H and I). As such, acetyl-CoA generation is reciprocally affected in MC38 versus $CD8^+$ T cells during glutamine blockade (Fig. 3J). Notably, down-regulation of ACSS1 and ACSS2 in response to glutamine blockade was observed in multiple cancer cell lines, suggesting a general effect of glutamine blockade on cancer cell metabolism (fig. S4F).

Oxidation of acetyl-CoA through the TCA cycle is a central process in cellular energy metabolism. However, highly proliferative cells, such as activated effector T cells, also use TCA cycle intermediates as building blocks for a range of biomolecules, including nonessential amino acids, lipids, and nucleic acids. Through a process called anaplerosis, these intermediates must be replenished to maintain TCA cycle function. Highly proliferative cells often fuel anaplerosis through glutamine metabolism, or glutaminolysis, which replenishes α-ketoglutarate in the TCA cycle through glutaminase (GLS) activity. Because glutamine contribution to TCA intermediates (as well as other glutamine-requiring processes) is suppressed in DON-treated T cells and MC38 cells (Fig. 4A and fig. S5), we hypothesized that T cells used an alternative anaplerotic source to maintain TCA activity. In this regard, pyruvate carboxylase (PC), which generates oxaloacetate from pyruvate, provides an important alternative anaplerotic pathway by allowing glucose to replenish TCA intermediates (*39*). As such, it has been reported that PC activity is required for cancer cell survival during targeted GLS inhibition (*40*). By performing stable-isotope tracing studies with glucose labeled with ^{13}C at all six carbons, we found that DON-treated MC38 cells show a profound suppression of glucose-derived carbons contributing to the TCA cycle, an effect that is not apparent in DON-treated T cells nor in MC38 cells treated with the GLS inhibitor CB839 (Fig. 4B and fig. S6A). In sharp contrast, DON-treated T cells show considerable flexibility in glucose metabolism, glucose-derived carbons, both as acetyl-CoA as well as oxaloacetate through PC-mediated anaplerosis. As such, the pattern of glucose carbons entering the TCA cycle in DON-treated T cells (and GLS-inhibited MC38 cells)—including increased M+3 isotopologues in citrate, malate, and aspartate and M+5 isotopologue in citrate—is indicative of enhanced PC activity (Fig. 4, C and D, and fig. S6B) (*41*). Unlike DON-treated T cells, this signature is broadly attenuated in DON-treated MC38 cells (Fig. 4D). Notably, PC activity is critically dependent on activation by acetyl-CoA (*42*, *43*). Accordingly, the differential activity of PC (which is expressed equally across each of these cell types and treatments) observed in these cells (fig. S6C) correlates with increased acetyl-CoA levels in DON-treated T cells and decreased acetyl-CoA in DON-treated MC38 cells (Fig. 3J).

The inability of DON-treated MC38 cells to supply glucose-derived metabolites for the TCA cycle (Fig. 4C) or to maintain Warburg physiology (Fig. 3A) is likely due to the suppressive effects of DON on proximal glycolytic metabolism in these cells. Unlike DON-treated T cells and MC38 cells treated with CB839, DON-treated MC38 cancer cells exhibit decreased glucose uptake and diminished levels of early glycolytic intermediates (Fig. 4, E and F, and fig. S6D). We were interested in the mechanism of glycolysis suppression in MC38 cells treated with glutamine antagonism. Unlike T cells, wherein metabolic adaptations to DON treatment (including up-regulated acetate metabolism and PC activity) allow them to maintain the cellular adenosine monophosphate (AMP)/ATP ratio, MC38 cells are unable to compensate energetically, leading to dramatically increased AMP/ATP ratios and robust activation of AMP kinase (AMPK) (Fig. 4, G and H). Furthermore, MC38 cells exposed to glutamine blockade markedly down-regulate c-MYC (Fig. 4H), a phenomenon previously reported in cancer cells in response to glutamine deprivation (*44*). We observed this effect in a range of murine cancer cell lines (fig. S4F). As c-MYC and AMPK are critical regulators of glycolytic flux (*45*–*47*), the down-regulation of

Downloaded from http://science.sciencemag.org/ on December 2, 2019

c-MYC and concurrent activation of AMPK play crucial roles in the suppression of glycolytic pathways in DON-treated MC38 cells. Like DON-treated T cells, however, CB839-treated MC38 cells maintain c-MYC expression and AMP/ATP ratios and show no increased AMPK signaling (fig. S6, C, E, and F).

Given the effects of glutamine antagonism on suppressing proximal glycolysis and Warburg physiology in MC38 cells, we hypothesized that parallel metabolic pathways may also be affected. We were particularly interested in the activity of the pentose phosphate pathway (PPP). The PPP is a branching pathway from glycolysis and is the major cellular source for reduced nicotinamide adenine dinucleotide phosphate (NADPH) production, which is vital for maintenance of redox homeostasis. In addition to their known role in regulating glycolytic flux, c-MYC and AMPK also regulate the activity of the PPP (*46*, *48*, *49*). Indeed, unlike DON-treated T cells and CB839-treated MC38 cells, DON-treated MC38 cells are unable to maintain NADP(H) homeostasis (Fig. 4I and fig. S6G), indicative of a defect in this pathway. This is further supported by measurement of the activity of the PPP rate-limiting enzyme glucose-6-phosphate dehydrogenase (G6PD), which is correspondingly suppressed in DON-treated MC38 cancer cells but increased in DON-treated $CD8^+$ T cells. (Fig. 4J).

Although tracing experiments confirmed blockade of glutamine pathways during DON or JHU083 treatment (fig. S5), we were interested in assessing the specific contribution of key glutamine-utilizing enzymes to the observed metabolic phenotype (fig. S7A). We used short hairpin RNA (shRNA) lentiviral constructs to generate stable knockdowns of glutamine-utilizing enzymes in the MC38 cell line (fig. S7, B and C). We found that knocking down several of these genes individually (e.g., *Gfpt1*, *Ctps1*, and *Asns*) contributed to c-MYC down-regulation, reduced glucose uptake, and suppressed proliferation (fig. S7, D to F). However, when each of these lines was treated with DON, further down-regulation of c-MYC was observed (fig. S7D). That is, although several knockdown cell lines demonstrated suppressed glucose uptake, down-regulated c-MYC, and attenuated proliferation, no single gene completely accounted for the phenotype that was observed through inhibition of all glutamine-utilizing pathways simultaneously (with DON or JHU083) (fig. S7, D and G).

In contrast to cancer cells, activated $CD8^+$ T cells adapt to glutamine blockade through the up-regulation of acetate metabolism, generating high levels of acetyl-CoA to fuel the TCA cycle directly (as a two-carbon source) and indirectly (as an activator of glucose anaplerosis through increased PC activity) (Fig. 4K). In addition to maintaining energy homeostasis (i.e., AMP/ATP ratio) (Fig. 4G), the ability of T cells to utilize acetate to fuel OXPHOS and produce ATP enables them to divert glucose into the PPP and maintain NADP(H) homeostasis (Fig. 4J). Overall, our studies demonstrate that for cancer cells, the interdependence of glycolysis, OXPHOS, and glutamine metabolism lacks plasticity, such that targeting glutamine metabolism leads to wide-ranging metabolic inhibition, disruption of NADP(H) homeostasis, and stymied growth. Conversely, targeting glutamine metabolism in T cells leads to adaptive metabolic reprogramming with enhanced survival, proliferation, and effector function.

From the time of the pioneering work of Otto Warburg, the study of immune cell metabolism has been linked to our understanding of cancer cell metabolism as a paradigm for the behavior of highly proliferative cells (*5*, *50*–*54*). Although both activated T cells and cancer cells use Warburg physiology and glutamine metabolism to fuel notably high proliferative rates (*2*, *54*, *55*), the precise role of Warburg physiology, an inherently less efficient mode of ATP generation than oxidative respiration, has not been established. Recent theories have suggested that highly proliferative cells engage Warburg physiology to maintain the activity of critical pathways such as the PPP that emanate from the glycolytic pathway (*56*). We show here that, although increased glucose uptake may fuel increased PPP activity (as observed in DON-treated T cells), the function of this pathway does not require that glucose is metabolized to lactic acid, that is, Warburg physiology. In other words, the PPP runs in parallel with glycolysis but is not necessarily dependent on it. As such, in addition to establishing glutamine metabolism as a metabolic checkpoint for cancer immunotherapy, our studies suggest fundamental insights into the nature and role of Warburg physiology. Moreover, through the application of glutamine blockade, we demonstrate the possibility of differentially modulating the metabolism of cancer cells and antitumor immune cells by exploiting the differential metabolic plasticity of each cell type.

REFERENCES AND NOTES

1. N. N. Pavlova, C. B. Thompson, *Cell Metab.* **23**, 27–47 (2016).
2. A. J. Wolpaw, C. V. Dang, *Trends Cell Biol.* **28**, 201–212 (2018).
3. B. J. Altman, Z. E. Stine, C. V. Dang, *Nat. Rev. Cancer* **16**, 619–634 (2016).
4. J. Zhang, N. N. Pavlova, C. B. Thompson, *EMBO J.* **36**, 1302–1315 (2017).
5. K. A. Frauwirth *et al.*, *Immunity* **16**, 769–777 (2002).
6. R. Wang *et al.*, *Immunity* **35**, 871–882 (2011).
7. S. Kouidhi, F. Ben Ayed, A. Benammar Elgaaied, *Front. Immunol.* **9**, 353 (2018).
8. S. Kouidhi, M. Z. Noman, C. Kieda, A. B. Elgaaied, S. Chouaib, *Front. Immunol.* **7**, 114 (2016).
9. I. Shevchenko, A. V. Bazhin, *Front. Immunol.* **9**, 1816 (2018).
10. G. C. Prendergast, W. J. Malachowski, A. Mondal, P. Scherle, A. J. Muller, *Int. Rev. Cell Mol. Biol.* **336**, 175–203 (2018).
11. E. Zhao *et al.*, *Nat. Immunol.* **17**, 95–103 (2016).
12. D. B. Rivadeneira, G. M. Delgoffe, *Clin. Cancer Res.* **24**, 2473–2481 (2018).
13. C. H. Chang *et al.*, *Cell* **162**, 1229–1241 (2015).
14. H. Dong, T. N. Bullock, *Front. Immunol.* **5**, 24 (2014).
15. S. M. Davidson *et al.*, *Cell Metab.* **23**, 517–528 (2016).
16. M. Gross *et al.*, *Cancer Res.* **76**, 2329 (2016).
17. C. T. Hensley, A. T. Wasti, R. J. DeBerardinis, *J. Clin. Invest.* **123**, 3678–3684 (2013).
18. L. M. Pinkus, *Methods Enzymol.* **46**, 414–427 (1977).
19. K. M. Lemberg, J. J. Vornov, R. Rais, B. S. Slusher, *Mol. Cancer Ther.* **17**, 1824–1832 (2018).
20. G. Lynch, N. Kemeny, E. Casper, *Am. J. Clin. Oncol.* **5**, 541–543 (1982).
21. R. H. Earhart *et al.*, *Invest. New Drugs* **8**, 113–119 (1990).
22. R. T. Eagan, S. Frytak, W. C. Nichols, E. T. Creagan, J. N. Ingle, *Cancer Treat. Rep.* **66**, 1665–1666 (1982).
23. G. B. Magill *et al.*, *Cancer* **10**, 1138–1150 (1957).
24. S. Krantz *et al.*, *J. Natl. Cancer Inst.* **22**, 433–439 (1959).
25. R. Rais *et al.*, *J. Med. Chem.* **59**, 8621–8633 (2016).
26. J. Fu *et al.*, *J. Pharm. Sci.* **105**, 989–995 (2016).
27. J. Van Gelder *et al.*, *Int. J. Pharm.* **205**, 93–100 (2000).
28. JHU083 is referred to as "compound 3" in (*19*) and "compound 4a" in (*25*).
29. X. Zhu *et al.*, *Neuropsychopharmacology* **44**, 683–694 (2019).
30. M. T. Nedelcovych *et al.*, *J. Neuroimmune Pharmacol.* **14**, 391–400 (2019).
31. B. Murter *et al.*, *Cancer Immunol. Res.* **7**, 244–256 (2019).
32. B. L. Horton, J. B. Williams, A. Cabanov, S. Spranger, T. F. Gajewski, *Cancer Immunol. Res.* **6**, 14–24 (2018).
33. D. A. Chisolm *et al.*, *Immunity* **47**, 251–267.e7 (2017).
34. P. A. Tyrakis *et al.*, *Nature* **540**, 236–241 (2016).
35. G. J. van der Windt *et al.*, *Immunity* **36**, 68–78 (2012).
36. M. L. Balmer *et al.*, *Immunity* **44**, 1312–1324 (2016).
37. J. Qiu *et al.*, *Cell Reports* **27**, 2063–2074.e5 (2019).
38. S. K. Vodnala *et al.*, *Science* **363**, eaau0135 (2019).
39. S. Jitrapakdee, A. Vidal-Puig, J. C. Wallace, *Cell. Mol. Life Sci.* **63**, 843–854 (2006).
39. T. Cheng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8674–8679 (2011).
40. K. Sellers *et al.*, *J. Clin. Invest.* **125**, 687–698 (2015).
41. M. St. Maurice *et al.*, *Science* **317**, 1076–1079 (2007).
42. M. F. Utter, D. B. Keech, *J. Biol. Chem.* **238**, 2603–2608 (1963).
43. F. R. Dejure *et al.*, *EMBO J.* **36**, 1854–1868 (2017).
44. B. Faubert *et al.*, *Cell Metab.* **17**, 113–124 (2013).
45. H. M. Haikala, J. M. Anttila, J. Klefström, *Front. Cell Dev. Biol.* **5**, 38 (2017).
46. J. W. Kim, C. V. Dang, *Cancer Res.* **66**, 8927–8930 (2006).
47. A. B. Kohan, I. Talukdar, C. M. Walsh, L. M. Salati, *Biochem. Biophys. Res. Commun.* **388**, 117–121 (2009).
48. Z. E. Stine, Z. E. Walton, B. J. Altman, A. L. Hsieh, C. V. Dang, *Cancer Discov.* **5**, 1024–1039 (2015).
49. O. Warburg, K. Gawehn, A. W. Geissler, *Z. Naturforsch. B* **13**, 515–516 (1958).
50. A. Bakker, *Klin. Wochenschr.* **6**, 252–254 (1927).
51. V. A. Gerriets, J. C. Rathmell, *Trends Immunol.* **33**, 168–173 (2012).
52. N. J. MacIver, R. D. Michalek, J. C. Rathmell, *Annu. Rev. Immunol.* **31**, 259–283 (2013).
53. E. L. Pearce, M. C. Poffenberger, C. H. Chang, R. G. Jones, *Science* **342**, 1242454 (2013).
54. P. J. Siska *et al.*, *JCI Insight* **2**, e93411 (2017).
55. M. V. Liberti, J. W. Locasale, *Trends Biochem. Sci.* **41**, 211–218 (2016).

ACKNOWLEDGMENTS

We thank W. Xu, E. Thompson, S. Collins, R. Helms, M. Claiborne, and other members of the Powell lab for critical discussion of the manuscript. **Funding:** This work was supported by the National Institutes of Health (R01CA226765 to J.D.P. and R01CA229451 to J.D.P. and B.S.S.) and The Bloomberg–Kimmel Institute for Cancer Immunotherapy. **Author contributions:** R.D.L. and J.M.E. designed and conducted tumor growth and survival experiments. R.D.L., L.Z., J.M.E., I.-M.S., M.-H.O., I.-H.S., M.L.A., I.A.B., C.H.P., J.W., A.T., R.L.B., and J.A. all were involved in performing experiments. E.P. and R.R. performed drug synthesis. B.S.S. supervised drug synthesis and development. R.D.L. and J.D.P. designed and supervised experiments. R.D.L. and J.D.P. wrote the manuscript. **Competing interests:** J.D.P., B.S.S., and R.R. are scientific founders, paid consultants (J.D.P. and B.S.S.), and have

Downloaded from http://science.sciencemag.org/ on December 2, 2019

equity in Dracen Pharmaceuticals. Technology arising in part from the studies described herein were patented by Johns Hopkins University and subsequently licensed to Dracen Pharmaceuticals (JHU083 is currently labeled as DRP-083). R.D.L., J.M.E., B.S.S., R.R., J.A., and J.D.P. are inventors for pending patent application no. PCT/US16/44829, submitted by Johns Hopkins University that covers the use of glutamine analogs, such as JHU083 (DRP-083), for cancer immunotherapy. **Data and materials availability:** The chemical structure of JHU083 [ethyl 2-(2-amino-4-methylpentanamido)-DON] has previously been published as (*25*) and "compound 3" in (*19*) and "prodrug 4a" in (*25*). RNA-seq data have been deposited in the National Institutes of Health (NIH) Gene Expression Omnibus (GEO) repository, accession no. GSE120345.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/1013/suppl/DC1
Materials and Methods
Figs. S1 to S7
References
Data S1

View/request a protocol for this paper from *Bio-protocol*.

23 September 2018; resubmitted 21 July 2019
Accepted 25 October 2019
Published online 7 November 2019
10.1126/science.aav2588

Downloaded from http://science.sciencemag.org/ on December 2, 2019

PLANT SCIENCE

A shared gene drives lateral root development and root nodule symbiosis pathways in *Lotus*

Takashi Soyano[1,2,3]*, Yoshikazu Shimoda[4], Masayoshi Kawaguchi[1,2], Makoto Hayashi[3,4]*

Legumes develop root nodules in symbiosis with nitrogen-fixing rhizobial bacteria. Rhizobia evoke cell division of differentiated cortical cells into root nodule primordia for accommodating bacterial symbionts. In this study, we show that NODULE INCEPTION (NIN), a transcription factor in *Lotus japonicus* that is essential for initiating cortical cell divisions during nodulation, regulates the gene *ASYMMETRIC LEAVES 2-LIKE 18/LATERAL ORGAN BOUNDARIES DOMAIN 16a* (*ASL18/LBD16a*). Orthologs of *ASL18/LBD16a* in nonlegume plants are required for lateral root development. Coexpression of *ASL18a* and the *CCAAT box–binding protein Nuclear Factor-Y* (*NF-Y*) subunits, which are also directly targeted by NIN, partially suppressed the nodulation-defective phenotype of *L. japonicus daphne* mutants, in which cortical expression of *NIN* was attenuated. Our results demonstrate that ASL18a and NF-Y together regulate nodule organogenesis. Thus, a lateral root developmental pathway is incorporated downstream of NIN to drive nodule symbiosis.

Root nodule symbiosis in legumes allows host survival in nitrogen-limiting conditions and partakes in the nitrogen cycle in terrestrial ecosystems. This symbiosis has evolved through the co-option and rearrangement of signaling pathways, following predisposition in a single ancestor of the nitrogen-fixing angiosperm clade (*1*–*3*). It is presumed that nodulation-specific factors, such as *Lotus japonicus* NODULE INCEPTION (NIN), function downstream of early signaling modules (common symbiosis pathway) recruited from arbuscular mycorrhizal symbiosis, which is widely distributed in plants (Fig. 1A) (*4*–*6*). NIN is related to proteins involved in responses to nitrate (*7*). Ectopic expression of *NIN* and its target *Nuclear Factor-Y* (*NF-Y*) subunit genes *NF-YA1* and *NF-YB1* alters development of lateral root primordia and activates cortical cell division for nodule organogenesis (*8*), implying that NIN and its target factors link nodule development programs with lateral root developmental programs.

To identify transcription factors that influence cell division with NF-Y, we searched for genes whose transcription is induced in response to rhizobia among NIN target candidates found by a chromatin immunoprecipitation sequencing (ChIP-seq) analysis (*9*), and we further overexpressed them in *L. japonicus* roots. Only one gene, *ASYMMETRIC LEAVES 2-LIKE 18/LATERAL ORGAN BOUNDARIES DOMAIN 16a* (*ASL18/LBD16a*), stimulated cell division when co-overexpressed with *NF-YA1*.

[1]National Institute for Basic Biology, Nishigonaka 38, Myodaiji, Okazaki, Aichi, 444-8585, Japan. [2]Department of Basic Biology, School of Life Science, SOKENDAI (The Graduate University for Advanced Studies), Nishigonaka 38, Myodaiji, Okazaki, Aichi, 444-8585, Japan. [3]Center for Sustainable Resource Science, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan. [4]Institute of Agrobiological Sciences, National Agriculture and Food Research Organization, Tsukuba, Ibaraki, 305-8634, Japan.
*Corresponding author. Email: soyano@nibb.ac.jp (T.S.); makoto.hayashi@riken.jp (M.H.)

ASL18 genes have been duplicated at least once in an ancestral legume lineage (fig. S1). One or two NIN-binding nucleotide sequences (NBS-S1, or both NBS-S1 and NBS-S2) were found in *ASL18b* and *ASL18a* introns, respectively (Fig. 1B and fig. S2) (*10*). NBS-S1 and its flanking nucleotide sequences were conserved in leguminous *ASL18* introns—particularly in Papilionoideae, with the exception of a few species—but were not observed in nonleguminous orthologs (fig. S1). Hence, the evolution of NBS in *ASL18* intron sequences could have played a key role in recruitment of this lateral root regulator into the nodule signaling pathway in legumes. *NIN* was required for *ASL18a* expression in response to inoculation with *Mesorhizobium loti* (Fig. 1C). Furthermore, dexamethasone treatment of roots expressing NIN fused with a glucocorticoid receptor (NIN-GR) (*8*) increased *ASL18a* expression within 4 hours (Fig. 1D and fig. S3).

Spatial expression patterns of *ASL18* genes were investigated using translational fusion with *β-glucuronidase* (*GUS*) reporter (fig. S4A; *ProASL18a:ASL18a-GUS* and *ProASL18b:ASL18b-GUS*). Both translational fusions were expressed in early lateral root primordia derived from the pericycle (Fig. 2, A and B). Lateral root densities exhibited by *asl18a* knockout plants were lower than those of wild-type plants (Fig. 3A and fig. S5), which was consistent with the general function of ASL18/LBD16 (*11*–*13*). The *ASL18a* promoter was responsible for expression in lateral root primordia and for response to auxin (Fig. 2, C and D, and fig. S6). In the presence of rhizobia, *ProASL18a:ASL18a-GUS* was expressed at infection foci in the root epidermis and nodule primordia formed in the cortex, similar to expression patterns of *NIN* and *NF-Y* subunit genes (Fig. 2, E and F, and fig. S4) (*8*). *ProASL18b:ASL18b-GUS* showed less expression at the basal region of the nodule primordia (Fig. 2G and fig. S4). The *ASL18a* intron was sufficient for conferring expression in nodule primordia and its transcription induced by *NIN*, whereas the *ASL18a* promoter was also active in primordia (Fig. 2, H to J, and figs. S4 and S7). Thus, multiple pathways are connected with *ASL18a* transcription downstream of *NIN*. The number and size of *asl18a* mutant nodules were reduced compared with those of the wild type, when KNO_3 was supplemented for partial inhibition of nodulation (Fig. 3B). This suggested that *ASL18a* was involved in nodule growth. The weakness of the *asl18a* phenotype is probably due to redundancy as observed in *Arabidopsis* (*14*). Indeed, nodule and lateral root development were inhibited when *ASL18a* was expressed as a fusion protein with an artificial repressor domain, SRDX, in hairy roots (fig. S8). *ASL18a* fused with a *35S minimal* promoter partially suppressed the *asl18a* nodule phenotype (fig. S9).

The *asl18a* mutations enhanced nodulation phenotypes of *nf-y* subunit mutants. Nodule

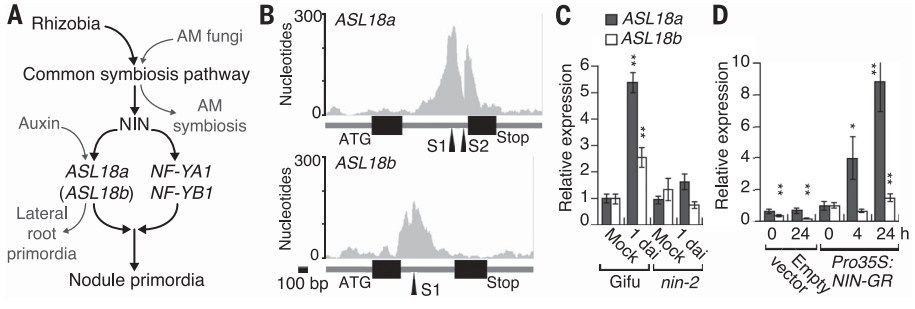


Fig. 1. Rhizobial infection activated ASL18a depending on NIN. (**A**) Schematic of a working hypothesis. AM, arbuscular mycorrhizal. (**B**) Read coverage in *ASL18* genes obtained by ChIP-seq analysis with NIN precipitation. Boxes indicate exons and arrowheads indicate putative NBSs (S1 and S2). bp, base pairs. (**C** and **D**) Quantitative reverse transcription polymerase chain reaction analyses of *ASL18* expression. (C) Wild-type (Gifu B-129) and *nin-2* roots inoculated with *M. loti* ($n > 10$ plants for each biological replicate). dai, days after inoculation. (D) Gifu B-129 roots transformed with either an empty vector or *Pro35S:NIN-GR* were treated with 10 μM dexamethasone ($n > 10$). $^*P < 0.05$, $^{**}P < 0.01$ [one-way analysis of variance (ANOVA) with Tukey's post hoc test] versus wild-type mock (C) and *Pro35S:NIN-GR* 0 hours (D). Data are mean ± SD of three biological repeats.

Downloaded from http://science.sciencemag.org/ on November 29, 2019

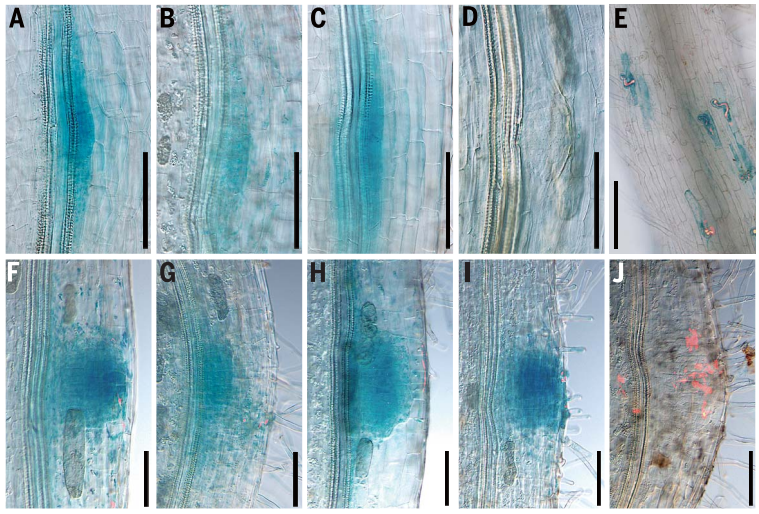


Fig. 2. Spatial expression patterns of ASL18 genes. GUS expression in lateral root primordia (**A** to **D**), infected root hairs (**E**), and nodule primordia (**F** to **J**). [(E) to (J)] Images merged with fluorescence from DsRed-labeled *M. loti*. Roots were transformed with *ProASL18a:ASL18a-GUS* [(A), (E), and (F)]; *ProASL18b:ASL18b-GUS* [(B) and (G)]; *ProASL18a:ASL18a (cDNA)-GUS* [(C) and (H)]; *Pro35Sminimal:ASL18a-GUS* [(D) and (I)]; and *Pro35Sminimal:ASL18a(cDNA)-GUS* (J) (see fig. S4A). Scale bars: 0.2 mm.

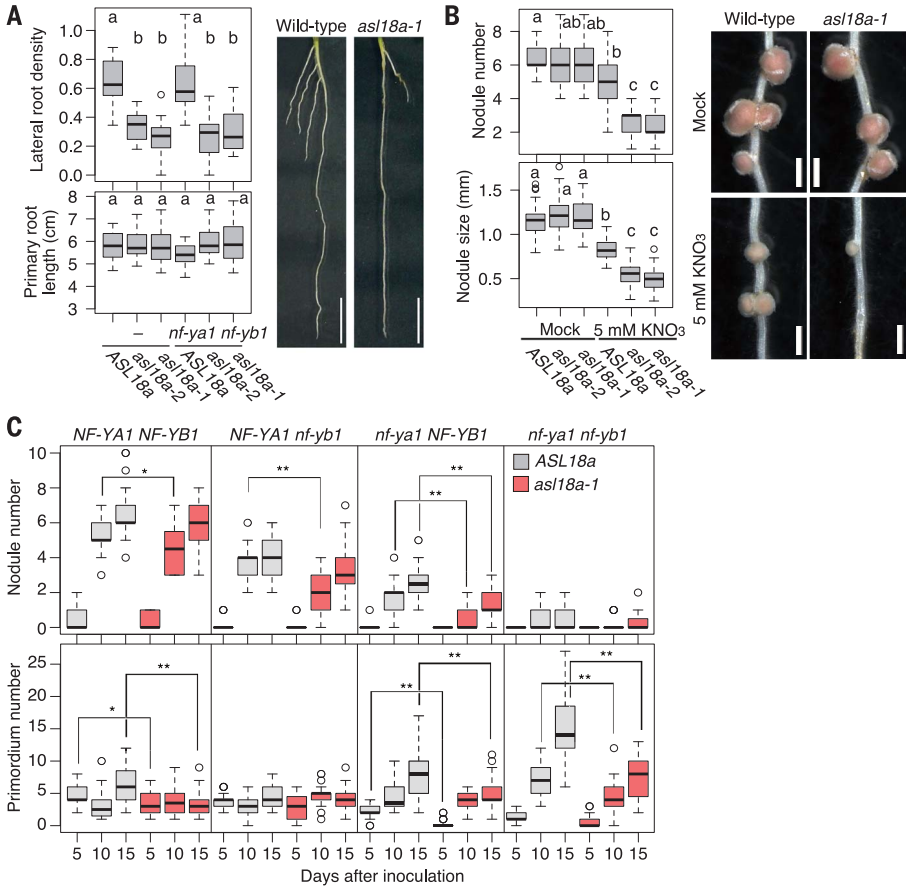


Fig. 3. ASL18a is involved in both lateral root and nodule development. (**A**) Lateral root densities and primary root length (n > 15 plants) of wild type, *als18a*, and *nf-ya1-5 nf-yb1-1* (14-day-old). (**B**) Numbers (n = 15 plants) and diameters (n > 37 nodules) of nodules formed under conditions supplemented with KNO_3 (15 dai). (**C**) Nodule and primordium numbers (n > 20 plants) of multiple mutants between *asl18a-1*, *nf-ya1-5*, and *nf-yb1-1*. One-way ANOVA with Tukey's post hoc test was used. Different letters represent classes with significant difference (P < 0.05). *P < 0.05, **P < 0.01. Scale bars: 1 cm for (A), 1 mm for (B).

development was delayed and nodule number was reduced in *nf-ya1* (*8*, *15*). Nodule development was affected more severely in *nf-ya1 nf-yb1* double mutants (figs. S10 and S11). We attribute enhancement of the nodulation phenotype to functional redundancies with other NF-Y subunits (*16*). Development of nodule primordia in *asl18a nf-ya1 nf-yb1* triple mutants was delayed further, and the numbers of primordia and visible cortical division sites were approximately half of those observed in *nf-y* double mutants (Fig. 3C and fig. S11). This suggested the involvement of *ASL18a* in nodule development from early stages. Further, *ASL18a* seemed to genetically interact with *NF-Y* during nodule development. In contrast, *nf-y* mutations did not influence lateral root densities (Fig. 3A).

NF-Y requires other factors, including pioneer transcription factors, for an activation of its targets (*17*). *Lotus* NF-Y subunits interacted with ASL18 proteins in vitro and in planta (Fig. 4A and fig. S12). *NF-Y* subunits were overexpressed with or without *ASL18a* in roots (fig. S13). Double expression of *NF-YA1* and *NF-YB1* increased lateral root densities to twice those of empty vector controls (Fig. 4C and fig. S14) (*8*). *ASL18a* alone exerted no effect. Coexpression of *ASL18a* with both *NF-Y* subunits increased lateral root densities sixfold

Downloaded from http://science.sciencemag.org/ on November 29, 2019

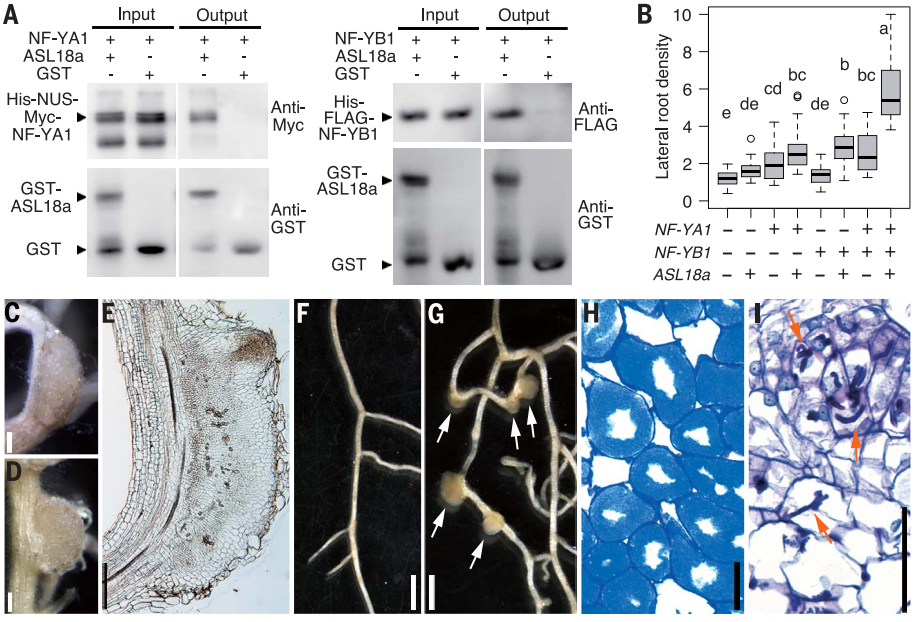


Fig. 4. Interaction of ASL18a with NF-Y stimulated lateral root formation and ectopic cell division. (**A**) Pulldown assay in vitro. (**B**) Lateral root densities of wild-type (MG-20) roots constitutively expressing *ASL18a*, *NF-YA1*, and *NF-YB1* (n = 30 roots). Different letters represent classes with significant difference (P < 0.05, one-way ANOVA with Tukey's post hoc test). (**C** to **E**) Bumps formed in MG-20 (C) and *nin-9* (D) roots coexpressing *ASL18a* with both *NF-Y* subunits. (E) A longitudinal section of (C). (**F** to **I**) Empty vector–transformed *daphne* roots (F) or a construct to constitutively express *ASL18a* with either *NF-Y* subunits [(G) and (I)] or *NF-YA1* (H) were inoculated with *M. loti* (4 weeks). [(H) and (I)] Sections of infected nodules stained with toluidine blue. Arrows in (G) and (I) indicate infected nodules and infection threads, respectively. Scale bars: 0.2 mm for (C) to (E), 2 mm for (F) and (G), 0.1 mm for (H) and (I).

over controls, which was compatible with the protein interactions. Furthermore, roots ectopically expressing *ASL18a* and *NF-YA1* generated bumps (fig. S14 and table S1). Likewise, triple overexpression showed bumps in both wild type and *nin-9* (Fig. 4, C to E). Thus, ASL18a stimulated cell division in collaboration with NF-Y subunits. This effect was not specific to legumes (fig. S15). However, it did not increase nodule numbers (fig. S14). To examine whether ectopic cell division is associated with symbiotic events, we expressed *ASL18a* and *NF-Y* subunits in *daphne* mutants, in which a chromosomal translocation upstream of *NIN* diminishes its expression in root cortex, thereby the mutant roots host infection threads in root epidermis but do not produce nodule primordia (*18*). Expression of *ASL18a* with *NF-Y* subunits led to the appearance of infected nodules on *daphne* roots (Fig. 4, F and G, and table S1). Infection threads penetrated into nodules formed in *daphne* roots, and rhizobia were released into host cells when *ASL18a* was expressed with *NF-YA1* (Fig. 4H). Although the efficiency of production of infected nodules was higher in roots coexpressing *ASL18a* and both *NF-Y* subunits, inhibition of rhizobial release suggested that a correct expression pattern is required for endosymbiosis (Fig. 4I).

The evolutionary origin of root nodules has been previously discussed (*19*, *20*). Here, we show that a gene involved in lateral root development is co-opted for nodule organogenesis downstream of NIN. Replacement of NIN function by ASL18a in collaboration with NF-Y suggested the recruitment of *ASL18a* in organogenesis. An organogenesis-regulating molecular network has evolved through the interplay between the nodulation-specific and lateral root developmental pathways. Our findings thus clarify how legumes acquired the ability to produce root nodules.

REFERENCES AND NOTES

1. J. J. Doyle, *Mol. Plant Microbe Interact.* **24**, 1289–1295 (2011).
2. G. D. A. Werner, W. K. Cornwell, J. I. Sprent, J. Kattge, E. T. Kiers, *Nat. Commun.* **5**, 4087 (2014).
3. M. Griesmann *et al.*, *Science* **361**, eaat1743 (2018).
4. C. Kistner, M. Parniske, *Trends Plant Sci.* **7**, 511–518 (2002).
5. L. Schauser, A. Roussis, J. Stiller, J. Stougaard, *Nature* **402**, 191–195 (1999).
6. S. Singh, K. Katzer, J. Lambert, M. Cerri, M. Parniske, *Cell Host Microbe* **15**, 139–152 (2014).
7. L. Schauser, W. Wieloch, J. Stougaard, *J. Mol. Evol.* **60**, 229–237 (2005).
8. T. Soyano, H. Kouchi, A. Hirota, M. Hayashi, *PLOS Genet.* **9**, e1003352 (2013).
9. T. Soyano, H. Hirakawa, S. Sato, M. Hayashi, M. Kawaguchi, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 14607–14612 (2014).
10. T. Soyano, Y. Shimoda, M. Hayashi, *Plant Cell Physiol.* **56**, 368–376 (2015).
11. Y. Inukai *et al.*, *Plant Cell* **17**, 1387–1396 (2005).
12. Y. Okushima, H. Fukaki, M. Onoda, A. Theologis, M. Tasaka, *Plant Cell* **19**, 118–130 (2007).
13. G. Taramino *et al.*, *Plant J.* **50**, 649–659 (2007).
14. T. Goh, S. Joi, T. Mimura, H. Fukaki, *Development* **139**, 883–893 (2012).
15. M. S. Hossain *et al.*, *Mol. Plant Microbe Interact.* **29**, 950–964 (2016).
16. M. Baudin *et al.*, *Plant Physiol.* **169**, 2761–2773 (2015).
17. A. J. Oldfield *et al.*, *Mol. Cell* **55**, 708–722 (2014).
18. E. Yoro *et al.*, *Plant Physiol.* **165**, 747–758 (2014).
19. A. M. Hirsch, T. A. Larue, J. Doyle, *Crit. Rev. Plant Sci.* **16**, 361–392 (1997).
20. K. Magne *et al.*, *Plant Physiol.* **178**, 295–316 (2018).

ACKNOWLEDGMENTS

We thank the National Institute for Basic Biology's Functional Genomics Facility, Model Plant Research Facility, and Spectrography and Bioimaging Facility for technical support and A. Tokairin and A. Oda for technical assistance. **Funding:** This work was funded by a RIKEN Incentive Research Grant to T.S., and by KAKENHI (16K08149) and the Next Generation World-Leading Researchers grant (GS029) from JSPS to T.S. and M.H., respectively. **Author contributions:** T.S. and M.H. designed the study. T.S. performed experiments, with support from Y.S., M.K., and M.H., and analyzed data. T.S. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/1021/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S3
References (*21*–*30*)

View/request a protocol for this paper from *Bio-protocol*.

4 March 2019; accepted 18 October 2019
10.1126/science.aax2153

Downloaded from http://science.sciencemag.org/ on November 29, 2019

PROTEIN DESIGN

Computational design of a modular protein sense-response system

Anum A. Glasgow[1]*, Yao-Ming Huang[1]*†, Daniel J. Mandell[1,2]*‡, Michael Thompson[1], Ryan Ritterson[1]§, Amanda L. Loshbaugh[1,3], Jenna Pellegrino[1,3], Cody Krivacic[1,4], Roland A. Pache[1]‖, Kyle A. Barlow[1,2]¶, Noah Ollikainen[1,2], Deborah Jeon[1], Mark J. S. Kelly[5], James S. Fraser[1,3,6], Tanja Kortemme[1,2,3,4,6,7]#

Sensing and responding to signals is a fundamental ability of living systems, but despite substantial progress in the computational design of new protein structures, there is no general approach for engineering arbitrary new protein sensors. Here, we describe a generalizable computational strategy for designing sensor-actuator proteins by building binding sites de novo into heterodimeric protein-protein interfaces and coupling ligand sensing to modular actuation through split reporters. Using this approach, we designed protein sensors that respond to farnesyl pyrophosphate, a metabolic intermediate in the production of valuable compounds. The sensors are functional in vitro and in cells, and the crystal structure of the engineered binding site closely matches the design model. Our computational design strategy opens broad avenues to link biological outputs to new signals.

In the past two decades, computational protein design has created diverse new protein structures spanning helical (*1*–*5*), alpha-beta (*6*–*8*), and beta-sheet (*9*, *10*) folds. By contrast, our ability to computationally design arbitrary protein functions de novo lags far behind, with relatively few examples that often require screening of many variants (*11*, *12*). One unsolved challenge is the de novo design of small-molecule sensor-actuators in which ligand binding by a protein directly controls changes in downstream functions, a key aspect of cellular signal transduction (*13*).

Sensing and responding to a small-molecule signal requires both recognition of the target and linking target recognition to an output response. Exciting progress has been made with the design of proteins recognizing new ligands (*10*, *11*, *14*–*16*). A general solution to the second problem, coupling ligand recognition to diverse output responses, has remained challenging. Existing approaches have used a ligand that fluoresces upon binding (*10*), engineered the sensor components to be unstable and hence inactive in the absence of the ligand (*14*, *17*), or repurposed an allosteric transcription factor (*18*). Each of these strategies places constraints on the input signals or output responses that can be used.

Here, we describe a computational strategy to engineer protein complexes that can sense a small molecule and respond directly using different biological outputs, creating modular sensor-actuator systems. Unlike previous work (*10*, *11*, *14*, *15*) that reengineered existing binding sites or placed ligands into preformed cavities, we build small-molecule binding sites de novo into heterodimeric protein-protein interfaces to create new and programmable chemically induced dimerization systems (CIDs). This strategy is inspired by naturally occurring and reengineered CID systems (*19*) that have been widely used but are limited to a small number of existing or similar input molecules. We aimed to design synthetic CIDs that could similarly link binding of a small molecule to modular cellular responses through genetically encoded fusions of each sensor monomer to a split reporter (Fig. 1A) but respond to new, user-defined inputs.

To demonstrate this strategy, we chose farnesyl pyrophosphate (FPP) as the target ligand. FPP is an attractive target because it is a toxic intermediate in a commonly engineered biosynthesis pathway for the production of valuable terpenoid compounds (*20*). Sensors for FPP could be used, for example, to optimize pathway enzymes or, when linked to appropriate outputs, to regulate pathway gene expression in response to changes in metabolite concentrations (*21*). Our computational strategy (Fig. 1B and supplementary materials and methods) proceeds in four main steps: (i) defining the geometries of minimal FPP-binding sites composed of three to four side chains (termed "motif residues") that form key interactions with the target ligand; (ii) modeling these geometries into a dataset of heterodimeric protein-protein interfaces (termed "scaffolds") and computationally screening for coarsely compatible scaffolds (*22*); (iii) optimizing the binding sites in these scaffolds using flexible backbone design methods previously used to predict ligand-binding specificities (*23*–*25*) but not tested in the de novo design of binding sites ("reshaping"); and (iv) ranking individual designs for experimental testing according to several design metrics, including ligand-binding energy predicted using the Rosetta force field (*26*) and ligand burial.

Starting with five FPP-binding-site geometries and up to 3462 heterodimeric scaffolds, we selected the highest-ranked designs across three engineered scaffolds for experimental testing (Fig. 1B and supplementary materials and methods): (i) the FKBP-FRB complex originally responsive to rapamycin (*27*) (one design), (ii) a complex of the bacterial proteins RapF and ComA (*28*) (four designs), and (iii) an engineered complex of maltose-binding protein (MBP) and an ankyrin repeat (AR) protein (*29*) (four designs) (Fig. 2A, table S1, and fig. S1). The ligand was placed into the rapamycin site in FKBP-FRB, but the binding sites in the other two complexes were modeled de novo.

To test these computationally designed FPP sensors, we genetically fused the engineered sensor proteins to a well-studied split reporter, the enzyme murine dihydrofolate reductase [mDHFR (*30*); Fig. 2B and supplementary materials, appendix 1] and expressed the fusion constructs in *Escherichia coli*. We reasoned that functional sensors should exhibit increased growth through FPP-driven dimerization of the sensor proteins and resulting complementation of functional mDHFR under conditions in which endogenous *E. coli* DHFR was specifically inhibited by trimethoprim. Because FPP does not efficiently enter *E. coli*, we added its metabolic precursor, mevalonate, to the growth medium and coexpressed an engineered pathway of five enzymes (*20*) (Fig. 2B, pMBIS) to produce FPP from mevalonate in the cells. We then monitored sensor function as change in growth in the presence or absence of mevalonate under otherwise identical conditions (Fig. 2B and supplementary materials and methods). In the following, we denote designs by their scaffold (S1, S2, and S3), design generation (1, 2, and 3), and successive letter (A, B, C, etc.) (table S1 and fig. S1).

[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA. [2]Bioinformatics Graduate Program, University of California San Francisco, San Francisco, CA, USA. [3]Biophysics Graduate Program, University of California San Francisco, San Francisco, CA, USA. [4]UC Berkeley–UCSF Graduate Program in Bioengineering, University of California San Francisco, San Francisco, CA, USA. [5]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA. [6]Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA, USA. [7]Chan Zuckerberg Biohub, San Francisco, CA, USA.
*These authors contributed equally to this work. †Present address: 23andMe, Inc., Therapeutics, South San Francisco, CA, USA. ‡Present address: GRO Biosciences, Boston, MA, USA. §Present address: Gryphon Scientific, LLC, Takoma Park, MD, USA. ‖Present address: Novozymes A/S, Biologiens Vej 2, 2800 Kgs. Lyngby, Denmark. ¶Present address: Adimab LLC, Palo Alto, CA, USA.
#Corresponding author. Email: tanjakortemme@gmail.com

Downloaded from http://science.sciencemag.org/ on November 29, 2019

Although seven of the nine selected designs showed only a small signal (S2-1A, B, C, D, S3-1A, B) or no signal (S1-1A), two designs (S3-1C, D) displayed a robust signal response to FPP (Fig. 2C and fig. S2). Both designs resulted from the AR-MBP scaffold (Fig. 2A, S3). For this scaffold, we also generated two libraries: library 1 based on our ensemble design predictions (Fig. 2A and table S2) and library 2 using error-prone polymerase chain reaction (epPCR) starting from design S3-1C. After an initial growth-based selection and subsequent plate-based screens in the presence and absence of FPP (supplementary materials and methods, fig. S3), we identified 36 hits from which we confirmed 27 FPP-responsive sequences by individual growth assays (figs. S4 and S5). One of the most active designs identified across both libraries (S3-2A) was a variant of design S3-1C with two additional mutations distal from the designed FPP-binding site introduced by epPCR. This variant displayed essentially equal activity as the original S3-1C design when tested under identical conditions (Fig. 2C, table S1, and fig. S2). These results show that library screening or epPCR was not necessary to identify functional sensors; instead, we obtained functional sensors directly by computational design. However, library 1 provided additional active sequences from the sequence tolerance predicted in the ensemble design simulations (Fig. 2A, table S2, and fig. S4).

To further characterize the identified best design, S3-2A (table S1), we performed single-site saturation mutagenesis at (SSM) 11 positions (table S3). We tested the resulting mutants with the growth-based split mDHFR reporter in the presence and absence of FPP under more stringent conditions by increasing the trimethoprim concentration (fig. S6). Whereas at most positions the originally designed amino acid (Fig. 2A, design S3-1C) appeared to be optimal under these conditions, we saw considerable improvements for mutations at two positions, R194A (Fig. 2A, design S3-2B) and R194A/L85G (Fig. 2A, design S3-2C). These two designs displayed increasing responses to mevalonate at higher trimethoprim concentrations (Fig. 2D). For the most active design, S3-2C, we confirmed that the sensor signal was dependent on the expression of the sensor proteins [Fig. 2E, -IPTG (isopropyl-β-D-thiogalactopyranoside)] and the metabolic pathway that converts added mevalonate to FPP (Fig. 2E, pMBIS). To test for specificity for FPP, we confirmed that the sensor signal was absent when preventing the accumulation of FPP either by inactivating the fifth enzyme in the pathway by a single point mutation (Fig. 2, B and E, *ispA* R116A) or by adding a sixth enzyme that converts FPP to amorphadiene (Fig. 2, B and E, pB5K). To test whether the sensor signal was dependent on the original four motif side chains, we mutated each individually to alanine and observed decreased sensitivity to the presence of mevalonate for three of the four motif side chains (Fig. 2F, L89, F133, and R145, but not W114). Finally, we tested whether the sensor signal of design

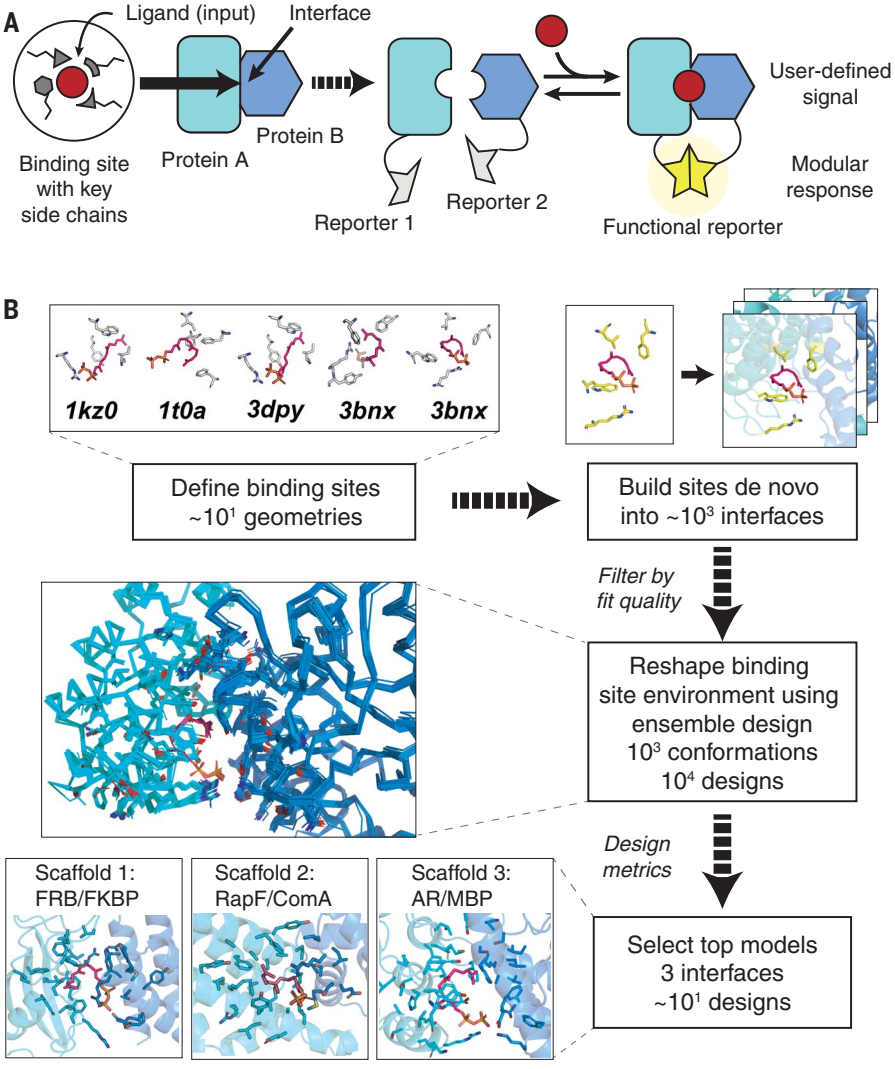


Fig. 1. Computational design. (**A**) Diagram of the design strategy. A small-molecule binding site is built de novo into protein-protein interfaces (left) to create synthetic CIDs (right). Linking the designed sensor proteins to split reporters yields modular CID systems in which different reporter outputs can be coupled to user-defined small-molecule input signals. (**B**) Steps in the design of a synthetic CID system sensing FPP. Top: Binding-site geometries with key interacting side chains selected from FPP-binding proteins (PDB codes indicated) are computationally modeled into a large number of protein-protein interfaces. Middle: Binding sites with feasible geometries are reshaped and optimized by flexible backbone design; shown is a conformational ensemble for a single sequence. Bottom: Top designs from three different scaffolds selected for experimental tests (Fig. 2).

Downloaded from http://science.sciencemag.org/ on November 29, 2019

S3-2C was dependent on the concentration of FPP by titrating the extracellular concentration of the mevalonate precursor (Fig. 2G). Although the sensor signal initially increased with increasing mevalonate concentrations, as expected, the signal decreased at the highest mevalonate concentration tested. This behavior likely arises from FPP-mediated toxicity previously observed at this mevalonate concentration using the same FPP-biosynthesis pathway (*20*). We confirmed a consistent dependency of the sensor signal both on sensor expression (by varying the concentration of the inducer, IPTG) and on mevalonate concentration in the growth medium for seven of our designs (fig. S7, S3-1A, B, C, D, S3-2A, B, C). These results confirm that sensor function in *E. coli* is specific to FPP produced by an engineered pathway, dependent on key residues in the de novo designed binding site, dose dependent in *E. coli*, and sensitive to FPP concentrations in a relevant range (i.e., below the toxicity level).

To confirm biochemically that FPP increases the binding affinity of the AR-MBP complex as designed, we purified the designed AR and MBP proteins without attached reporters [see the supplementary materials and methods; these constructs contained several previously published mutations to stabilize AR (*31*), which when tested in the split mDHFR reporter assay led to active sensor S3-2D (table S1, fig. S8, and supplementary materials, appendix 2)]. We determined the apparent binding affinity of the designed AR and MBP proteins comprising the S3-2D sensor (Fig. 3A, table S1, and fig. S1) in the absence and presence of 200 μM FPP using biolayer interferometry (Fig. 3B, fig. S9, and supplementary materials and methods). The presence of FPP led to a >100-fold stabilization of the interaction between the AR and MBP proteins comprising sensor S3-2D [dissociation constant (K_D) from >200 to 2.1 μM, Fig. 3C; for comparison, the original AR-MBP scaffold had a K_D of 4.4 nM (*29*)]. Binding of FPP to the designed AR component of S3-2D alone was weak, and binding of FPP to the designed MBP component of S3-2D alone was not detectable (Fig. 3D). These results confirm in vitro with purified components that design S3-2D functions as a CID system responding to FPP.

To determine whether FPP is recognized in the de novo engineered binding site as predicted by the design model, we determined a 2.2-Å resolution crystal structure of the ternary complex of FPP bound in the engineered AR-MBP interface (supplementary materials and methods; table S4). The crystal structure of the bound complex is in excellent overall agreement with the design model (Fig. 4, A to C). Despite twinning in the crystals, examining unbiased omit maps allowed modeling of unexplained density in the engineered binding site as FPP (Fig. 4B and fig. S10) and confirmed the side-chain conformations in the designed binding pocket (Fig. 4, C and D). Overall, in a 10-Å shell around FPP in the binding pocket, the Cα root mean square

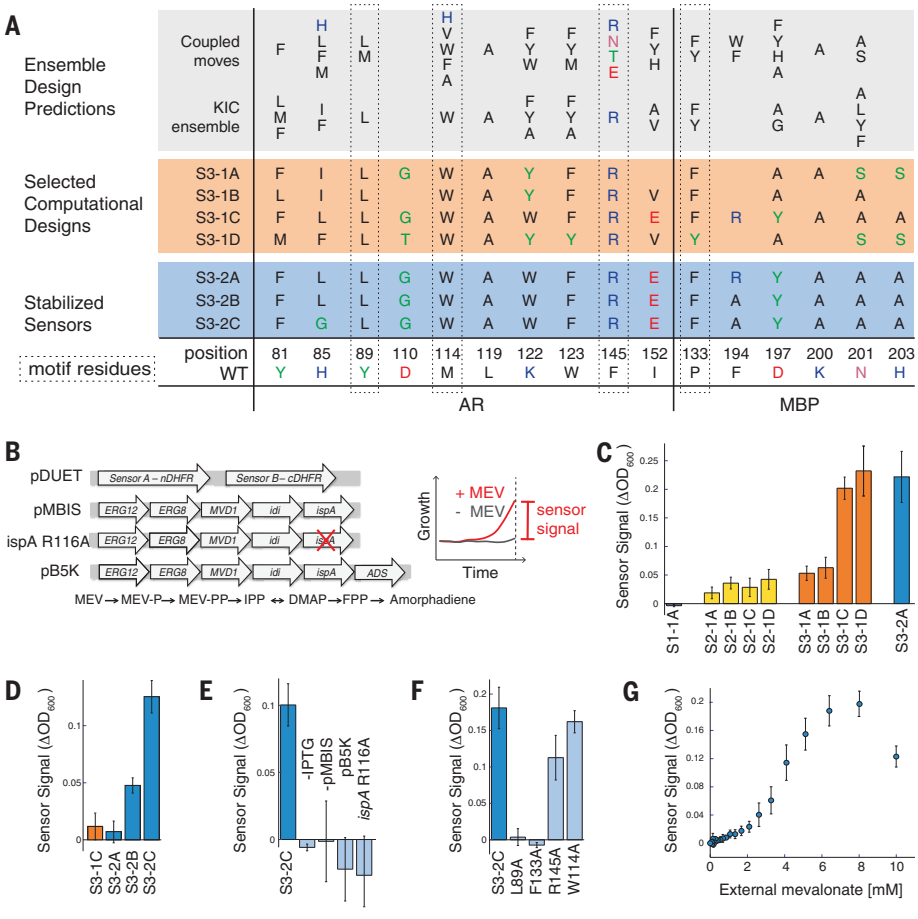


Fig. 2. Sensor function in bacteria. (**A**) Designed sequences at key positions for scaffold 3. Gray shading indicates preferred residues from flexible backbone reshaping by kinematic closure [KIC (*23*, *24*)] or coupled moves (*25*). Orange shading indicates individual computational designs selected based on ligand burial (S3-1A), consensus (S3-1B), optimized ligand packing (S3-1C), and predicted ligand-binding score (S3-1D). Blue shading indicates sensors stabilized by two additional mutations from SSM (S3-2B and S3-2C also contained two mutations from epPCR that were not in the designed FPP-binding site; fig. S1). (**B**) Constructs (left; for details, see supplementary materials, appendix 1) used in the split mDHFR reporter assay (right). pDUET, sensor proteins linked to the split mDHFR reporter; pMBIS, engineered pathway of five enzymes to convert mevalonate (MEV) into FPP (*20*); *ispA* R116A, pMBIS containing R116A mutation in *ispA* that reduces catalytic activity 13-fold (*37*); pB5K, pMBIS with amorphadiene synthase (ADS) (*20*). Sensor signal is quantified as the change in optical density at wavelength 600 nm (OD_{600}) in the presence and absence of mevalonate. (**C**) Sensor signal in the split mDHFR assay for computational designs based on scaffold 1 (FKBP-FRB12, purple bar), scaffold 2 (RapF-ComA, yellow bars), and scaffold 3 (AR-MBP, orange bars). Sensor S3-2A (identified from library 2 with two mutations distal from the designed FPP-binding site; table S1), is shown for comparison (blue bar). (**D**) Improvement of sensor signal by stability-enhancing mutations in S3-2B and S3-2C at increased stringency [trimethoprim concentration 6 μM versus 1 μM in (C)]. (**E**) Dependence of S3-2C sensor signal on sensor expression (-IPTG) and FPP production (-pMBIS, pB5K, *ispA* R116A). (**F**) Dependence of the S3-2C sensor signal on motif residues. (**G**) Dependence of the S3-2C sensor signal on the concentration of the FPP precursor mevalonate added extracellularly. Error bars indicate standard deviation from at least four biological replicates and eight technical replicates for each biological replicate.

Downloaded from http://science.sciencemag.org/ on November 29, 2019

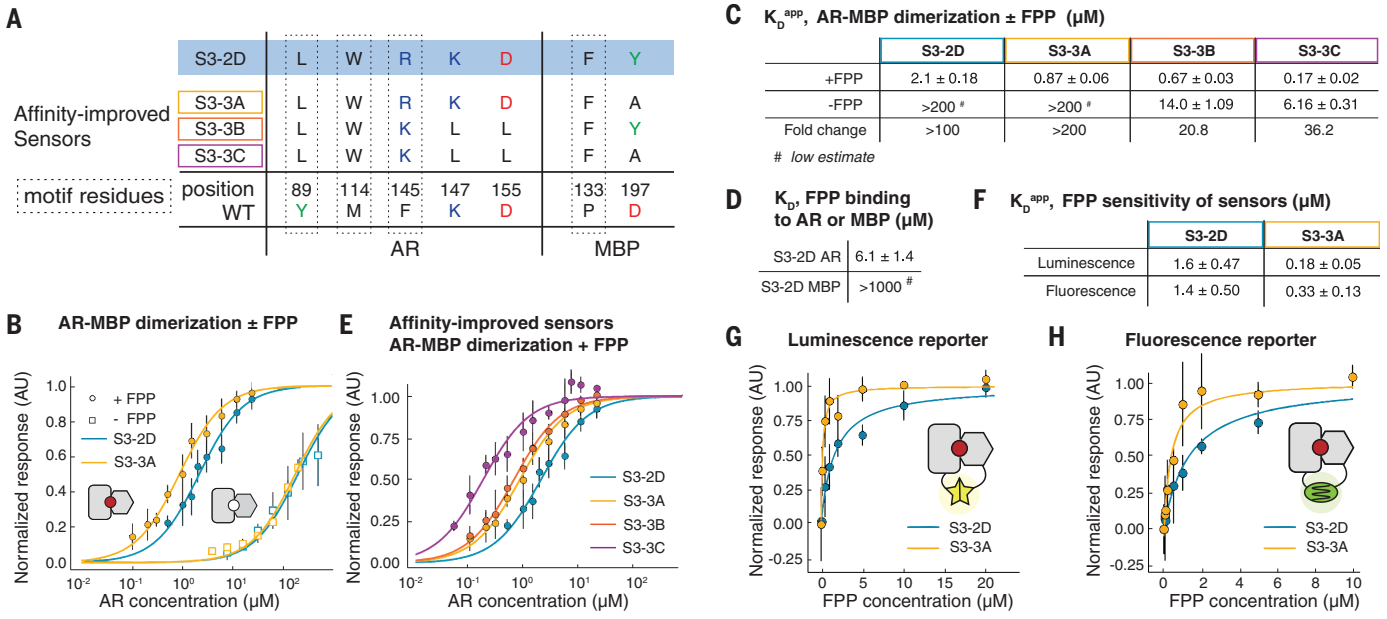


Fig. 3. In vitro sensor characterization and output modularity. (**A**) Sequence changes in sensor constructs tested in vitro. Motif residues are also shown. The starting construct, S3-2D (blue), is identical to S3-2C in the engineered FPP-binding site but contains additional previously published stabilizing mutations in AR (*31*) (shown in table S1). (**B** to **H**) In vitro binding measurements from biolayer interferometry (BLI) using purified protein (B to E) or FPP titrations with sensors expressed using in vitro TxTl (F to H). (B) Apparent AR interaction with immobilized MBP in the presence (closed circles) or absence (open squares) of 200 μM FPP, comparing designs S3-2D (blue) and S3-3A containing the Y197A mutation (orange). (C) Summary of BLI results for apparent AR-MBP dimerization with and without FPP. (D) Summary of BLI results for FPP binding to the individual designed AR and MBP proteins comprising design S3-2D (table S1). (E) Apparent AR interaction with immobilized MBP for a computationally designed variant using the S3-2D crystal structure as the input with (purple, S3-3C) or without (red, S3-3B) the Y197A mutation. (F) Apparent affinity of the S3-2D and S3-3A sensors for FPP using luminescent or fluorescent reporters in TxTl experiments. (G and H) FPP titrations in TxTl using the luminescent reporter (G) or the fluorescent reporter (H). Error bars indicate standard deviations for $n \geq 3$.

deviation (rmsd) between the model and the structure was 0.53 Å and the all-heavy-atom rmsd was 1.13 Å. Although crystals formed only in the presence of FPP, only one of the two complexes in the asymmetric unit contained FPP in the binding site (fig. S11). This behavior allowed us to compare *apo* and *holo* states of the complex. Most of the designed side chains are in identical conformations in the FPP-bound *holo* and FPP-minus *apo* states (Fig. 4E), suggesting favorable preorganization of the designed binding site. An exception is W114 on AR, which is partly disordered in the *apo* state (fig. S11), providing a potential explanation for why a W114A mutation is less detrimental for sensor activity (Fig. 2F) than expected based on the observed packing interactions between W114 and FPP in the *holo* state. A second slight deviation between the model and the crystal structure appeared to be caused by potential steric clashes of the engineered Y197 on MBP with the modeled FPP conformer, which led to rearrangements in the FPP structure and a rotamer change in designed residue F133 on MBP (Fig. 4D). Many of the original models from computational design favored a smaller alanine side chain at this position (Fig. 2A). These observations led to the prediction that a Y197A mutation might stabilize the ternary complex and, indeed, design S3-3A containing the Y197A mutation showed an increased (>200 fold) stabilization of the complex with FPP, with an apparent dissociation constant of 870 ± 60 nM for the designed AR and MBP proteins comprising sensor S3-3A in the presence of 200 μM FPP (Fig. 3, B and C). We also confirmed that design S3-3A (table S1) is active in *E. coli* (fig. S12). To further improve the design based on the crystal structure of design S3-2D, we used an additional round of flexible backbone design using the Rosetta coupled-moves method (*25*) starting from the FPP-bound crystal structure. These simulations suggested three additional mutations leading to design S3-3B: R145K, K147L, and D155L (Fig. 3A). These mutations, when combined with the Y197A mutation (design S3-3C), enhanced the apparent binding affinity of the designed AR and MBP proteins comprising sensor S3-3C in the presence of 200 μM FPP to 170 ± 20 nM (Fig. 3, C and E), which is within 40-fold of the original scaffold AR-MBP interaction affinity (*29*), but also strengthened the binding affinity of the protein–protein dimer in the absence of FPP to 6.2 ± 0.3 μM (fig. S13). The design simulations optimized sequences for stability of the ternary complex without also destabilizing the dimer in the absence of the small molecule. Methods integrating negative design (*32*) could be incorporated to improve the dynamic range of the system (supplementary text).

A key advantage of our CID design strategy is the ability to link an engineered sensor, the input of which is specific to a user-defined small-molecule signal, to a modular output that can in principle be chosen from many available split reporters (Fig. 1A). To test this concept, we linked the engineered CID sensors S3-2D and S3-3A to two additional outputs: a dimerization-dependent fluorescent protein (*33*) and a split luciferase (*34*) (Fig. 3, G and H, and supplementary materials, appendix 3). We tested input-output responses using an in vitro transcription-translation system (TxTl) (*35*) in which FPP can be added at defined concentrations to the assay extract, in contrast to the cell-based split mDHFR assay. The TxTl assay revealed a nanomolar FPP sensitivity (K_D^{app}) for our best sensor, S3-3A (Fig. 3F), that was essentially identical for both reporters (180 ± 50 and 330 ± 130 nM by luminescence and fluorescence detection, respectively; Fig. 3, G and H, and fig. S14), and additionally confirms the improvements in design S3-3A containing the Y197A mutation over design S3-2D (the K_D^{app} for S3-2D was 1.6 ± 0.5 and 1.4 ± 0.5 μM for the luminescence and fluorescence reporters, respectively; Fig. 3, F to H). These results show that our CID

Downloaded from http://science.sciencemag.org/ on November 29, 2019

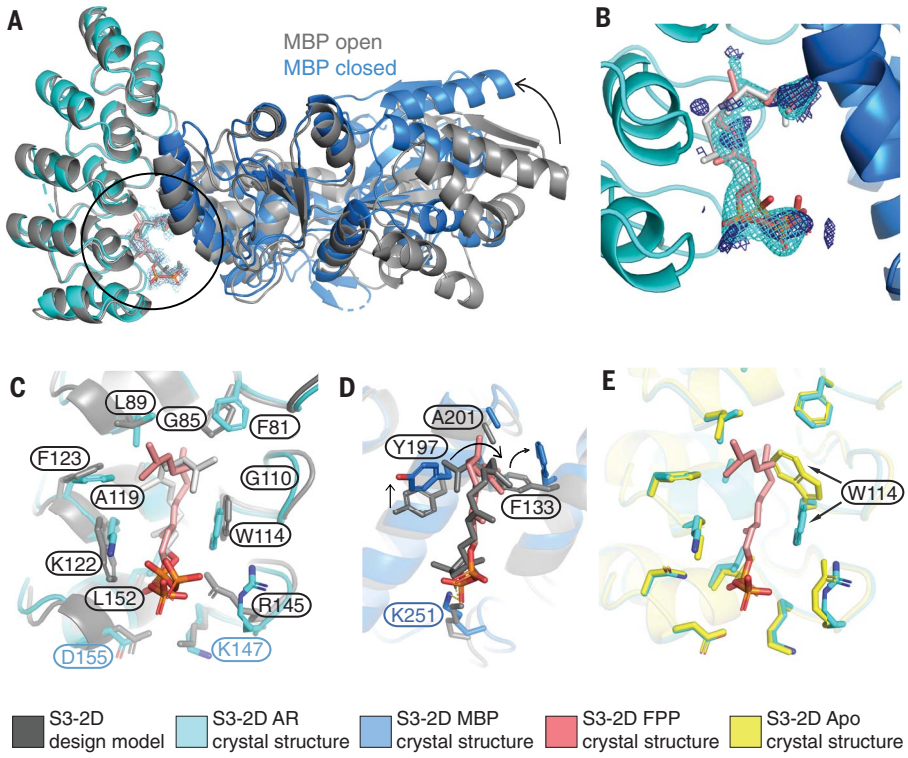


Fig. 4. The S3-2D crystal structure closely matches the computational design model. (**A**) Overlay of the design model (gray) with the crystal structure (designed AR, cyan; designed MBP, blue; FPP, pink) showing FPP binding in the computationally designed binding site at the AR-MBP interface (circle). The design crystallized in the closed MBP conformation, whereas MBP was in the open conformation in the original scaffold on which the model was based, leading to a difference in rigid-body orientation (arrow) of one lobe of MBP distal to the FPP-binding site. (**B**) FPP overlaid with $2mF_o - DF_c$ electron density map (1.2σ, cyan) and ligand $2mF_o - DF_c$ omit map (1.0σ, dark blue). Strong density peaks were present in both maps for the phosphates and several anchoring hydrophobic groups. (**C**) Open-book representation of the FPP-binding site on AR showing close match of designed side-chain conformations to the crystal structure. (**D**) Open-book representation of the FPP-binding site on MBP indicating a clash between the position of MBP Y197 in the crystal structure (blue) and the designed FPP orientation in the model (gray), causing slight rearrangements of FPP and F133 (arrows). (**E**) Alignment of the *holo* (cyan) and *apo* (yellow) structures of S3-2D showing overall agreement with the exception of the side chain of W114 (arrows). In (C) to (E), residues are labeled black when designed and cyan or blue when present in the original scaffold complex.

sensor design strategy is compatible with modular outputs.

The most critical feature of our approach is the ability to computationally design small-molecule binding sites de novo into protein–protein interfaces. A previous computational analysis suggested that the appearance of pockets around artificially generated protein–protein interfaces may be an intrinsic geometric feature of protein structure (*36*), lending support to the idea that our approach is extensible to many other small molecules and interfaces. The design method presented here thus introduces a generalizable way to create synthetic sensing systems with different outputs that can be used in diverse biological contexts to respond to user-specified molecular signals.

REFERENCES AND NOTES

1. P. S. Huang *et al.*, *Science* **346**, 481–485 (2014).
2. T. M. Jacobs *et al.*, *Science* **352**, 687–690 (2016).
3. A. R. Thomson *et al.*, *Science* **346**, 485–488 (2014).
4. R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. DeGrado, *Acc. Chem. Res.* **33**, 745–754 (2000).
5. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **282**, 1462–1467 (1998).
6. G. J. Rocklin *et al.*, *Science* **357**, 168–175 (2017).
7. N. Koga *et al.*, *Nature* **491**, 222–227 (2012).
8. P. S. Huang *et al.*, *Nat. Chem. Biol.* **12**, 29–34 (2016).
9. E. Marcos *et al.*, *Nat. Struct. Mol. Biol.* **25**, 1028–1034 (2018).
10. J. Dou *et al.*, *Nature* **561**, 485–491 (2018).
11. H. Lechner, N. Ferruz, B. Höcker, *Curr. Opin. Chem. Biol.* **47**, 67–76 (2018).
12. J. Dou *et al.*, *Protein Sci.* **26**, 2426–2437 (2017).
13. R. M. Gordley, L. J. Bugaj, W. A. Lim, *Curr. Opin. Struct. Biol.* **39**, 106–114 (2016).
14. M. J. Bick *et al.*, *eLife* **6**, e28909 (2017).
15. C. E. Tinberg *et al.*, *Nature* **501**, 212–216 (2013).
16. N. F. Polizzi *et al.*, *Nat. Chem.* **9**, 1157–1164 (2017).
17. J. Feng *et al.*, *eLife* **4**, e10606 (2015).
18. N. D. Taylor *et al.*, *Nat. Methods* **13**, 177–183 (2016).
19. D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, *Science* **262**, 1019–1024 (1993).
20. V. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman, J. D. Keasling, *Nat. Biotechnol.* **21**, 796–802 (2003).
21. F. Zhang, J. Keasling, *Trends Microbiol.* **19**, 323–329 (2011).
22. A. Zanghellini *et al.*, *Protein Sci.* **15**, 2785–2794 (2006).
23. M. Babor, D. J. Mandell, T. Kortemme, *Protein Sci.* **20**, 1082–1089 (2011).
24. D. J. Mandell, E. A. Coutsias, T. Kortemme, *Nat. Methods* **6**, 551–552 (2009).
25. N. Ollikainen, R. M. de Jong, T. Kortemme, *PLOS Comput. Biol.* **11**, e1004335 (2015).
26. R. F. Alford *et al.*, *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
27. J. Choi, J. Chen, S. L. Schreiber, J. Clardy, *Science* **273**, 239–242 (1996).
28. M. D. Baker, M. B. Neiditch, *PLOS Biol.* **9**, e1001226 (2011).
29. H. K. Binz *et al.*, *Nat. Biotechnol.* **22**, 575–582 (2004).
30. J. N. Pelletier, F. X. Campbell-Valois, S. W. Michnick, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12141–12146 (1998).
31. M. A. Kramer, S. K. Wetzel, A. Plückthun, P. R. Mittl, M. G. Grütter, *J. Mol. Biol.* **404**, 381–391 (2010).
32. J. A. Davey, R. A. Chica, *Protein Sci.* **21**, 1241–1252 (2012).
33. S. C. Alford, Y. Ding, T. Simmen, R. E. Campbell, *ACS Synth. Biol.* **1**, 569–575 (2012).
34. A. S. Dixon *et al.*, *ACS Chem. Biol.* **11**, 400–408 (2016).
35. R. Marshall, V. Noireaux, *Methods Mol. Biol.* **1772**, 61–93 (2018).
36. M. Gao, J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3784–3789 (2012).
37. L. Song, C. D. Poulter, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3044–3048 (1994).

ACKNOWLEDGMENTS

We thank J. Keasling and F. Zhang for advice on FPP production in microbes and pathway constructs; E. de los Santos, Z. Sun, V. Noireux, and R. Murray for TxTl advice and reagents; S. Alford and R. Campbell for dimerization-dependent fluorescent protein constructs; A. Anand, V. Ruiz, B. Adler, and A. Maxwell for contributions to computational design and characterization; S. O'Connor for developing a database for design models; and members of the Kortemme lab for discussion. **Funding:** This work was supported by a grant from the National Institutes of Health (NIH) (R01-GM110089) and a W.M.F. Keck Foundation Medical Research Award to T.K. We additionally acknowledge the following fellowships: NIH IRACDA and UC Chancellor's Postdoctoral Fellowships (A.A.G.), PhRMA Foundation Predoctoral Fellowship in Informatics (D.J.M.), NIH F32 Postdoctoral Fellowship (M.T.), and National Science Foundation Graduate Research Fellowships (J.P. and N.O.). **Author contributions:** D.J.M. and T.K. conceived the idea for the project; D.J.M. developed and performed the majority of the computational design with contributions from A.A.G., R.A.P., K.A.B., N.O., J.P., and T.K.; A.A.G. and Y.-M.H. designed the experimental approach and performed the majority of the experimental characterization with contributions from R.R., A.L.L., C.K., D.J., and M.J.S.K.; M.T. and J.S.F. determined the crystal structure; M.J.S.K., J.S.F., and T.K. provided guidance, mentorship, and resources; and A.A.G. and T.K. wrote the manuscript with contributions from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Coordinates and structure files have been deposited to the Protein Data Bank (PDB) with accession code 6OB5. All other relevant data are available in the main text or the supplementary materials. Rosetta source code is available from rosettacommons.org. Upon publication, constructs will be made available by Addgene.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/1024/suppl/DC1
Materials and Methods
Supplementary Text
Tables S1 to S4
Figs. S1 to S14
References (*38*–*73*)

1 May 2019; accepted 7 October 2019
10.1126/science.aax8780

Downloaded from http://science.sciencemag.org/ on November 29, 2019

CANCER

Peritumoral activation of the Hippo pathway effectors YAP and TAZ suppresses liver cancer in mice

Iván M. Moya[1,2]*, Stéphanie A. Castaldo[1]*, Laura Van den Mooter[1]*, Soheil Soheily[1]*, Leticia Sansores-Garcia[1], Jelle Jacobs[3], Inge Mannaerts[4], Jun Xie[1], Elisabeth Verboven[1], Hanne Hillen[1], Ana Algueró-Nadal[1], Ruchan Karaman[1], Matthias Van Haele[5], Weronika Kowalczyk[1], Maxime De Waegeneer[3], Stefaan Verhulst[4], Panagiotis Karras[6], Leen van Huffel[1], Lars Zender[7,8,9], Jean-Christophe Marine[6], Tania Roskams[5], Randy Johnson[10], Stein Aerts[3], Leo A. van Grunsven[4], Georg Halder[1]†

The Hippo signaling pathway and its two downstream effectors, the YAP and TAZ transcriptional coactivators, are drivers of tumor growth in experimental models. Studying mouse models, we show that YAP and TAZ can also exert a tumor-suppressive function. We found that normal hepatocytes surrounding liver tumors displayed activation of YAP and TAZ and that deletion of *Yap* and *Taz* in these peritumoral hepatocytes accelerated tumor growth. Conversely, experimental hyperactivation of YAP in peritumoral hepatocytes triggered regression of primary liver tumors and melanoma-derived liver metastases. Furthermore, whereas tumor cells growing in wild-type livers required YAP and TAZ for their survival, those surrounded by *Yap*- and *Taz*-deficient hepatocytes were not dependent on YAP and TAZ. Tumor cell survival thus depends on the relative activity of YAP and TAZ in tumor cells and their surrounding tissue, suggesting that YAP and TAZ act through a mechanism of cell competition to eliminate tumor cells.

The Hippo signaling pathway has been implicated in cancer development in humans and mice. Its two downstream effectors, the homologous YAP and TAZ transcriptional coactivators, are upregulated in a wide range of human cancers (*1*, *2*). When active, YAP and TAZ accumulate in the nucleus and, through their binding to transcription factors such as TEAD, they regulate the expression of target genes that promote cell proliferation, stemness, and cell survival. Their hyperactivation can trigger the expansion of progenitor cell populations, induce organ overgrowth, and lead to cancer initiation (*2*, *3*). YAP and TAZ are considered attractive targets for cancer therapy because they are required for the proliferation and survival of cancer cells while being largely dispensable for homeostasis of many adult mouse tissues (*1*–*3*). However, little is known about the function of YAP and TAZ in normal cells of the immediate and larger tumor environments.

To study the function of the Hippo pathway in tumor and peritumoral cells, we induced the development of intrahepatic cholangiocarcinoma (CCA) in adult mice by hydrodynamic tail vein injection of genome-integrating sleeping beauty (SB) plasmids that express activated versions of the Notch receptor [MYC-tagged Notch intracellular domain (N^{ICD})] and Akt [myristoylated and hemagglutinin-tagged (HA-Akt)] (*4*). This procedure results in transfection of scattered hepatocytes and leads to multiple macroscopic tumors 6 to 7 weeks after DNA injection into C57BL/6 mice (fig. S1), hereafter referred to as N-Akt tumors (*4*). High levels of YAP and TAZ were present in N-Akt tumor cells (Fig. 1A and fig. S2A), as is seen in human cholangiocarcinoma (*5*). YAP levels in tumor cells were as high as those in bile duct and endothelial cells, where YAP is normally expressed (Fig. 1A) (*6*, *7*).

We also observed YAP accumulation in hepatocytes around N-Akt tumors; in contrast, YAP was barely detectable in hepatocytes of normal livers (Fig. 1A). This accumulation of YAP in peritumoral hepatocytes was not due to direct effects of N^{ICD} or HA-Akt expression in these cells, because all peritumoral hepatocytes showed YAP accumulation whereas only a few of them expressed N^{ICD} or HA-Akt (fig. S2, B to C). Moreover, N^{ICD} or HA-Akt expression alone was not sufficient to induce YAP accumulation in peritumoral hepatocytes (fig. S2D). YAP accumulation was not due to elevated levels of *Yap* mRNA in hepatocytes (fig. S2E) but to posttranslational regulation, because the presence of N-Akt tumors triggered nuclear accumulation of ectopically expressed HA-tagged YAP in peritumoral hepatocytes (Fig. 1, B and C).

We also observed YAP and TAZ accumulation in peritumoral hepatocytes of ~50% of human hepatocellular carcinoma (HCC) (44 out of 82) and CCA (13 out of 26), but not in hepatocytes of healthy human livers (fig. S3). When present, peritumoral YAP was strongly nuclear in HCC patients (fig. S4). In addition, YAP accumulated in peritumoral zones of colorectal cancer and melanoma metastases to the liver (fig. S5). Notably, unlike most primary liver cancers, these metastases occurred in normal livers, demonstrating that the peritumoral activation of YAP was due to the tumor and not to underlying liver disease.

To address whether YAP and TAZ (YAP/TAZ) were active in peritumoral hepatocytes, we used RNA sequencing (RNA-seq) to study the gene expression profiles of purified hepatocytes from normal livers and livers with N-Akt tumors (fig. S6, A to C, and materials and methods). This analysis identified 3273 and 523 genes that were significantly up- or down-regulated, respectively [$\log_2$FC > 1, false discovery rate (FDR) < 0.05]. The up-regulated genes were enriched for those encoding factors that function in cell proliferation, stress response, and wound healing (fig. S6D), and gene set enrichment analysis (GSEA) detected prominent Hippo pathway gene expression signatures (*8*–*10*) [normalized enrichment score (NES) = 2.57 and 2.33, FDR < 0.05] (Fig. 1, D and E). Classic YAP targets, including *Ctgf*, *Cyr61*, *Pdgfr*, *Fbn1*, *Ankrd1*, and *Birc5*, were among the up-regulated genes, which was confirmed by quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) (Fig. 1, D and F). Consistent with YAP activation, ~6% of peritumoral hepatocytes expressed a marker of cell proliferation ($HNF4\alpha^+$ $Ki67^+$) in tumor-bearing livers, whereas <0.2% of hepatocytes expressed Ki67 in normal livers (fig. S7). These data show that YAP/TAZ were ectopically activated in peritumoral hepatocytes.

To determine the function of the Hippo pathway in peritumoral hepatocytes, we simultaneously deleted *Yap* and its redundantly acting homolog *Taz* in hepatocytes but not in tumor cells. This was achieved by injecting a Cre-expressing adeno-associated virus 8 (AAV-Cre) into $Yap^{fl/fl}$;$Taz^{fl/fl}$ double-floxed mice (*11*, *12*) that had N-Akt tumors (Fig. 2A). AAV-Cre expresses Cre under the hepatocyte-specific thyroxine-binding globulin promoter (*4*, *13*), and its injection into *R26-LoxP-STOP-LoxP-tdTomato* reporter mice bearing N-Akt tumors triggered recombination of the *tdTomato* reporter in essentially all hepatocytes but in <0.05% of cholangiocarcinoma cells (fig. S8, A

[1]VIB Center for Cancer Biology and KU Leuven Department of Oncology, University of Leuven, Leuven, Belgium. [2]Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Americas, Quito, Ecuador. [3]VIB Center for Brain and Disease Research and KU Leuven Center for Human Genetics, University of Leuven, Leuven, Belgium. [4]Liver Cell Biology Research Group, Vrije Universiteit Brussel, 1090 Brussels, Belgium. [5]Department of Imaging and Pathology, KU Leuven and University Hospitals Leuven, Leuven, Belgium. [6]Laboratory for Molecular Cancer Biology, VIB Center for Cancer Biology and Department of Oncology, KU Leuven, Leuven, Belgium. [7]Department of Medical Oncology and Pneumology, University Hospital Tuebingen, 72076 Tuebingen, Germany. [8]German Cancer Research Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [9]iFIT Cluster of Excellence EXC 2180 "Image Guided and Functionally Instructed Tumour Therapies," University of Tuebingen, Tuebingen, Germany. [10]Department of Cancer Biology, Program in Cancer Biology, Graduate School for Biological Sciences GSBS, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: georg.halder@vib.be

Downloaded from http://science.sciencemag.org/ on November 29, 2019

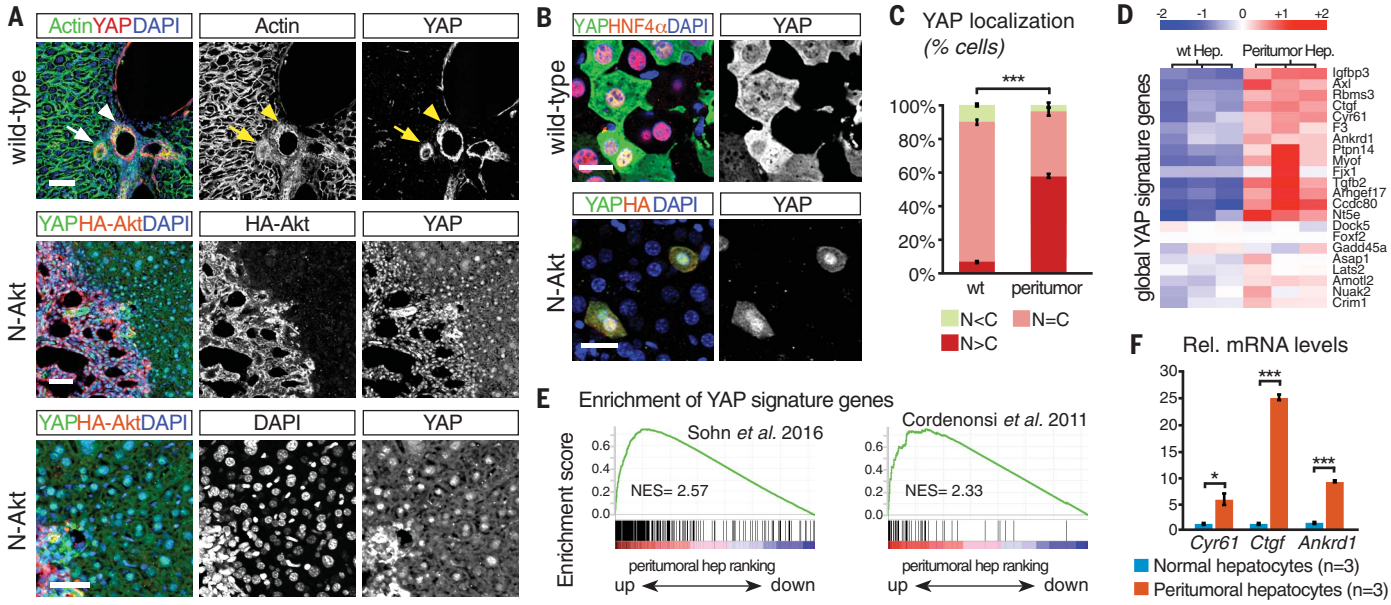


Fig. 1. YAP is up-regulated in peritumoral hepatocytes. (**A**) Immunofluorescent detection of YAP on sections of WT mouse livers and livers with N-Akt tumors. Tumor cells were detected by HA-Akt expression (red). Arrows and arrowheads indicate YAP in bile duct and portal vein cells, respectively. DAPI, 4′,6-diamidino-2-phenylindole. Scale bars, 100 mm. (**B** and **C**) Immunofluorescent images and quantification of the localization of ectopically expressed HA-tagged YAP, 2 days after transfection into a few hepatocytes of normal (WT) livers and livers with N-Akt tumors. N, nuclear; C, cytoplasmic. Scale bars, 50 mm. (**D**) Heat map showing up-regulation of global YAP signature genes in peritumoral hepatocytes (Hep.) relative to hepatocytes from normal livers. (**E**) GSEA plots showing the distribution of two other more inclusive sets of YAP signature genes identified from human HCC samples (*10*) and cultured cells overexpressing YAP (*8*, *28*). (**F**) Quantitative RT-PCR for YAP/TAZ target genes in purified normal and peritumoral hepatocytes. Rel., relative. Data are means ± SEM. *P < 0.05, ***P < 0.001; n.s., not significant.

to D). AAV-Cre was administered 4 weeks after N^{ICD} and *HA-Akt* plasmid injection, and mice were analyzed 3 weeks later (Fig. 2A). Analysis of *Yap* and *Taz* mRNA and protein levels confirmed robust deletion in peritumoral hepatocytes (fig. S8, E and F). Deletion of *Yap*/*Taz* in peritumoral hepatocytes resulted in increased tumor burden (Fig. 2, B to E). We established this by quantifying the relative tumor area on liver sections (excluding the tumor luminal spaces) and by determining the absolute tumor mass (by multiplying the relative tumor area with the liver weight) (Fig. 2, B to D). The increase in tumor burden was due to an increase in tumor cell proliferation, as evidenced by the increased number of Ki67-positive tumor cells (Fig. 2, F and G). This effect was a result of *Yap*/*Taz* deletion and not due to nonspecific effects of AAV-Cre infection because AAV-Cre injection did not affect the tumor burden in C57BL/6 control mice (fig. S9). These data highlight an unexpected activity of YAP/TAZ in peritumoral hepatocytes that nonautonomously restrains tumor growth.

The finding that endogenous activation of YAP/TAZ in peritumoral hepatocytes restrains tumor growth prompted us to test whether additional activation of YAP/TAZ in peritumoral hepatocytes could cause tumor elimination. To study this question, we activated YAP/TAZ by conditional deletion of the genes encoding the upstream large tumor suppressor kinases 1 and 2 (LATS1/2), which inhibit YAP/TAZ by phosphorylation (*3*). We injected the N^{ICD} and *HA-Akt* plasmids into $Lats1^{fl/fl}$;$Lats2^{fl/fl}$ double-floxed mice (*14*) and triggered *Lats1/2* deletion in peritumoral hepatocytes by AAV-Cre injection 4 weeks later when livers already had macroscopic tumors (Fig. 3A and fig. S10A). Peritumoral deletion of *Lats1/2* caused a decrease in phospho-S112-YAP, an increase in TAZ (fig. S10B), strong activation of hepatocyte proliferation (fig. S10, C and D), and liver overgrowth (Fig. 3B). Thus, deletion of *Lats1/2* hyperactivated YAP/TAZ above the levels present in wild-type (WT) peritumoral hepatocytes.

Two weeks after *Lats1/2* deletion, most mutant livers showed markedly reduced tumor loads compared with controls (Fig. 3, C to E). In control mice, tumors occupied 40% of the liver volume whereas in mice with *Lats1/2* mutant livers, tumors occupied <5% of the liver volume (fig. S10E). This represents a >70% reduction in tumor load (Fig. 3C). This effect was mediated by YAP/TAZ, as simultaneous deletion of *Yap*, *Taz*, *Lats1*, and *Lats2* in peritumoral hepatocytes abolished the ectopic hepatocyte proliferation and tumor elimination observed upon *Lats1/2* deletion (Fig. 3, B and C, and fig. S10, C to E). This experiment confirmed that YAP/TAZ act as tumor suppressors in peritumoral hepatocytes, because mice with *Yap*, *Taz*, and *Lats1/2* quadruple mutant hepatocytes had increased cancer cell proliferation and tumor burden when compared with non-AAV-Cre–injected siblings (fig. S10, F to L).

In a second approach to activate YAP/TAZ in peritumoral hepatocytes, we used transgenic mice that conditionally overexpress a constitutively active form of human YAP ($hYAP^{1SA}$), which contains a mutation of the major LATS1/2 phosphorylation site, under the doxycycline (Dox)–inducible TetON system (*15*). These mice showed induced expression of $hYAP^{1SA}$ specifically in hepatocytes but not in cholangiocytes (fig. S11A) because they express the reverse tetracycline transactivator (rtTA) under the hepatocyte-specific *ApoE* promoter (hereafter called *Apo>hYAP*^{ISA}) (*15*). Indeed, Dox feeding led to an increase in YAP protein levels (fig. S11B) and induced hepatocyte proliferation and liver overgrowth in *Apo>hYAP*^{ISA} mice (fig. S11, C to E) (*15*). We next induced N-Akt tumors in *Apo>hYAP*^{ISA} mice and activated $hYAP^{1SA}$ expression at week 4 of tumor development by adding Dox to their drinking water (fig. S11F). Two weeks later, these mice displayed a prominent reduction in tumor load compared with nontreated *Apo>hYAP*^{ISA} siblings and Dox-treated C57BL/6 mice with

Downloaded from http://science.sciencemag.org/ on November 29, 2019

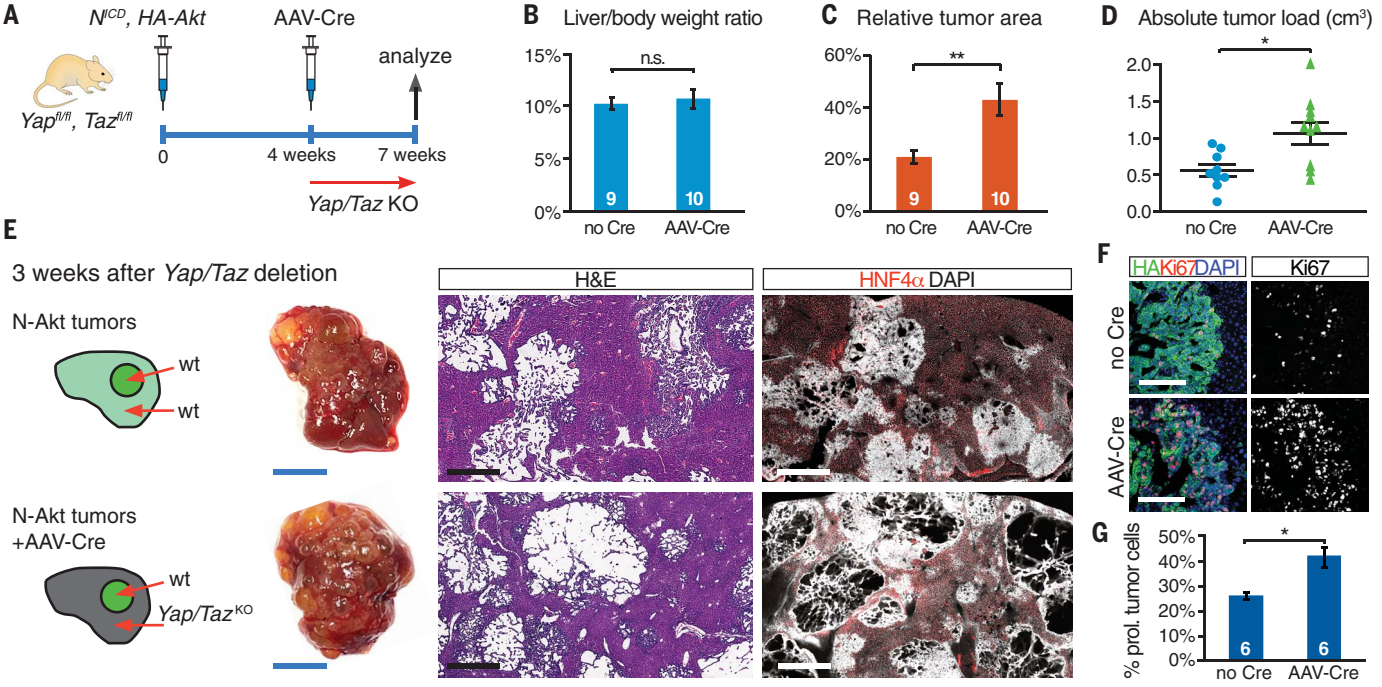


Fig. 2. YAP/TAZ in peritumoral hepatocytes restrain tumor growth. (**A**) Experimental outline: *Yap*$^{fl/fl}$;*Taz*$^{fl/fl}$ mice were hydrodynamically injected with the *N*^{ICD} and *HA-Akt* plasmids. After 4 weeks, half of these mice were injected with AAV-Cre, and all were euthanized at 7 weeks. KO, knockout. (**B** to **D**) Quantifications of liver to body weight ratios, relative tumor area, and absolute tumor load of these mice. Numbers in the bars represent number of mice; blue dots and green triangles represent individual mice. (**E**) From left to right: Schematics depicting mosaic of genotypes, where shade of green indicates levels of YAP/TAZ activity; whole liver pictures (scale bars, 1 cm); liver sections stained with hematoxylin and eosin (H&E) (scale bars, 1 mm); and liver sections stained for HNF4α (hepatocytes, red) and DAPI (white) (scale bars, 500 μm). (**F** and **G**) Ki67 staining (red) to detect proliferating cells and their quantification. Tumor cells are marked in green (anti-HA-tag). Scale bars, 100 μm. Data are means ± SEM. *P < 0.05, **P < 0.01; n.s., not significant.

N-Akt tumors (Fig. 3F and fig. S11, G to I). Mice that were fed Dox for 2 weeks survived longer, although they eventually died from recurrent cholangiocarcinoma (Fig. 3G). Thus, ectopic YAP activation in peritumoral hepatocytes is sufficient to induce tumor regression.

To investigate whether tumor elimination is due simply to peritumoral hepatocyte overproliferation, we simultaneously overexpressed oncogenic BRAF^{V600E} and deleted *PTEN* in peritumoral hepatocytes to trigger ectopic hepatocyte proliferation (*16*). Although *BRaf*^{V600E};*PTEN*^{CKO} livers overgrew at the same rate and to the same size as *Apo>hYAP*^{ISA} livers, this did not affect N-Akt tumor growth (fig. S12). Therefore, tumor elimination is not caused by excessive proliferation of peritumoral hepatocytes, but rather it is a specific effect driven by YAP/TAZ activation.

The decrease in tumor burden was likely due to the elimination of tumor cells by programmed cell death. N-Akt tumors surrounded by *Lats1/2* mutant or hYAP^{1SA}-expressing hepatocytes had more cells that were positive for terminal deoxynucleotidyl transferase–mediated deoxyuridine triphosphate nick end labeling (TUNEL) staining (which labels fragmented DNA in dying cells) (Fig. 4, A and B, and fig. S13A) but displayed no change in proliferation (fig. S13, B and C). Regressing tumors had lower levels of the anti-apoptotic protein BCL2 compared with tumors from control mice (Fig. 4C), but there were no significant increases in cleaved caspase-3, poly(ADP-ribose) polymerase 1 (PARP-1), tumor hypoxia (HIF2A and GLUT1), phosphorylated RIPK3 or p38, or immune cell infiltration (CD45 and CD3; fig. S13, D and E). Nor were there decreases in readouts for nutrient status of tumor cells such as phosphorylated S6, AKT, and AMPK, or growth factor signaling (phosphorylated extracellular signal–regulated kinase) (Fig. 4C and fig. S13, G and H). We also did not observe substantial changes in the amounts of YAP, TAZ, or phosphorylated YAP compared to control tumors (Fig. 4D), thereby excluding the possibility that tumor regression was due to feedback inhibition of the Hippo pathway. Notably, Dox-induced conditional overexpression of BCL2 in tumor cells, starting at the time of *Lats1/2* deletion or hYAP^{1SA} overexpression in surrounding hepatocytes, abolished tumor elimination (figs. S14 and S15 and materials and methods). Thus, YAP/TAZ activation in peritumoral hepatocytes triggers nonapoptotic programmed cell death in tumor cells, which is prevented by BCL2 overexpression.

The finding that YAP-activated hepatocytes can suppress the growth of liver tumors suggested a competitive interaction between tumor cells and their surrounding tissue. This led us to investigate how YAP and TAZ affect the competitive fitness of established tumors against peritumoral hepatocytes. First, we deleted *Yap* and *Taz* in tumor cells but not in surrounding hepatocytes. We co-injected a plasmid that expressed the tamoxifen inducible Cre^{ERT2} (*SB-Cre*^{ERT2}) together with the *N*^{ICD} and *HA-Akt* plasmids into *Yap*$^{fl/fl}$;*Taz*$^{fl/fl}$ double-floxed mice, and we then triggered *Yap/Taz* deletion by tamoxifen administration 4 weeks later when macroscopic tumors had formed (fig. S16A). Deletion of *Yap* and *Taz* in tumor cells strongly reduced tumor burden (Fig. 4, E to H). Three weeks after *Yap/Taz* deletion, no macroscopic tumors were visible, and the liver parenchyma was largely composed of normal hepatocytes and contained only a few tumor remnants (Fig. 4E). As controls, tamoxifen injection did not affect tumor burden in WT C57BL/6 mice that had *SB-Cre*^{ERT2}–expressing N-Akt tumors or in *Yap*$^{fl/fl}$;*Taz*$^{fl/fl}$ mice with N-Akt tumors that did not have *SB-Cre*^{ERT2} (fig. S9). YAP and TAZ are therefore required for N-Akt tumor maintenance when surrounded by WT hepatocytes.

Downloaded from http://science.sciencemag.org/ on November 29, 2019

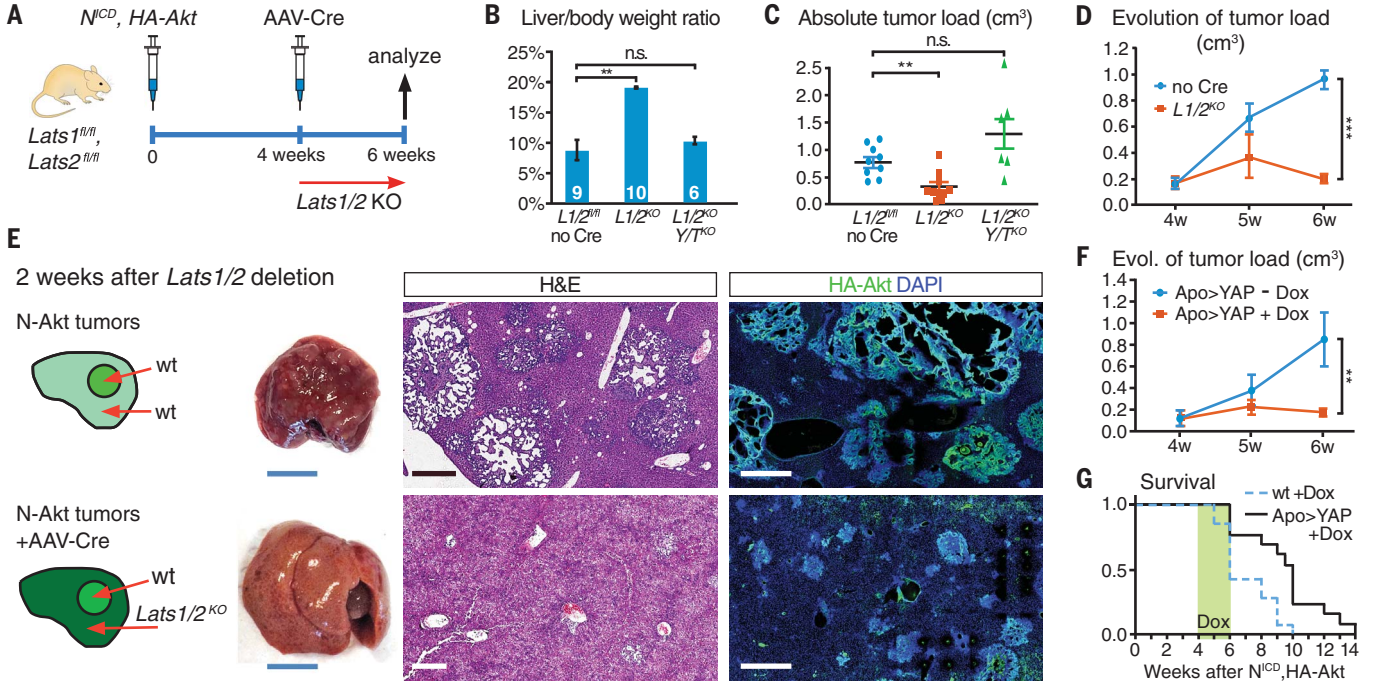


Fig. 3. Hyperactivation of YAP/TAZ in peritumoral hepatocytes causes tumor elimination. (**A**) Experimental outline: *Lats1*$^{fl/fl}$;*Lats2*$^{fl/fl}$ mice were hydrodynamically injected with the *N*^{ICD} and *HA-Akt* plasmids. After 4 weeks, half of these mice were injected with AAV-Cre, and all were euthanized at 6 weeks. (**B** to **D**) Quantifications of liver to body weight ratio, absolute tumor load, and evolution of tumor load in these mice. $P = 0.0011$. (**E**) From left to right: Schematics depicting mosaic of genotypes where shade of green indicates levels of YAP/TAZ activity; whole liver pictures (scale bars, 1 cm); liver sections stained with H&E (scale bars, 1 mm); liver sections stained for tumor cells (HA-Akt expression, green) and nuclei (DAPI, blue). Scale bars, 500 μm. (**F**) Evolution of tumor load. $P = 0.028$. (**G**) Survival of mice after 2 weeks of Dox. Data are means ± SEM. $^{**}P < 0.01$, $^{***}P < 0.001$; n.s., not significant.

In a second experiment, we simultaneously deleted *Yap* and *Taz* in tumor cells and in peritumoral hepatocytes. To do this, we co-injected *SB-Cre*^{ERT2} together with the *N*^{ICD} and *HA-Akt* plasmids into *Yap*$^{fl/fl}$;*Taz*$^{fl/fl}$ mice, as above, but then injected tamoxifen together with AAV-Cre 4 weeks after plasmid injection (fig. S16A). Monitoring of Cre activity with the *Rosa26-LoxP-STOP-LoxP-tdTomato* reporter confirmed ubiquitous tdTomato expression in virtually all tumor cells and peritumoral hepatocytes (fig. S16B). Efficient reduction of YAP/TAZ protein in tumor cells and normal hepatocytes was confirmed by immunohistochemistry and Western blot (fig. S16, C and D). Simultaneous deletion of *Yap/Taz* in tumor cells and in peritumoral hepatocytes prevented tumor elimination and resulted in a similar tumor load as obtained in WT mice (Fig. 4, E to H, and fig. S16, E and F). Thus, N-Akt tumor cells required YAP/TAZ for survival, but only when surrounded by WT hepatocytes.

Finally, we examined whether this tumor suppressor mechanism affects the survival of other types of liver tumors. We tested the effects of peritumoral YAP activation on mouse hepatocellular carcinoma induced by coexpressing *Myc* and *NRas*^{G12V} (*17*) (hereafter called Myc-Ras tumors) in *Apo>hYAP*^{ISA} mice. To prevent *hYAP*^{ISA} expression in Myc-Ras tumor cells, we co-injected plasmids expressing short hairpin RNAs (shRNAs) targeting the *rtTA* and *hYAP*^{ISA} transgenes together with the *Myc* and *NRas*^{G12V} plasmids (figs. S17 and S18) (*18*). Notably, the *YAP*^{ISA} transgene encodes human YAP, which allowed us to design shRNAs that specifically targeted the human *YAP*^{ISA} transgene but not the endogenous mouse *Yap* gene. Control mice that expressed *Myc-Ras* and shRNAs targeting *rtTA* and *hYAP*, but that did not receive Dox, developed a high tumor load 6 to 7 weeks after plasmid injection (Fig. 4, I and J). In contrast, Dox-treated *Apo>hYAP*^{ISA} mice expressing the *rtTA* or *hYAP* shRNA had reduced tumor loads, and mice expressing both shRNAs showed nearly complete tumor elimination (Fig. 4, I and J, and figs. S17 and S18). Therefore, YAP activation in peritumoral hepatocytes is sufficient to eliminate hepatocellular carcinoma.

Many types of cancer metastasize to the liver (*19*). Melanoma liver metastases are lethal, especially for patients whose tumors harbor an activating *RAS* mutation. To test whether YAP-driven cell competition in hepatocytes can suppress the growth of such aggressive metastatic melanomas, we injected syngeneic mouse melanoma cells with an activating *NRAS* mutation and a deletion that abrogates *p16Ink4A* expression, by hydrodynamic tail vein injection into *Apo>hYAP*^{ISA} mice. Hydrodynamic injection of tumor cells is highly effective in establishing tumor growth in the liver (*20*). Mice injected with 10,000 melanoma cells developed macroscopic tumors after 5 weeks and had to be euthanized after 7 weeks (Fig. 4K). We administered Dox in drinking water 3 weeks after the cell injection to induce hYAP^{1SA} expression in hepatocytes (fig. S19A). Two weeks after YAP activation, the *Apo>YAP*^{ISA} mice showed a 98% reduction in tumor load (Fig. 4, K and L, and fig. S19, B to D). These melanoma lesions did not trigger endogenous YAP activation in peritumoral hepatocytes. Consistent with this finding, *Yap/Taz* deletion in peritumoral hepatocytes did not alter melanoma growth in the liver (fig. S19, E to J). Thus, YAP-activated hepatocytes can outcompete highly aggressive *NRAS*-mutant melanoma cells.

The tumor suppression effect described here is reminiscent of "cell competition," a phenomenon originally described in *Drosophila* but recently also observed in mammalian systems (*21*, *22*). During competition, intact viable cells are eliminated from a tissue when they are adjacent to cells that have higher fitness (referred to as winner and loser cells) (*21*, *22*). Notably, winner or loser status is not

Downloaded from http://science.sciencemag.org/ on November 29, 2019

an inherent property of a cell but can change depending upon alterations in the neighboring cells. In *Drosophila* imaginal discs, for instance, WT cells are winners when confronting Myc-deficient cells but are losers when confronted with cells that overexpress Myc, a condition that converts cells into supercompetitors (*23*, *24*). Thus, it is the relative level, not the absolute level, of Myc that determines whether a cell is a winner or a loser. Similarly, we found that the reliance of liver tumor cells on YAP/TAZ for their survival was not absolute but instead was relative to the levels of YAP/TAZ in neighboring hepatocytes. Tumor cells died when surrounded by

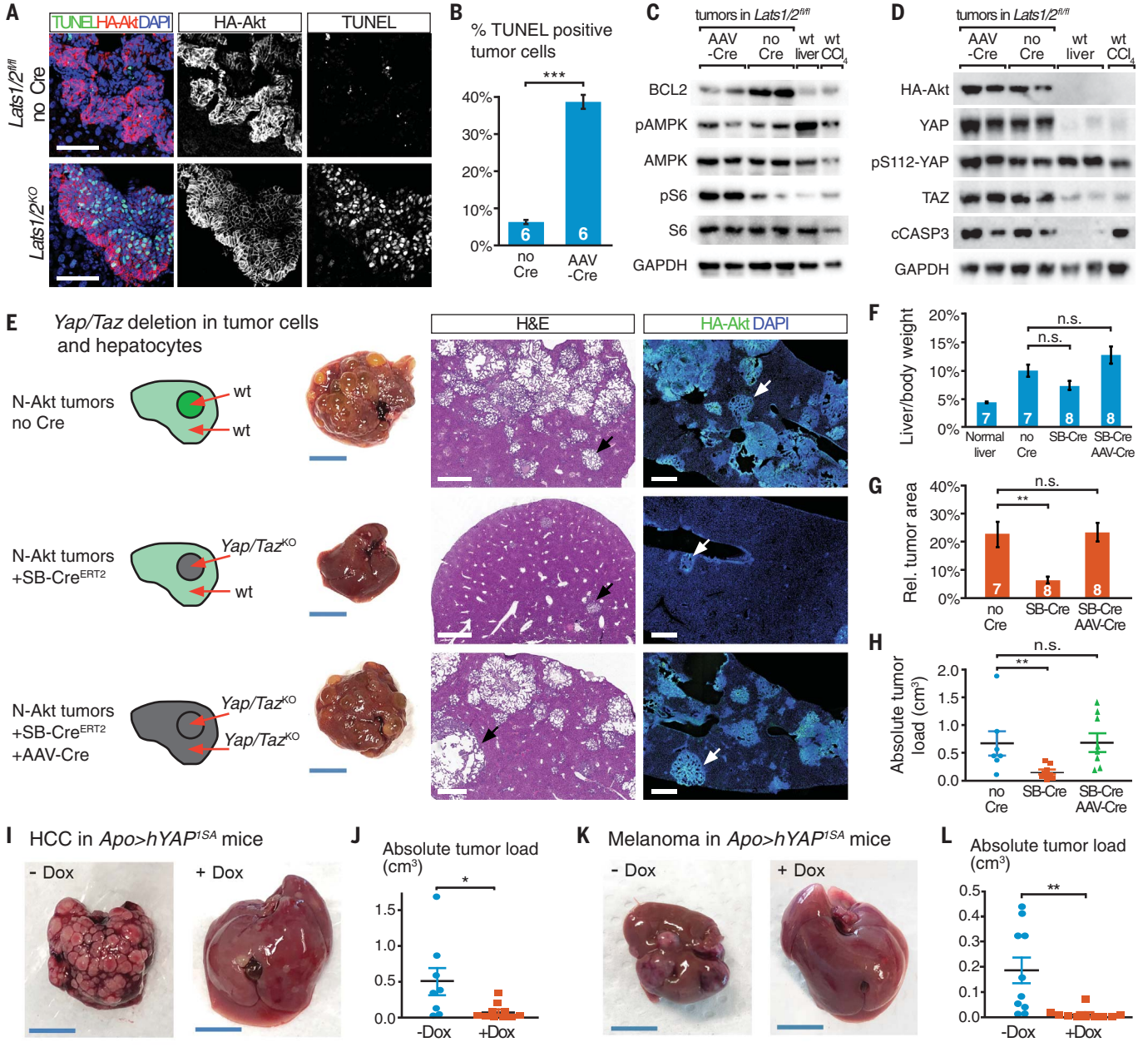


Fig. 4. Tumor cell survival depends on YAP/TAZ levels in surrounding tissue. (**A** and **B**) TUNEL staining (green) and its quantification of liver sections of WT and *Lats1/2*^{CKO} mice with N-Akt tumors 6 days after AAV-Cre administration. Tumor cells are visualized by HA-Akt expression (red). Scale bars, 100 μm. (**C**) Western blots of lysates of tumors dissected from WT and *Lats1/2*^{CKO} livers detecting markers of apoptosis and nutrient status. GAPDH, glyceraldehyde-3-phosphate dehydrogenase. (**D**) Western blots of tumor and whole-liver lysates for YAP, TAZ, and cleaved Caspase 3. Liver injury (CCl_4) was used as positive control for cell death markers. (**E**) From left to right: Schematics depicting mosaic of genotypes of *Yap*$^{fl/fl}$;*Taz*$^{fl/fl}$ mice hydrodynamically injected with the *N*^{ICD} and *HA-Akt* together with or without *SB-Cre*^{ERT2}. At 4 weeks, a portion of these mice received tamoxifen (5 consecutive days) together with or without AAV-Cre, and all were euthanized at 7 weeks. Whole liver pictures are shown next to the schematics (scale bars, 1 cm), and liver sections stained with H&E (scale bars, 1 mm) and for tumor cells (HA-Akt expression, green) and nuclei (DAPI, blue) (scale bars, 500 μm) are on the far right. Arrows indicate HA-positive tumors. (**F** to **H**) Quantifications of liver to body weight ratios, relative tumor area, and absolute tumor load of these mice. (**I** to **L**) Myc-Ras HCC and *NRas-INK4a*$^{-}$ melanoma cells growing in *Apo>hYAP*^{1SA} mice with and without Dox administration starting 4 weeks after tumor initiation. Mice were euthanized and analyzed at 6 weeks. Whole liver pictures (scale bars, 1 cm) and tumor quantifications are shown. Myc-Ras tumor cells were prevented from activating *Apo>hYAP*^{1SA} by expression of shRNAs targeting *rtTA* and *hYAP*^{1SA} (more details and controls are in figs. S17 to S19). Data are means ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001; n.s., not significant.

Downloaded from http://science.sciencemag.org/ on November 29, 2019

hepatocytes that had higher levels of YAP/TAZ activity but survived when competition was neutralized—for example, when tumor cells and hepatocytes simultaneously deleted *Yap/Taz* or when both overexpressed hYAP^{1SA}, such as in the shRNA control mice of the Apo>hYAP^{1SA} experiment with Myc-Ras tumors (fig. S18, B to D). Therefore, a major function of YAP/TAZ in tumor cells is to elevate their competitive fitness and to protect them from the tumor-suppressive action of the surrounding parenchyma.

We have identified in our mouse models a mechanism of non–cell autonomous tumor suppression driven by the Hippo pathway effectors YAP and TAZ, whereby the normal tissue surrounding liver tumors suppresses tumor growth and can cause its regression. Notably, this mechanism of tumor elimination is different from other previously described cell autonomous tumor-suppressive functions of YAP/TAZ (*25*–*27*). Here, YAP and TAZ act non–cell autonomously in normal peritumoral cells and not by direct regulation of target genes in cancer cells. Whether analogous tumor-suppressive mechanisms operate in the human liver—or in other mouse organs—is unknown. Given that YAP and TAZ are hyperactivated in many human cancers (*2*), there is a growing interest in developing drugs that inhibit YAP/TAZ as a possible cancer treatment. The data from the mouse models studied in this work raise the possibility that systemic inhibition of YAP/TAZ could have undesirable protumorigenic effects.

REFERENCES AND NOTES

1. K. F. Harvey, X. Zhang, D. M. Thomas, *Nat. Rev. Cancer* **13**, 246–257 (2013).
2. T. Moroishi, C. G. Hansen, K. L. Guan, *Nat. Rev. Cancer* **15**, 73–79 (2015).
3. R. Johnson, G. Halder, *Nat. Rev. Drug Discov.* **13**, 63–79 (2014).
4. B. Fan *et al.*, *J. Clin. Invest.* **122**, 2911–2915 (2012).
5. P. Marti *et al.*, *Hepatology* **62**, 1497–1510 (2015).
6. N. Zhang *et al.*, *Dev. Cell* **19**, 27–38 (2010).
7. X. Wang *et al.*, *Dev. Cell* **42**, 462–478.e7 (2017).
8. M. Cordenonsi *et al.*, *Cell* **147**, 759–772 (2011).
9. B. J. Pepe-Mooney *et al.*, *Cell Stem Cell* **25**, 23–38.e8 (2019).
10. B. H. Sohn *et al.*, *Clin. Cancer Res.* **22**, 1256–1264 (2016).
11. M. Xin *et al.*, *Sci. Signal.* **4**, ra70 (2011).
12. M. Xin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13839–13844 (2013).
13. K. Yanger *et al.*, *Genes Dev.* **27**, 719–724 (2013).
14. J. Yi *et al.*, *Hepatology* **64**, 1757–1772 (2016).
15. J. Dong *et al.*, *Cell* **130**, 1120–1133 (2007).
16. C. Köhler *et al.*, *Cell Stem Cell* **21**, 679–693.e6 (2017).
17. M. Seehawer *et al.*, *Nature* **562**, 69–75 (2018).
18. C. Fellmann *et al.*, *Cell Rep.* **5**, 1704–1713 (2013).
19. S. S. Agarwala, A. M. Eggermont, S. O'Day, J. S. Zager, *Cancer* **120**, 781–789 (2014).
20. J. Li, Q. Yao, D. Liu, *Cancer Biol. Ther.* **12**, 737–741 (2011).
21. R. Nagata, T. Igaki, *Dev. Growth Differ.* **60**, 522–530 (2018).
22. T. Maruyama, Y. Fujita, *Curr. Opin. Cell Biol.* **48**, 106–112 (2017).
23. C. de la Cova, M. Abril, P. Bellosta, P. Gallant, L. A. Johnston, *Cell* **117**, 107–116 (2004).
24. E. Moreno, K. Basler, *Cell* **117**, 117–129 (2004).
25. S. Strano *et al.*, *Mol. Cell* **18**, 447–459 (2005).
26. M. Yuan *et al.*, *Cell Death Differ.* **15**, 1752–1759 (2008).
27. E. R. Barry *et al.*, *Nature* **493**, 106–110 (2013).
28. B. Zhao *et al.*, *Genes Dev.* **22**, 1962–1971 (2008).

ACKNOWLEDGMENTS

We thank X. Chen, P. Hackett, C. Der, C. Distelhorst, and E. Kowarz for depositing plasmids at Addgene and D. Pan for the *Apo>hYAP^{1SA}* mouse strain. **Funding:** S.A.C., L.V.d.M., and E.V. were supported by doctoral fellowships from the Research Foundation Flanders (FWO). I.M. was supported by a postdoctoral fellowship from FWO. L.Z. was supported by the German Research Foundation: DFG EXC 2180 – 390900677 [Image Guided and Functionally Instructed Tumour Therapies (iFIT)], FOR2314, SFB-TR209, Gottfried Wilhelm Leibniz Program, the German Ministry for Education and Research (BMBF): eMed/Multiscale HCC, the European Research Council: Consolidator grant "CholangioConcept," and the German Center for Translational Cancer Research (DKTK). L.A.v.G. and G.H. were supported by grants from FWO, and G.H. by a grant from Stichting tegen Kanker (FAF-F/2016/867). **Author contributions:** I.M.M., S.A.C., L.V.d.M., S.S., L.S.-G., I.M., J.X., E.V., H.H., A.A.-N., R.K., M.V.H., S.V., P.K., and L.v.H. conducted experiments. J.J., M.D.W., W.K., and S.A. conducted bioinformatics analyses and provided conceptual advice. T.R. performed pathology analysis and provided conceptual advice. R.J. provided genetic strains of mice. L.Z. provided genetic tools and conceptual advice. L.A.v.G and J.-C.M. edited the manuscript and provided conceptual advice. I.M.M. and G.H. designed the study and wrote the manuscript. **Competing interests:** G.H. and I.M.M. are co-inventors on a patent ("Cancer regression by inducing a regeneration-like response," GB 1819659.2) filed by the VIB and KU University of Leuven. **Data and materials availability:** The accession number for the cell sequencing datasets reported in this paper is GEO: GSE103788. The *Apo>hYAP^{1SA}* mouse strain is available from D. Pan at the University of Texas Southwestern Medical Center in Dallas under a material transfer agreement with the Johns Hopkins University, Baltimore. Mice carrying the *Yap* and *Taz* floxed alleles are available from E. N. Olson under a material transfer agreement with the University of Texas Southwestern Medical Center in Dallas.

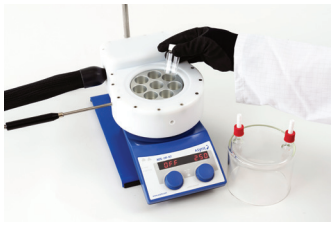
SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/1029/suppl/DC1
Materials and Methods
Figs. S1 to S19

View/request a protocol for this paper from *Bio-protocol*.

12 February 2019; accepted 25 October 2019
10.1126/science.aaw9886

Downloaded from http://science.sciencemag.org/ on November 29, 2019



Cooling Plate

Asynt announces FroSyn, an innovative fluid-free cooling plate that enables rapid cooling of samples in vials or round-bottom flasks (up to 500 mL) to below -60 °C. Unlike dry ice cooling baths that require constant attention, the stable and controlled sub-ambient performance of the FroSyn means that even overnight reactions can be performed with complete confidence. Specialized insulation around the cooling probe and plate on the FroSyn improves performance, keeps the apparatus ice-free, and helps eliminate the possibility of operator cold burns during operation. Positioning of the FroSyn on top of a conventional magnetic stirrer plate allows for efficient sample mixing during cooled experiments. The use of either weights and seals for the flask inserts, or an inerting dome with the vial plates, prevents the formation of ice on the glassware. Other custom cooling accessories are also available to compliment different applications, for example, a 4-position holding module for battery research and development.

Asynt
For info: +44-(0)-1638-781-709
www.asynt.com/product/frosyn-cooling-station

Cell Culture Inserts for Multi-Well Plates

The BRAND Insert 2in1 can be used as both a hanging insert and a standing insert in 2D and 3D cell culture. When used as a hanging insert, it fits nearly all standard multi-well plates. In larger wells and culture dishes, the insert can be used in a standing position. The cell culture insert is available for 6-, 12-, and 24-well plates and in two pore sizes (0.4 µm and 8.0 µm) with polycarbonate (PC) or polyethylene terephthalate (PET) membrane. The cellGrade plus surface of the membranes ensures optimal cell attachment and growth. They are available individually packed or in multi-packs. Common applications include co-culture, transmigration, chemotaxis, transport/barrier analysis (TEER), cytotoxicity, and air-lift culture assays.

BrandTech Scientific
For info: 888-522-2726
www.brandtech.com/product/new-brand-insert-2in1

Alkaline Phosphatase Animal-Free Antibody Labeling Kits

Lightning-Link Alkaline Phosphatase AF (Animal Free) enables scalable, direct conjugation of antibodies, proteins, and peptides (or any other biomolecule with an available amine group), to microbial alkaline phosphatase with only 30 s hands-on time and with 100% antibody recovery. AF Alkaline Phosphatase is expressed in a bacterial host rather than bovine-sourced. This enzyme catalyzes the same reaction as the conventional alkaline phosphatase, without compromising on quality, stability, or accuracy. It works with the same substrates commonly used with conventional alkaline phosphatase (e.g., pNPP and 4-MUP), thus not requiring any workflow changes or special costly reagent. Furthermore, as the kit is BSA free, there are less paperwork and safety restrictions for transport and handling.

Expedeon
For info: 844-611-3656
www.expedeon.com

Freeze Dryer

The LyoConstellation S10 is a Line of Sight–enabled freeze-drying system for undertaking early-stage development under aseptic and controlled conditions. Offering a capacity of 9.8 ft^2 (0.85 m^2) and precise shelf temperature control from -55°C to +65°C, the LyoConstellation S10 is designed, even when fully loaded, to hold 100 mTorr vacuum and shelf temperatures of over +40°C, enabling unmatched sublimation rates with high moisture applications with no evidence of choked flow. Isolator and cleanroom integratable, the LyoConstellation S10 is capable of clean-in-place (CIP) and steam-in-place (SIP) for aseptic manufacturing. Configurable with the latest technologies including SMART cycle optimization technology, ControLyo Ice Nucleation Technology, and LyoFlux TDLAS (Tunable Diode Laser Absorption Spectroscopy) for vapor mass flow measurement and inference of critical data, the LyoConstellation S10 is able to deliver an unmatched array of process analytical data.

SP Scientific
For info: 800-523-2327
www.spscientific.com/LyoConstellationFreezeDryer

Single-Tube Scanners and Rack Readers

The Ziath range of single-tube scanners, compact rack readers, and large multi-rack readers are all capable of reading 2D-barcoded tubes straight from the freezer. A common problem when using 2D-barcoded tube racks straight from the freezer is "fogging" of the scanner or reader glass as condensation forms on the surface which has been cooled by the application of cold racks or tubes. Ziath's proprietary "cryoprotection" coating is proven to stop this fogging from occurring. An additional advantage of Ziath's cryoprotection coating, which is available as an option on all their tube readers and scanners, is that unlike physical methods, such as blowing heated air over the surface, it will not heat up and potentially cause damage to delicate thermally sensitive samples. Ziath's 2D-barcoded tube scanners and rack readers are intuitive to use, reliable, robust, and used by organizations worldwide for biobanking, compound management, and all other sample tracking applications.

Ziath
For info: 858-880-6920
www.ziath.com

Desktop Nanoprinters

Rapidly pattern a wide variety of materials over cm-scale areas with sub-250 nm resolution right from your desktop. The patterning can be done in either direct-write or photolithographic mode. You can change your pattern design instantaneously from the software, as no photomask or mold is needed. With these tools, you can generate patterns of biomolecules, create single-cell arrays, explore new compositions of nanoparticles, pattern polymers, and perform highly localized photochemical and photolithographic applications. You can also discover new materials with unique physicochemical properties for catalysis, photonics, and medical applications at an unprecedented rate. Think Nano? Print Tera!

TERA-print
For info: 847-467-7389
www.tera-print.com

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.

Downloaded from http://science.sciencemag.org/ on November 28, 2019

PINS

RESTORING HOPE



Since the first Deep Brain Stimulation initiative of Tsinghua Univeristy in 2000, PINS Medical has gradually established a multinational corporation with headquarter based in Beijing and international business center in Singapore. As an innovative high-tech enterprise with focus on neuromodulation, a variety of clinical products have been developed to date, which include stimulators for deep brain, vagus nerve, spinal cord and sacral nerve stimulation therapies. PINS Medical devotes itself to providing cutting-edge treatments for patients who suffer from neurological disorders such as Parkinson's Disease, Epilepsy, Chronic Pain and OAB, etc.

As part of the "National Engineering Laboratory for Neuromodulation", PINS Medical works in close cooperation with Tsinghua University and the numerous affiliated clinical centers, becoming a center of attraction for a wide range of professional talents in areas of clinical research, innovative R&D and business management. Since 2008, PINS Medical has developed rapidly in becoming a leading brand in neuromodulation within the Chinese market, due to the success of its creative research platform that efficiency links basic research, R&D of novel products, clinical testing and market entry.

With an outstanding reputation as a high-tech healthcare corporation, PINS Medical has a primary mission for providing innovative, high-quality products and services for patients to improve quality of life. PINS, which stands for Programmable Implanted Neuromodulation Stimulator, is also an abbreviation of "Patient Is No.1 alwayS". This clearly presents the goal of PINS Medical for "restoring hope", not simply as an innovation company but also across society to citizens.

Looking into the future with the continuous rise in incidence of neuropsychiatric disorders and increased social burden across the globe, PINS Medical along with local governments, research centers, companies and top academic scientists, are now developing and promoting innovative therapies worldwide.

www.pinsmedical.com info@pinsmedical.cn

Milestone Celebrations



Image: Library of Congress

19th Amendment

In August 1920, Congress ratified the 19th Amendment. The legal validation for women's voting rights has resulted, directly or indirectly, in the great advances women and other underrepresented groups have made in all spheres of public and private life.

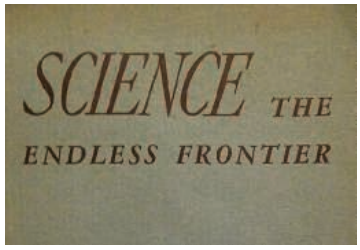


Image: National Science Foundation

Science—The Endless Frontier Report

In 1945, engineer Vannevar Bush created this extensive report for then-president Franklin D. Roosevelt. It provided many justifications for the value of the scientific endeavor and the ways it can support U.S. policy, and has had global and wide-ranging impact in government funding for the sciences in the years since its publication.



Image: NASA

Earth Day

In 1970, Senator Gaylord Nelson of Wisconsin established the first Earth Day, an occasion to raise awareness about environmental concerns for individuals from all walks of life. In doing so, he began a national conversation about how best to protect America's natural resources in the years to come.

AAAS | ANNUAL MEETING
Seattle, WA | February 13–16, 2020

MEETING PROGRAM AND REGISTRATION ARE AVAILABLE ONLINE: AAAS.ORG/MEETINGS

engineering) to design and construct a microbial cell factory that will enable a clean, green, and efficient bioprocess for developing new drugs and for the biosynthesis of natural products.

As these research groups and the SCCE play a major role in developing China's industries and defense systems, the school has received a lot of support from China's central government. The SCCE has been awarded 126 grants totaling over RMB 100 million (USD 14 million) in the last five years from a number of government bodies, including the National Natural Science Foundation, the Ministry of Science and Technology, and the Ministry of Education.

Stellar students and faculty

The growing SCCE can now boast 514 Chinese undergraduates and nearly 50 international students.

It offers five undergraduate programs: physical chemistry, applied chemistry, chemical engineering and technology, pharmaceutical engineering, and energy and chemical industry. Classes are taught in both Chinese and English.

The applied chemistry program has been categorized as a national key program of China, while the chemical engineering and technology and physical chemistry undergraduates belong to the key programs of the Ministry of Industry and Information Technology.

The SCCE runs over three campuses and has 123 faculty members. It also has 245 Ph.D. candidates, over 430 Master's students, and offers the Undergraduate Mentoring System network to help link students with professional mentors.

In its nearly 80-year history, the school has graduated more than 300,000 talented professionals including scholars, entrepreneurs, and government leaders. Graduates include the China's former Premier, Li Peng, and Chinese Minister of Energy Yicheng Huang. A number of alumni have become academics at the Chinese Academy of Engineering and other well-known organizations.

Successful collaborations

The school has won a number of awards for its research, including two second-class prizes of the National Natural Science Award, a first-class prize in the Natural Science Award of the Ministry of Education, and an Outstanding Young Chemist Award.

Graduates from the SCCE have an employment rate of over 98%, and in the past five years, the school has undertaken more than 120 scientific research projects. The school been awarded nearly 100 national invention patents.

The SCCE's undergraduate and graduate students have published 727 Science Citation Index (SCI) papers in last five years, of which 31 were ESI highly cited papers; and 8 Ph.D.s and 14 Master's students received Excellent Thesis Awards from BIT.

"The SCCE's fast-growing international reputation has qualified it as one of China's most important platforms for scientific research and personnel training in chemical sciences and engineering," says Zhang.

As part of its expanding global outreach, the SCCE has collaborations and partnerships with several research institutions. It has recently implemented "2+2" joint education projects with Canada's largest engineering school, the University of Waterloo, and with Queen's University in the United Kingdom. Students who are part of these projects will be trained and granted bachelor degrees from both universities.

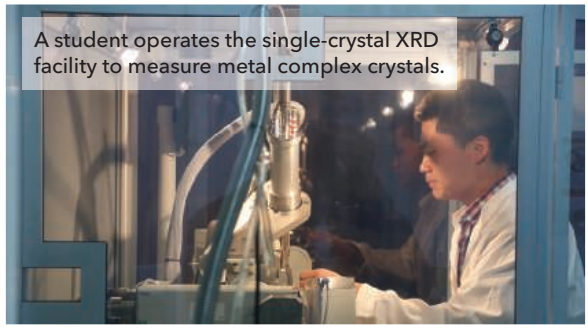
"We have also launched China's first 'Triple One Center,' focused on special molecules, nanoscience, and technology," adds Zhang.

The Triple One Center, based at the Beijing campus, brings together the expertise of international academics to stimulate joint research projects.

"Because of this center, several foreign professors, including Omar M. Yaghi from the University of California, Berkeley Wolfgang Knoll from the Austrian Institute of Technology, and J. Fraser Stoddart from Northwestern University in the United States, were recruited to our program as adjunct professors to stimulate scientific interaction and collaboration," says Zhang.

Zhang says that the SCCE is preparing to bring even more international academics to the Triple One Center in the near future to boost its research.

The SCCE programs also have close ties with other schools at BIT, including the Schools of Materials Science and Engineering, Physics, Mechanical Engineering, and Mechatronical Engineering.



A student operates the single-crystal XRD facility to measure metal complex crystals.

Big ambitions

With a long-running history that has seen a number of mergers and reinventions, the school has great ambitions for new chapters of growth and development.

Zhang says that the SCCE's goals include strengthening the international visibility of BIT's chemistry and chemical engineering programs, attracting more high-level experts, and strengthening undergraduate and graduate education.

Plans include creating an international advisory board, expanding the number of English-language taught courses, and exploring more formal partnerships with foreign universities as well as providing more opportunity for overseas conferences.

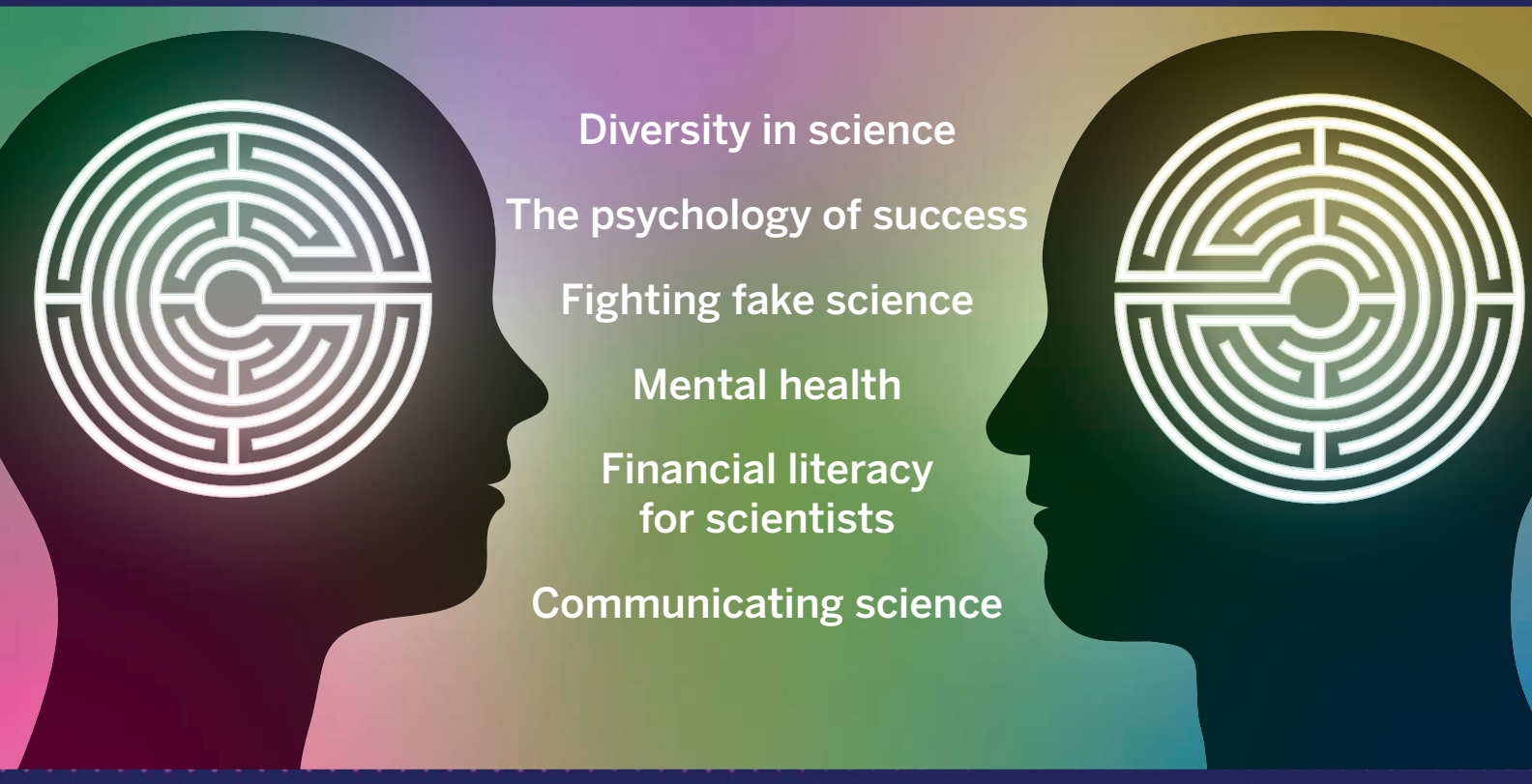
"Our objectives in the next 10 years are to reach world-class level with leading original work in certain research fields, to raise a group of prominent scholars and outstanding young talents, and to establish an interdisciplinary center with top universities for both research and education," says Zhang.

Sponsored by

PHOTO: PEIFENG GAO

The science of being a scientist

Science and Life Webinar Series



Throughout 2019, *Science* and Fondation Ipsen are offering free Science and Life webinars that tackle the issues researchers face in the field.

Sign Up Today

https://scim.ag/35mLvNv

FONDATION IPSEN

ScienceWebinars

AAAS

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS

spj.sciencemag.org/bmef



BME Frontiers

OPEN ACCESS

Biomedical Engineering (BME) Frontiers is a **Science Partner Journal** distributed by the **American Association for the Advancement of Science** (AAAS) in collaboration with the **Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences** (SIBET CAS). *BME Frontiers* aims to serve as an effective platform for the multidisciplinary community of biomedical engineering. The journal will publish breakthrough research in the fields of pathogenic mechanisms as well as disease prevention, diagnosis, treatment, and assessment.

The Science Partner Journals (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, open access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS 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 general science membership society**.

Submit your research to *Biomedical Engineering Frontiers* today!
Learn more at: **spj.sciencemag.org/bmef**

ARTICLE PROCESSING CHARGES WAIVED UNTIL 2021

NEVER STOP

ACCELERATING REGENERATIVE MEDICINE

We're applying our photographic film innovations to help advance new treatments in the revolutionary field of regenerative medicine. Over the last 80-plus years, we've developed advanced technology that controls complex chemical reactions in photographic film that's a mere 20 microns[*1] thick. And today, that technologyis being applied to research and the world's first clinical trial[*2] of medical treatments that use high-quality iPS cells. And in the future, we'll strive to help those suffering from a range of medical conditions, such as those of the eyes, nerves, heart and more. Of course, the challenges are endless, but so are the possibilities. Which is why we'll never stop accelerating regenerative medicine to help build astronger, healthier future for all.

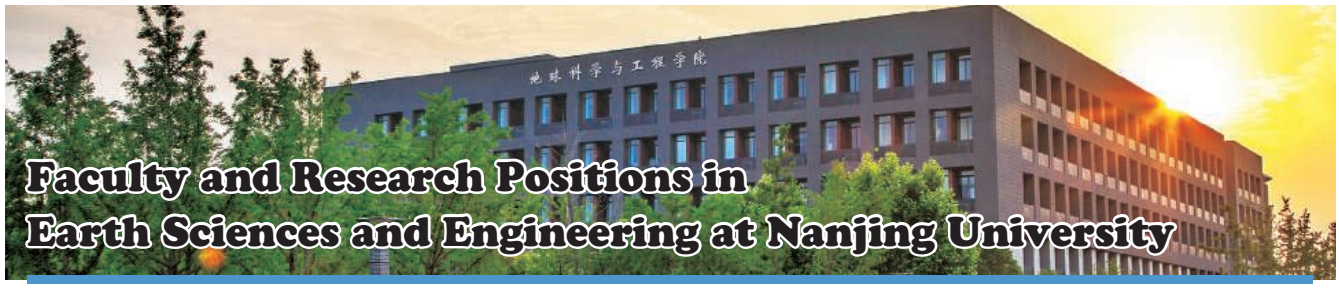
FUJIFILM
Value from Innovation

Follow Fujifilm Life Sciences at in



*1 Thickness of layers excluding the base.
*2 Fujifilm's iPS cells are being utilized in the world's first clinical trial using iPS cells conducted in the UK by the Australian company Cynata.

FUJIFILM and Fujifilm Value from Innovation are trademarks of FUJIFILM Corporation.
©2019 FUJIFILM Corporation. All rights reserved.



Faculty and Research Positions in Earth Sciences and Engineering at Nanjing University

Nanjing University, the School of Earth Sciences and Engineering, invites applications for multiple faculty and research positions at all levels covering the broad field of earth sciences! Nanjing University is a prestigious world-class research institute, and a member of the elite C9 League of Chinese universities. The university is located in the city of Nanjing, which is renowned for its long history, rich culture, and beautiful natural scenery. The earth and environmental sciences of the university are among the top ones in the field, being ranked No. 1 in China and No. 2 worldwide by Nature Index (2019). The school of earth sciences and engineering at Nanjing University traces back to the geology department of the National Southeast University established in 1921, and has now developed into a major institute consisting of three departments: earth and planetary sciences, water sciences, and geological engineering and information technology. The school is committed to developing world-class academic programs, to training leading earth scientists and engineers, to conducting cutting-edge research, and to contributing to serving the country and people. In the fourth-round national evaluation of all universities in 2017, the geology program (level-one rank) in the school was ranked as A+ by the Ministry of Education of the People's Republic of China. In September 2019, two programs of the school, geology and mineralogical engineering, were successfully selected to the "Double First-Class" project, a highly selective government initiative that aims at developing world-leading academic programs and institutes.

The school invites talented overseas researchers to apply for faculty and research positions at all ranks, and welcomes those interested in high-level university/government programs (e.g. the Thousand Talents Young program). The recruitment fields include but are not limited to: Crystal Chemistry of Minerals, High-pressure Mineral Physics, Magma Dynamics, Isotope Geochemistry Geofluids and Ore-forming Fluids, Sedimentology, Tectonics, Geodynamics, Paleoceanography, Environmental Geochemistry and Global Change, Hydrology and Water Resources, Seismology and Applied Geophysics, Engineering Geology and Geohazards Mitigation & Prevention, Paleobiology and biodiversity in deep time, and Planetary Sciences. The university and the school provide competitive salary, all-round support including office room, laboratory space, equipment purchase, student resources, and assist in family issues including daycare and school enrollment for children depending on specific cases.

For more information, please visit the university webpage (https://rczp.nju.edu.cn/17410/list.htm) and school webpage (https://es.nju.edu.cn/). Contact information attached below as well.
Get in touch if you are interested or have any questions! Join us!

Contact Information: Ms. Yao
Email: earthrs@nju.edu.cn.
TEL: 86-25-89680717



School of the Environment at Nanjing University

Established in 1978, school of the environment at Nanjing University is the oldest and the most prestigious environmental program in China. Currently, there are 82 faculty members and four disciplines in the school, including environmental chemistry, environmental biology, environmental engineering and environmental planning and management. Among the faculties, there are one Chinese Engineering Academician, one National Famous Teacher, four Chang Jiang Scholars and four Distinguished Young Researchers of National Science Foundation. According to the latest ranking by Nature Index (2019), Nanjing University ranked No. 1 in the earth and environmental sciences in China and No. 6 worldwide.

School of the environment conducted world-class environmental related research with annual research funding of ~100 million RMB. The school is equipped with state-of-the-art instruments and has set up one State Key Laboratory of Pollution Control and Resource Reuse and one National Engineering Research Center for Organic Pollution Control and Resource. Each year, the school published over 200 SCI papers, about one third of them are tier 1 high quality papers. In the last 10 years, school of the environment won several national awards, e.g., two Second Prize of National Natural Science Awards, four Second Prize of National Technological Innovation Awards, and three Second Prize of National Science and Technology Development Awards.

School of the environment at Nanjing University is enthusiastically engaged in international collaborations for interdisciplinary research and global education. As one of the two founding members of international institute for environmental studies (IIES), school of the environment is the home of IIES. Now IIES is becoming a global leader and platform with 18 members worldwide in the development of research and policy for the management of environmental issues having international dimensions. The IIES has offered online courses, annual science and policy workshop and graduate student forum, et al.

School of the environment is now embracing top talents from all over the world, please join us.

Contact Information: Cheng Gu
Email: chenggu@nju.edu.cn
The Official Website:
http://hjxy.nju.edu.cn/en_wbe/



School of Geography and Ocean Science, Nanjing University

With a reputation for "Cradle of Geographers" in China, the School of Geography and Ocean Science (SGOS) of Nanjing University was founded by Professor ZHU Kezhen (竺可桢) in 1921. Many prestigious professors have devoted their talents and careers to SOGS. In early days, Professor ZHU Kezhen's ground-breaking research on Asian Monsoon; the population density boundary line in China, is also known as the "Hu Huanyong Line", proposed by Professor HU Huanyong (胡焕庸), and the pioneering research on historical geography by Professor ZHANG Qiyun (张其昀) established SOGS as the leader in geography.

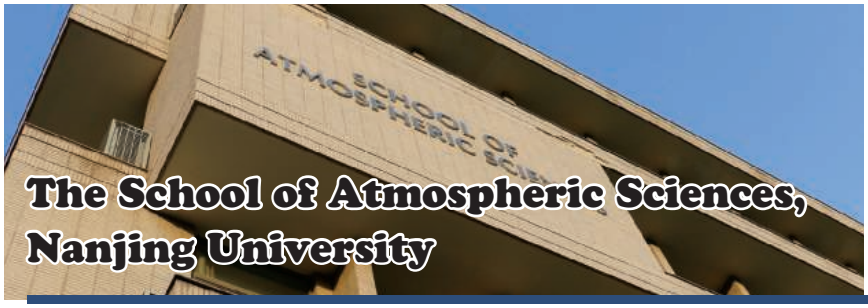
SOGS takes an active attitude to enhance its strengths and adjusts orientations of its disciplines in modern days, pivoting great research resources in the science of Earth surface systems. Due to its persistent excellence, the academic disciplines of SOGS are enlisted in the "Double First-Class" initiative of China. SOGS are currently undertaking numerous research projects funded by the National Natural Science Foundation of China (NSFC), the Ministry of Science and Technology China (MOST), as well as other collaborative projects with prestigious international institutions. The four on-going National Key Research and Development Projects broadly cover research fields in climate change, paleoclimate, carbon cycle and vegetation response, land planning, coastal landform processes, as well as land-use policy and its environmental impacts.

SOGS has 120 full-time faculty and staff, and 30 postdoctoral researchers. There are over 70 undergraduate students, 120 master students and 60 PhD candidates enrolled into SGOS in each academic year. SOGS devotes great efforts to the best education for all-level students. This also echoes with the goal by Nanjing University, that is to provide the best undergraduate education in China

SOGS is equipped with the state-of-art research laboratories and field bases, including the Huangshan National Park Ecosystem Field Scientific Observation and Research Station (MOE), laboratories of AMS C-14 dating, ICP-MS U-Pb dating, OSL dating, Pb-210 dating, tree ring chronology as well as laboratories for isotopic, spectral and geochemical analysis of sediments, ice core and organic matter.

SOGS received a series of awards at both national and provincial levels in the past five years. Annual publications by SOGS are around 500 in peer-review journals, with 60% of them published in international scientific journals, including some of the high-profile journals, such as Nature communications, PNAS, Science Bulletin, Science Advances.

Warmly welcome talents of all levels to join us!



The School of Atmospheric Sciences, Nanjing University



Atmospheric Sciences at Nanjing University, China

The School of Atmospheric Sciences at Nanjing University is the cradle of modern higher education in meteorology in China and has a long history that dates back to the country's first modern meteorology major founded in 1924. Over its near one century of history, the atmospheric science program at Nanjing University has always adhered to a high standard of excellence, and has developed into one of the best atmospheric science programs in China. The school currently consists of about 100 faculty and research staffs, over 300 undergraduate students and over 300 graduate students. While the school holds its traditional strength in meso-scale and climate dynamics, it has recently developed new cross-disciplinary research areas, such as atmospheric pollution and weather/climate interactions, regional and global climate change. The school has recently built new capability in measuring atmospheric and earth system processes and the Station for Observing Regional Processes of Earth System (SORPES) has become one of the most advanced supersites in eastern Asia that helps to monitor regional environment and climate changes. The campus is located in Nanjing, capital of Jiangsu Province, with a population over 8 million. It is recognized as one of the key research and education centers in China.

Position Description

The School of Atmospheric Sciences at Nanjing University is seeking outstanding applicants to fill multiple faculty positions at the ranks of associate and full professors in the areas of meso-scale meteorology, climate dynamics, atmospheric physics, atmospheric chemistry, climate change, physical oceanography, and Earth system science. These positions require a Ph.D. degree in atmospheric sciences or related disciplines, and strong potentials for research and teaching.

Qualifications and responsibilities

- Demonstrated research excellence with a record of publications in high quality journals
- Establish and maintain an externally funded, outstanding research team
- Commit to the school's undergraduate and graduate-level teaching and mentorship duties

Salary and benefits

A competitive startup package will be offered to successful candidates, and salary is commensurate with qualifications and experiences. Benefits include scientific research start-up grant, settling-in allowance, near-campus housing with discounted price, assistance in establishing research group. (See http://hr.nju.edu.cn for details)

Application Process

Application should include a cover letter, a full CV with a list of publications, a copy of five selected publications, a research and a teaching statement, a record of research funding, and professional experiences, and contact information of at least three professional referees. The complete application should be submitted electronically, as a single PDF document, to: Mr. Huan Wang (wangh@nju.edu.cn).

For more information, please feel free to contact Prof. Aijun Ding, Dean of the School.
Tel: +86-25-89683758
Email: dingaj@nju.edu.cn
Website: https://as.nju.edu.cn/as_en/

Science Careers online @sciencecareers.org
OPPORTUNITIES IN CHINA

HIGH-LEVEL GLOBAL TALENTS RECRUITMENT



Welcome back to hometown.
Thousands of academic job vacancies are in fast-developing China.

On-the-spot Recruitment in UK

Dec. 10, 2019 Imperial College London
Dec. 11, 2019 Queen Mary University of London

On-the-spot Recruitment in France

Dec. 14, 2019 Pierre and Marie Curie University
Dec. 15, 2019 Université Paris-Sud

Participating Universities

Jiangsu Normal University
Xuzhou Medical University
Shanghai Institute of Technology
The First Affiliated Hospital of Jinan University
Notheastern University
Shanghai University of Electric Power

Scan the QR code to apply for UK&France Job Fair

Online Job Fair

Dec. 21, 2019(GTM+8) www.edu.cn/cv

Participating Universities

Shanghai University of Political Science and Law
Shanghai International Studies University
Nanchang University
Xi'an Technological University
Inner Mongolia University of Technology
Inner Mongolia University of Finance and Economics
Hunan University of Technology and Business
Inner Mongolia University

Scan the QR code to apply for Online Job Fair

Qualification for Applicants

Overseas scholars, Doctor and Post-doctor

Participating Approach

Please send your CV to consultant@acabridge.edu.cn for on-the-spot Recruitment in UK and France and Online Job Fair.

Job Vacancies in China's Universities and Institutes

Please visit https://www.acabridge.edu.cn/
Contact consultant@acabridge.edu.cn


乡愁，
是那一汪大海，
我在这头，
家人在那头。
千万个
不回的理由，
难抵
一个归根的念头。
Nostalgia,
is like an ocean,
I am here,
the family is over there.
Thousands of reasons
to stay abroad,
but one decision to
return to the roots.
Overseas Chinese Scholars'
Visit to Top Chinese Universities
Check the Details from www.edu.cn/zgx
▶ 10,000+ academic job vacancies in China
▶ Free one-to-one consultation service
Send your CV to consultant@acabridge.edu.cn
上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY
貴州醫科大學
GUIZHOU MEDICAL UNIVERSITY
南京工業大學
NANJING TECH UNIVERSITY
天津师范大学
Tianjin Normal University
青島大学
山东第一医科大学（山东省医学科学院）
Shandong First Medical University & Shandong Academy of Medical Sciences

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Faculty Position in Plant Biology Waksman Institute of Microbiology

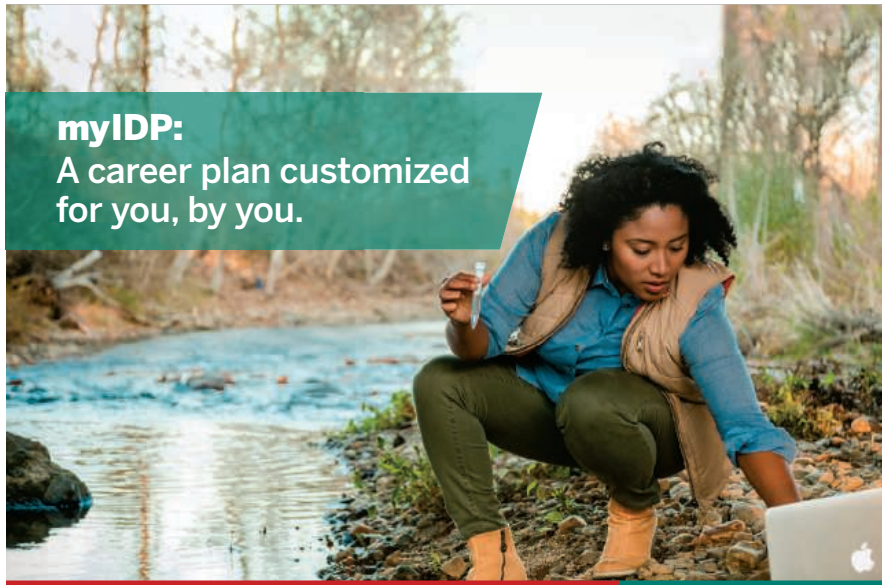
The Waksman Institute at Rutgers University seeks an outstanding scientist to fill a new tenure-track faculty position in Plant Biology at the Assistant Professor level, with a tentative starting date of September 1, 2020. The academic appointment will be in the Department of Plant Biology within the Rutgers School of Environmental and Biological Sciences (http://sebs.rutgers.edu).

We are seeking an individual with research interests in plant biology that complement and expand our existing strengths in genetics, molecular, cellular, and developmental biology, and genomics. Areas of interest include, but are not limited to, plant developmental genetics and epigenetics, cell-cell communication, signal transduction, gene regulation, plant-microbe interactions, plant microbiomes, and plant biotic/abiotic stress responses. The successful candidate is expected to develop a strong extramurally funded research program, teach plant molecular biology, genetics, or biotechnology at the graduate and undergraduate levels, as well as serve on departmental and university committees.

The Waksman Institute is home to 15 faculty members who use a broad range of approaches and experimental systems, including bacterial, animal and plant systems. The institute is equipped with the state-of-art Greenhouse facility featuring 4,200 square feet of growing space and an outdoor field station. Hosted on the Busch campus of Rutgers University, the Waksman Institute is part of a vibrant life science community, neighboring the Center for Advanced Biotechnology and Medicine, the Center for Integrative Proteomics Research, and the Robert Wood Johnson Medical School. The Busch campus is located in central New Jersey, within a short distance from New York City, the Atlantic Ocean, and several major international airports. A leading research university, Rutgers is a member of the AAU and the CIC. For more information, please visit our website https://www.waksman.rutgers.edu.

Applicants must have a Ph.D. or equivalent degree, and should submit a one-page cover letter, a curriculum vitae, a detailed statement of research interests and plans (maximum 4 pages), and full contact information for three individuals willing to provide confidential letters of reference. Applications must be submitted electronically at: **http://jobs.rutgers.edu/postings/105384**. All other inquiries may be made to Ms. Erin Sorge, esorge@waksman.rutgers.edu. Review of applications will begin on **December 16, 2019** and continue until the position is filled.

Rutgers, the State University of New Jersey, is an Equal Opportunity/Affirmative Action Employer. For additional information please see the Non-Discrimination Statement at: http://uhr.rutgers.edu/non-discrimination-statement.



For your career in science, there's only one Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- 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.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with:

UCSF University of California San Francisco

UIC

BURROUGHS WELLCOME FUND

FASEB Federation of American Societies for Experimental Biology

University of Massachusetts Medical School
UMASS

Yale Institute for BIOSPHERIC STUDIES

Hutchinson Environmental Postdoctoral Fellows
Yale Institute for Biospheric Studies
YALE UNIVERSITY

Yale

The Yale Institute for Biospheric Studies (YIBS) was established in 1990 through a generous bequest from Yale alumnus Edward P. Bass. YIBS is an umbrella environmental science organization that supports the environmental community at Yale University through research and training, grants and fellowships, and weekly seminars and events (yibs.yale.edu). The newly established Hutchinson Environmental Program, named in honor of G. Evelyn Hutchinson, the father of modern ecology, is intended to support the environmental priorities identified in Yale's strategy to advance the sciences.

We seek to attract as many as 10 Hutchinson Environmental Postdoctoral Fellows in the summer and fall of 2020. Cohorts of postdoctoral fellows, representing multiple disciplines, will be centered around the two broad themes outlined below: Environment and Evolution, and Climate and Greenhouse Gases. The Hutchinson Fellows will share space in the Osborn Memorial Laboratories on Yale's Science Hill. Postdoc cohorts will engage in research and synthesis activities with a network of Yale faculty mentors spanning multiple schools and departments. Priority will be given to applicants whose research incorporates data-driven synthesis, modeling, and/or conceptual unification of knowledge as a means to address these core themes. These are two-year postdoctoral fellowships (contingent on success in year one) with a starting salary of $62,000, plus $10,000 for research and travel. Funds will also be available to support cohort projects, symposia, workshops, and training in science communication.

THEME I Environment and Evolution. Human-accelerated environmental changes are affecting the flux of energy and nutrients in ecosystems, in many cases by altering the structure and function of species interactions. We are seeking candidates whose research addresses how species interactions are reshaped by ecological and evolutionary responses to environmental change and/or how this reshaping determines processes and patterns at larger scales. Successful applicants will work in a highly collaborative environment around themes including, but not limited to: (1) eco-evolutionary dynamics in ecosystems driven by human-accelerated change; (2) the genomic basis of evolving species interactions; and (3) contemporary effects of environmental change from a paleoecological perspective. *Faculty lead for Theme I: David Vasseur, Department of Ecology & Evolutionary Biology. Other relevant faculty are affiliated with the Yale School of Forestry & Environmental Studies, and the Departments of Geology & Geophysics and Anthropology. See **yibs.yale.edu/hutchinson-fellowship** for additional details and affiliated faculty.*

THEME II Climate and Greenhouse Gases. Atmospheric levels of CO_2, CH_4 and N_2O are increasing, leading to planetary warming. We are seeking candidates whose research can contribute to our general understanding of the generation and management of greenhouse gases (GHG). Because of growing interest in global methane among a group of Yale faculty, we are particularly interested in building a cohort of postdoctoral fellows who can advance knowledge of the production and control of CH_4 losses from natural and human-managed systems. Successful applicants will work in a highly collaborative environment around themes including, but not limited to: (1) GHG fluxes from ecosystems and the energy sector; (2) microbial ecology and evolution relevant to GHGs; (3) GHG pathways in plants; (4) GHG dynamics in deep time; and (5) natural climate solutions. *Faculty lead for Theme II: Peter Raymond, Yale School of Forestry & Environmental Studies. Other relevant faculty are affiliated with the Departments of Geology & Geophysics, Ecology & Evolutionary Biology, Chemical & Environmental Engineering, and the Yale School of Management. See **yibs.yale.edu/hutchinson-fellowship** for additional details and affiliated faculty.*

APPLICATIONS. Interested candidates should have, or will soon receive, a Ph.D. in a relevant discipline. Applications must be submitted by **Monday, January 13, 2020.** Submit a CV (with a complete bibliography), the names and email addresses of three references, and a research statement detailing potential connections to one (or both) of the themes described above; please also identify one or more potential Yale faculty mentors. Applications should be submitted to **http://apply.interfolio.com/71435.** Applicants will be notified of the results of the selection process by mid-February 2020.

Yale University considers applicants for employment without regard to, and does not discriminate on the basis of, an individual's sex, race, color, religion, age, disability, status as a veteran, or national or ethnic origin; nor does Yale discriminate on the basis of sexual orientation or gender identity or expression. Title IX of the Education Amendments of 1972 protects people from sex discrimination in educational programs and activities at institutions that receive federal financial assistance. Questions regarding Title IX may be referred to the University's Title IX Coordinator, at TitleIX@yale.edu, or to the U.S. Department of Education, Office for Civil Rights, 8th Floor, Five Post Office Square, Boston MA 02109-3921. Telephone: 617.289.0111, Fax: 617.289.0150, TDD: 800.877.8339, or Email: ocr.boston@ed.gov.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Confused about your next career move?

Download Free Career Advice Booklets!

ScienceCareers.org/booklets





VANDERBILT SCHOOL OF MEDICINE | Basic Sciences

Tenure-Track Faculty Positions – Molecular Physiology and Biophysics

The Department of Molecular Physiology & Biophysics (MPB) at Vanderbilt University invites applications for a tenure-track faculty appointment. MPB has a long and distinguished scientific history and currently has 23 primary faculty members with active research programs studying biophysics, gene regulation, genetics, membrane transport, metabolic/endocrine disorders, neuroscience, signal transduction, single-cell biology, and structural biology. Applicants must have a Ph.D. and/or M.D. degree and an outstanding record of scholarly achievement that demonstrates the potential to establish a highly productive and innovative research program. A generous start-up package will be provided to facilitate cutting-edge investigations into fundamentally significant biomedical research questions. Appointment is expected to be at the Assistant Professor (tenure-track) level, although applicants may also be considered for more senior appointment.

Vanderbilt University provides a rich research environment that strongly supports further career development and fosters opportunities for close multidisciplinary collaborations with faculty across the School of Medicine, Vanderbilt University Medical Center, the School of Engineering, the College of Arts & Science. This culture is enhanced by the diverse activities of several trans-institutional centers and institutes (https://tinyurl.com/yyqdq3t3; https://tinyurl.com/y2cwu27g). The appointee will have access to state-of-the-art core facilities (https://tinyurl.com/y3s24nds; https://tinyurl.com/y96ky46j), and will be eligible for supplementary internal funding mechanisms that foster team-science approaches. Trainees in the appointee's lab also may be eligible for stipend support from over 25 NIH-supported T32 training programs (https://tinyurl.com/y3fnhzst).

Please submit a *curriculum vitae*, a statement of research and teaching interests (maximum five pages) and three relevant publications and arrange for three letters of recommendation to be submitted. All materials should be submitted via Interfolio (**https://apply.interfolio.com/70139**). Review of applications will begin **December 1, 2019**, with applications being accepted until the position is filled. For informal enquiries, please email: mpbfaculty.search@vanderbilt.edu.

Vanderbilt University is located in Nashville, a rapidly growing cosmopolitan city rich in cultural activities with easy access to diverse outdoor recreational opportunities.

Vanderbilt University is an Affirmative Action/Equal Opportunity Employer committed to recruiting and retaining an academically and culturally diverse community of faculty. The university has a strong institutional commitment to diversity in all areas, and particularly encourages applications from minorities, women, individuals with disabilities, and members of other underrepresented groups.

ScienceCareers online @sciencecareers.org

By **Ragupathy Kannan**

Sidestepping politics to teach climate

I sat down at my desk to read the student evaluations for my undergraduate biology courses. "Please leave politics out of it. Not all Republicans deny science," one student wrote. "I didn't pay to take this class to hear him bash Republicans," wrote another. I had received similar comments in previous years, implying that my lectures about Earth's climate—and how it's changing—were unfairly infused with politics. But they still made me think. How can I teach controversial topics such as climate change at a state university in Arkansas, where denial of climate change is rampant, without sounding partisan to some students?

I teach climate science in my biology classes because I think it's important to lay out how the climate is changing before I talk about how those changes may impact plants and animals. Some of my students aren't science majors, so it may be the only time they hear about climate science in a university lecture.

Recognizing that science doesn't exist in a vacuum, I also try to make connections between the topics I'm teaching and the wider world. To do that, I set aside a few minutes at the beginning of lectures to talk about science in the news.

Earlier in the semester, I had started off a lecture by discussing a tweet by President Donald Trump. "Brutal and Extended Cold Blast could shatter ALL RECORDS - Whatever happened to Global Warming?" it read. I used it as a teachable moment to explain to the class that a cold spell doesn't mean global temperatures aren't trending upward.

Later, I brought up Trump again when he revoked California's authority to set stricter auto emissions rules. In response, a student raised a common talking point for climate change deniers. "But hasn't the climate always changed in the past?" he asked, implying that any discussion of auto emissions is moot if this is all part of a natural phenomenon. I answered his question—explaining why scientists are confident that humans have contributed to the recent warming—but I also wondered whether I'd triggered his defensiveness by bringing up politics.

A turning point came the following semester when I received a message from my dean, who regularly reads student evaluations. He had circled a comment from one of my students, writing, "Be careful here." After thinking it through more carefully, I came to realize that in a polarized political landscape, talking about politicians and the decisions they make is counterproductive. Students may put their guard up, thinking that I'm partisan, and tune me out when I'm lecturing about other things, such as climate modeling. So, I made a conscious decision to change my approach to teaching the subject.



"To open my students' minds to the science, I need to find common ground with them."

As part of my modified strategy, I joined a local bipartisan group that aims to bring people together by emphasizing the potential consequences, rather than causes, of climate change. The group taught me about tactful, nonconfrontational ways to discuss climate science.

Over the past year, I have experimented with my teaching methods, and I have learned that to open my students' minds to the science, I need to find common ground with them. Now, when I discuss climate science news, I focus on things that all of us care about and choose examples that illustrate how climate change might affect their lives. For instance, this summer Arkansas and neighboring states were devastated by catastrophic flooding. It was a topic that hit close to home for all of us, and it gave me the chance to tell my students that major floods are now happening more often than they did in the past.

Since I altered my teaching strategy, I've noticed a shift in my student evaluations. I still get the occasional comment that I'm a "climate doomsdayer," but overall the evaluations have been much more positive.

As teachers, we strive to make connections with our students. I've learned that delving into the quagmire of politics hinders those connections. But when I instead highlight the ways in which climate change may affect our shared experience on this planet, it's easier to communicate and connect with all of my students, regardless of their political affiliation. ■

Ragupathy Kannan is a professor at the University of Arkansas in Fort Smith. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Downloaded from http://science.sciencemag.org/ on November 28, 2019

Published by AAAS