**Attacking the many roots of epilepsy** p. 1300

**Evidence, alarm, and vaping regulation** p. 1318

**Tomorrow's Earth** pp. 1327, 1328, & 1329


# Science

$15
13 DECEMBER 2019
sciencemag.org

AAAS


## BECOMING GIANTS

The evolution of size in whales is driven by energetics

pp. 1316 & 1367

# CONTENTS

13 DECEMBER 2019 • VOLUME 366 • ISSUE 6471

1296

## NEWS

### IN BRIEF

**1292** News at a glance

### IN DEPTH

**1294 Donkeys face worldwide existential threat**
Surging demand for hides, used in Chinese medicine, sends populations crashing
*By C. Lesté-Lasserre*

**1295 AIs direct search for materials breakthroughs**
Decision-making algorithms transform how robots evaluate and synthesize solar cells and more *By R. F. Service*

**1296 DNA from Arctic lakes traces past climate impacts**
Ancient vegetation shifts offer clues to the impacts of future warming
*By P. Voosen*

**1298 Doubts persist for claimed Alzheimer's drug**
Once declared a failure, Biogen's antibody drug to be submitted for U.S. approval in 2020 *By K. Servick*

**1299 Cave painting suggests ancient origin of modern mind**
Half-human, half-animal hunters in mythical scene show early artists in Indonesia had sophisticated imaginations
*By M. Price*

### FEATURES

**1300 Epilepsy's next frontier**
Almost one in every three patients suffers from seizures that are resistant to drugs. Scientists are looking for new ways to help them *By J. Couzin-Frankel*

**1304 Foiling epilepsy in a brain at risk**
*By J. Couzin-Frankel*
PODCAST

## INSIGHTS

### LETTERS

**1306 Ingenuity: NextGen's Vision Foods of the future**
*By J. Sills*

### PERSPECTIVES

**1308 Two-dimensional polymers grow up**
Silicon wafer–sized single-layer films are synthesized with a new method
*By O. MacLean and F. Rosei*
REPORT p. 1379

**1309 Molecular messages in human microbiota**
Multiple approaches reveal the molecular wealth from the gut microbiome
*By M. T. Henke and J. Clardy*
RESEARCH ARTICLES pp. 1331 & 1332

**1311 Cell types exposed by social scent**
A function-guided expression profiling strategy uncovers how pheromones are detected *By S. L. Renninger*
REPORT p. 1384

**1312 Extreme diamond-based quantum sensors**
Nitrogen vacancies make for superlative sensors of material properties at high pressures *By J. J. Hamlin and B. B. Zhou*
REPORTS pp. 1349, 1355 & 1359

**1314 Making stronger carbon-fiber precursors**
Cross-linking multiple strands improves strength and toughness of polyacrylonitrile fibers *By B. Fox*
REPORT p. 1376

**1315 Surprising regulation of cell cycle entry**
Structures of heterotrimeric cyclin-dependent kinase 4 complexes reveal the mechanism of drug inhibition *By C. J. Sherr*
RESEARCH ARTICLE p. 1330

**1316 The biology of big**
Whales became the world's largest animals thanks to giant gulps of "bite-size" prey
*By T. M. Williams*
REPORT p. 1367

### POLICY FORUM

**1318 Evidence, alarm, and the debate over e-cigarettes**
Prohibitionist measures threaten public health *By A. Fairchild* et al.

### BOOKS *ET AL.*

**1321 Planning for a warmer world**
A pair of Obama advisers advocate for proactive climate solutions
*By D. A. Reifsnyder*

PHOTO: ZACH MONTES/ORIJIN MEDIA



Published by AAAS

**1322 Mastering the element of competition**
Players race to research the periodic table in an engaging new board game
*By J. V. Hines*

# RESEARCH

## IN BRIEF

**1324** From *Science* and other journals

## REVIEW

**1327 Global Conservation**
Pervasive human-driven decline of life on Earth points to the need for transformative change *S. Díaz* et al.
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAX3100
REVIEWS pp. 1328 & 1329

**1328 Agriculture**
Realizing the potential of digital development: The case of agricultural advice *R. Fabregas* et al.
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAY3038
REVIEWS pp. 1327 & 1329

**1329 Climate Change**
Measuring the success of climate change adaptation and mitigation in terrestrial ecosystems *M. D. Morecroft* et al.
REVIEW SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAW9256
REVIEWS pp. 1327 & 1328

## RESEARCH ARTICLES

**1330 Structural Biology**
p27 allosterically activates cyclin-dependent kinase 4 and antagonizes palbociclib inhibition *K. Z. Guiley* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAW2106
PERSPECTIVE p. 1315

**Microbiota**
**1331** Depletion of microbiome-derived molecules in the host using *Clostridium* genetics *C.-J. Guo* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAV1282
**1332** A metagenomic strategy for harnessing the chemical repertoire of the human microbiome *Y. Sugimoto* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DX.DOI.ORG/10.1126/SCIENCE.AAX9176
PERSPECTIVE p. 1309

**1333 Ancient Atmosphere**
Stepwise Earth oxygenation is an inherent property of global biogeochemical cycling *L. J. Alcott* et al.

**1338 Molecular Biology**
DNA loop extrusion by human cohesin *I. F. Davidson* et al.
REPORT p. 1345

## REPORTS

**1345 Molecular Biology**
Human cohesin compacts DNA by loop extrusion *Y. Kim* et al.
RESEARCH ARTICLE p. 1338

**High-Pressure Physics**
**1349** Imaging stress and magnetism at high pressures using a nanoscale quantum sensor *S. Hsieh* et al.
**1355** Measuring magnetic field texture in correlated electron systems under extreme conditions *K. Y. Yip* et al.
**1359** Magnetic measurements on micrometer-sized samples under high pressure using designed NV centers *M. Lesik* et al.
PERSPECTIVE p. 1312

**1363 Martian Atmosphere**
Global circulation of Mars' upper atmosphere *M. Benna* et al.

**1367 Body Size**
Why whales are big but not bigger: Physiological drivers and ecological limits in the age of ocean giants *J. A. Goldbogen* et al.
PERSPECTIVE p. 1316

**1372 Glycobiology**
Cryo–electron microscopy structures of human ligosaccharyltransferase complexes OST-A and OST-B *A. S. Ramírez* et al.

**1376 Polymers**
High strength in combination with high toughness in robust and sustainable polymeric materials *X. Liao* et al.
PERSPECTIVE p. 1314

**1379 Nanomaterials**
Wafer-scale synthesis of monolayer two-dimensional porphyrin polymers for hybrid superlattices *Y. Zhong* et al.
PERSPECTIVE p. 1308

**1384 Neuroscience**
Sensory coding mechanisms revealed by optical tagging of physiologically defined neuronal types *D. Lee* et al.
PERSPECTIVE p. 1311

## DEPARTMENTS

**1291 Editorial**
Protect global collaboration
*By Alan I. Leshner*

**1410 Working Life**
Turning a blind eye
*By Eva Kinnebrew*

## ON THE COVER

Gigantism evolved multiple times in the oceans, especially among filter-feeding organisms. Baleen whales, such as this humpback whale, represent the most recent radiation of gigantic filter feeders and rank among the largest animals ever. Researchers used direct measures of feeding performance and prey quality to determine the energetic drivers and limits that whales encounter at extremely large body sizes. See pages 1316 and 1367. *Photo: Karim Iliya*

Science Staff ..... 1290
New Products ..... 1390
*Science* Careers ..... 1391

ILLUSTRATION: ROBERT NEUBECKER

**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 © 2020 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $2148; Foreign postage extra: Air assist delivery: $98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. Printed in the U.S.A. **Change of address:** Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling available from backissues.sciencemag.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.



*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

# Protect global collaboration

Organized intellectual espionage is a concept that may not seem relevant to most American scientists because such instances are extremely rare, like most cases of scientific misconduct. But this type of theft from research institutions in the United States, although infrequent and isolated, is very real and very serious. It compromises the scientific values that uphold the openness of academic institutions in the United States as well as the nation's security. Over the past decade, U.S. academic, security, and intelligence communities have identified an increasing number of instances of such misconduct by foreign nations, primarily China but also Russia and Iran. As policies to aggressively deter these occurrences are under vigorous discussion, the scientific community needs to step up and participate in designing the legal path forward, no matter how unpalatable that might seem. Otherwise, scientists will simply be left accepting the cards dealt to them regarding openness and collaboration, which could be even more distasteful.

> "...the scientific community needs to step up and participate in designing the legal path forward..."

Currently, U.S. science funding and security agencies report that just under 200 cases of intellectual espionage are under investigation. Violations include the failure to disclose foreign financial support that is contingent on sharing scientific "secrets" and maintaining "shadow" or duplicate laboratories in two different countries but both working on the same problems and with duplicate support. In addition, granting agencies and scientific journals have been reporting increased violations of peer review confidentiality, in which research proposals or manuscripts under consideration for publication are inappropriately shared with colleagues or students, who then relay the ideas and other information to colleagues elsewhere. If this kind of strategic misconduct goes unchecked, essential scientific values will be undermined: openness and transparency, integrity and trust, fairness and an equal playing field for all, the confidentiality of peer review, and a welcoming environment for students and colleagues no matter their origins and ethnicities. As a result, the climate for science and scientific collaborations could become much more restrictive, and foreign scientists would be at risk of becoming victims of racial and ethnic profiling.

In October, the U.S. Federal Bureau of Investigation (FBI)—in partnership with the American Council on Education, the Association of American Universities, and the Association of Public and Land-grant Universities—convened a summit to discuss the balance between effective strategies to mitigate espionage risks and maintaining the country's openness to foreign students and scientists. The director of the White House Office of Science and Technology Policy, Kelvin Droegemeier, has held multisector discussions this year to explore potential solutions. And the U.S. National Academies of Science, Engineering, and Medicine has been conducting similar roundtables to discuss the nature of the problems and potential strategies for mitigating the risks in ways that will protect scientific values as well as intellectual property. A proposal is pending in the U.S. Congress to support a standing roundtable on these issues at the National Academies. At the FBI summit, director Christopher Wray called for a partnership that is not just about policing intellectual property theft but also better protecting the enterprise from unfair competition and other violations of scientific norms and values. He urged, for example, that all international scientific agreements include requirements for full transparency and reciprocity from all partners, and that there be consequences when those conditions are violated.

Engaging with security agencies is an uneasy concept for many American scientists. They fear excessive incursion into academic life. But it will be necessary to work with them to ensure that the core values of science, particularly open global collaboration, are not compromised by any proposed policies, regulations, or procedures.

**– Alan I. Leshner**

**Alan I. Leshner** is the interim chief executive officer of the American Association for the Advancement of Science (AAAS) and executive publisher of *Science*.
aleshner@aaas.org

PHOTO: PROFESSIONAL IMAGES PHOTOGRAPHY

10.1126/science.aba4762



Published by AAAS

# NEWS

> “It just got really warm. That means that the eggs didn’t survive.”
>
> **Fisheries biologist Steve Barbeaux,** in Alaska Public Media, about the first federal shutdown of the fishing season for Pacific cod in the Gulf of Alaska, after a marine heat wave cut the population.

## IN BRIEF

*Edited by* **Jeffrey Brainard**

An 11-year-old girl receives the measles vaccine; most of those sickened in Samoa have been children.

INFECTIOUS DISEASE

### Samoa battles measles, vaccine fears

The island nation of Samoa reported this week that a large measles outbreak, originally stoked by fears about the vaccine for the highly infectious disease, continued to surge. A total of 4819 cases and 70 deaths, nearly all in children, were recorded between mid-October and 10 December in a population of only about 200,000, the Ministry of Health reported; 87 new cases alone were reported in the 24 hours before the announcement. But the government also reported progress in a vaccination campaign that has covered about 91% of eligible people, from infants to adults. Vaccination rates had been declining for years, and distrust of the highly effective measles vaccine surged after July 2018, when two Samoan infants died because nurses administered doses that mistakenly contained an anesthetic. The World Health Organization estimates that between 2013 and 2018, the percentage of Samoan children who received the measles vaccine fell from 90% to 31% for the first dose and from 72% to 13% for the second and final dose. These levels are unusually low; vaccine coverage must reach 93% to 95% of the population to keep measles from spreading.

### Nuclear lab conversion halted

NONPROLIFERATION | Russia announced on 5 December it has suspended work to convert an Iranian military site into a civilian research center, a side effect of the gradual unraveling of the nuclear deal that world powers struck with Iran in 2015 to deter it from pursuing nuclear weapons. The accord called for reconfiguring 348 uranium-enrichment centrifuges at Iran’s Fordow nuclear site to produce stable isotopes for medicine and research. In a pilot effort, Russia’s nuclear fuel company TVEL had converted 11 centrifuges to enrich isotopes of xenon and tellurium. But last month, Iran resumed enriching uranium at Fordow; in announcing it was suspending work, TVEL declared it was “technologically impossible” to carry out stable isotope production near uranium enrichment.

### Open-access journals defined

PUBLISHING | The funders behind Plan S, a push centered in Europe to make journal articles free to read on publication, have defined the conditions that subscription-based journals must meet for researchers funded by Plan S signatories to publish there. Such “transformative” journals must increase the proportion of open-access articles they publish by at least eight percentage points per year on average over 3 years, according to guidelines proposed last month. The journals must commit to making all their articles immediately open access by the end of 2024 or when more than half of their content has already become so, and must provide transparent pricing. When all Plan S requirements take effect starting in 2021, grantees may also publish in repositories or in other journals in which all content is immediately open access. The draft guidelines are open for comment until 6 January 2020.

### France-Elsevier deal disappoints

PUBLISHING | Publishing giant Elsevier has signed a national license deal with France’s consortium of universities and research organizations that critics say doesn’t do enough to advance full open access to scientific journal articles. The terms are at odds

PHOTO: UNICEF/STEPHEN



Published by AAAS

PHOTO: DEA/C. DANI; I. JESKE/DEAGOSTINI/GETTY IMAGES

BOTANY

## Drought slays iconic Hawaiian silverswords

Despite conservation efforts, the population of silverswords (above) on the volcanic slopes of the Hawaiian island of Maui has dwindled by 60% since 1991—particularly at lower elevations, even though conditions there are less harsh. Now, scientists have an explanation that suggests the plants could be saved by sowing their seeds in areas that conserve moisture. It seems that silversword seeds that start out life in moist soil can't cope later with the dry, hot air that climate fluctuations are increasingly causing. Ecologist Paul Krushelnycky from the University of Hawaii in Honolulu and colleagues grew seeds from different elevations under drought and nondrought conditions; they found that wherever the seeds originated, those that sprouted in moister soil and then grew under drought conditions had the highest mortality, they reported on 29 November in *Ecological Monographs*. The finding suggests conservation efforts that transplant silverswords from high to low elevations won't work, Krushelnycky says, and that planting seeds in areas that retain water, such as those with gentler slopes, is more important.

with Plan S, a research funders' mandate to make publications immediately free to read, which France's National Research Agency has backed. The 4-year agreement announced last month includes modest discounts on journal subscription costs, which the government said will save €1.5 million this year. Unlike deals struck in other European countries that lump all publishing and reading fees, the agreement requires researchers to pay to publish individual articles as open access, but includes a 25% rebate. The agreement also creates a system to add papers to a government-backed open archive after 1 to 2 years—longer than the 6-month delay provided by French law for science papers.

## Space junk removal planned

SPACE FLIGHT | The European Space Agency (ESA) has tapped a Swiss company to launch in 2025 the first mission to remove a piece of debris from orbit. At a meeting last month, ESA member governments created a space safety program to counter threats from inactive satellites, rocket parts, and natural, near-Earth objects. The new program immediately adopted a proposal by ClearSpace to develop a spacecraft to home in on space junk and grasp it in a bear hug using four flexible arms. The vehicle would then descend with its prey into the atmosphere, where both would burn up. ClearSpace will target for ESA a 100-kilogram rocket stage leftover from the launch of the agency's Proba-V Earth-observing satellite in 2013.

## China biometrics studies probed

RESEARCH ETHICS | Publishers Springer Nature and Wiley are investigating whether authors failed to obtain informed consent from hundreds of Chinese minority group members in multiple published research studies, including at least three about technologies that China's police may be using to oppress minorities. The three used genetic or biometric data to predict height and facial features in Eurasian populations; one paper describes an algorithm for facial recognition of ethnic minorities including Uyghurs and Tibetans. Advocates for minorities in China fear that in its far west, where an estimated 1 million individuals, most of them Muslims, have been detained and where most samples are presumed to originate, it is nearly impossible to obtain true informed consent. Springer Nature said it would add notes of concern to the papers, published this year in *Human Genetics* and the *International Journal of Legal Medicine*.

## Database details financial ties

OVERSIGHT | The news organization ProPublica last week made public more than 8100 financial conflicts of interest reported to the U.S. National Institutes of Health by about 3000 researchers who have grants from the agency. The disclosures (online at projects.propublica.org/dollars-for-profs) cover compensation such as consulting fees and equity in companies between 2012 and 2019. The database, Dollars for Profs, also holds more than 29,000 disclosures of faculty and staff members' conflicts reported to public universities.



Published by AAAS

A man sprinkles salt on donkey hides to dry them at an abattoir in Nigeria's Delta state.

PHOTO: PIUS UTOMI EKPEI/AFP/GETTY IMAGES

ANIMAL SCIENCE

# Donkeys face worldwide existential threat

## Surging demand for hides, used in Chinese medicine, sends populations crashing

*By* **Christa Lesté-Lasserre**

Amy McLean has always had a soft spot for long ears. The animal scientist at the University of California, Davis, grew up on a Georgia farm that bred donkeys and mules and she has competed in mule riding world championships. Now, she studies donkey behavior and cognition, and she knows that despite popular negative stereotypes, donkeys are "highly intelligent and highly sensitive."

No wonder she calls the current plight of the world's donkeys "horrific."

Over the past 6 years, Chinese traders have been buying the hides of millions of butchered donkeys (*Equus asinus*) from developing countries and shipping them to China, where they're used to manufacture *ejiao*, a traditional Chinese medicine. The trade has led to an animal welfare nightmare, along with a threat to donkey populations, the severity of which is only now emerging. Without drastic measures, the number of donkeys worldwide will drop by half within 5 years, according to a 21 November report by the Donkey Sanctuary, a charity based in Sidmouth, U.K. The crisis threatens many of the world's rarer donkey breeds and a vital means of transport for the poor. But it is also spurring new studies of donkey biology—including how to speed their reproduction.

*Ejiao*, in use for thousands of years, purportedly treats or prevents many problems, including miscarriage, circulatory issues, and premature aging, although no rigorous clinical trials support those claims. The preparation combines mineral-rich water from China's Shandong province and collagen extracted from donkey hides, traditionally produced by boiling the skins in a 99-step process. Once reserved for China's elites, *ejiao* is now marketed to the country's booming middle class, causing demand to surge. One producer, Dong'e Ejiao in Liaocheng, China, touts it as "a creation of heaven and earth" that's now passing "from the royal tribute to the home of ordinary people."

Despite government incentives for new donkey farmers, farms in China can't keep up with the exploding demand, which the Donkey Sanctuary currently estimates at 4.8 million hides per year. Donkeys' gestation period is one full year, and they only reach their adult size after 2 years. So the industry has embarked on a frenzied hunt for donkeys elsewhere. This has triggered steep population declines. In Brazil, the population dropped by 28% between 2007 and 2017, according to the new report.

African populations are crashing, too, says Philip Mshelia, an equine veterinarian and researcher at Ahmadu Bello University in Zaria, Nigeria. After buying donkeys at markets, traders often drive large herds to slaughter, sometimes covering hundreds of kilometers with no rest, food, or water. Those transported by truck fare worse: Handlers tie their legs together and sling them onto piles or strap them to the top of the truck, Mshelia says. Animals that survive the journey—many with broken or severed limbs—are unloaded by the ears and tails and tossed in front of a slaughterhouse. Some meet their end in an open field where humans await them with hammers, axes, and knives.

For donkey owners, selling their animal means quick cash—now more than $200 in parts of Africa—but it's often a shortsighted deal: The report estimates working donkeys support the livelihood of half a billion people by carrying people and goods to markets, schools, and health clinics.

Ironically, the booming *ejiao* trade, along with a developing donkey dairy industry in Eastern Europe, has stirred scientific interest in donkeys. The number of PubMed-indexed papers with "donkeys" in the title shot up from six in 2000 to 46 so far this year, for example, including many from China.



Published by AAAS

Zhen Shenming, a reproductive biologist at the China Agricultural University in Beijing, says Chinese efforts are focused on increasing yields, for instance through artificial insemination—including better techniques to freeze and thaw sperm—and harmonization of the estrus in female donkeys, which allows farmers to inseminate them more efficiently. "We're improving techniques so jennies don't end up empty," Zhen says. He takes *ejiao* himself occasionally for "better color" and "more energy," although he says scientists "still don't know the bioactivity of the product."

Chinese breeders are also testing new nutrition programs that expedite growth, leading to an adult-size donkey in only 18 months. McLean believes improving welfare at donkey farms could further increase yields. During her visits to China, she encourages farmers to respect animal welfare rules from the Paris-based World Organisation for Animal Health. Some Chinese farms have already installed electric fans to cool the animals or added "enrichment," such as sand pits—donkeys love rolling in the dirt—and scratching posts. Next year, McLean hopes to start to screen Chinese populations for asinine and equine herpesviruses, three of which are known to cause abortion in donkeys.

Faith Burden, research director at the Donkey Sanctuary, hopes *ejiao* can one day be produced without donkeys. A few years ago, researchers at the University of Barcelona in Spain grew hairless, bloodless horse skin in a laboratory, suggesting lab-grown donkey skin might be feasible. A recent survey by the University of Reading in the United Kingdom showed that 48% of *ejiao* consumers in China were open to accepting such a product.

McLean's research so far has focused on behavior—and has made a misnomer of the word "asinine" in its common use, she says. "Donkeys have good reasoning skills," McLean says. "They are very observant and sentient animals, and they create very strong bonds with other donkeys." That's one reason she abhors the current slaughtering practice, in which the animals often await their turn while watching other donkeys being beaten unconscious, slaughtered, and skinned. "They're certainly quite well aware of what's happening and what's to come," McLean says. (Reputable slaughterhouses usually prevent animals from witnessing such scenes.)

In developing countries, education programs could help preserve the threatened herds, McLean says. "We need people to think of donkeys in terms of savings accounts," she says. "If they need to sell a donkey for the money, they need to keep one they can breed so that they always have at least one." ■

Christa Lesté-Lasserre is a science journalist based in Paris.

ARTIFICIAL INTELLIGENCE

# AIs direct search for materials breakthroughs

## Decision-making algorithms transform how robots evaluate and synthesize solar cells and more

*By* **Robert F. Service**, *in Boston*

In July 2018, Curtis Berlinguette, a materials scientist at the University of British Columbia in Vancouver, Canada, realized he was wasting his graduate student's time and talent. He had asked her to refine a key material in solar cells to boost its electrical conductivity. But the number of potential tweaks was overwhelming, from spiking the recipe with traces of metals and other additives to varying the heating and drying times. "There are so many things you can go change, you can quickly go through 10 million [designs] you can test," Berlinguette says.

> **"What used to take us 9 months now takes us 5 days."**
> **Curtis Berlinguette**, University of British Columbia

So he and colleagues outsourced the effort to a single-armed robot overseen by an artificial intelligence (AI) algorithm. Dubbed Ada, the robot mixed different solutions, cast them in films, performed heat treatments and other processing steps, tested the films' conductivity, evaluated their microstructure, and logged the results. The AI interpreted each experiment and determined what to synthesize next. At a meeting of the Materials Research Society (MRS) here last week, Berlinguette reported that the system quickly homed in on a recipe and heating conditions that created defect-free films ideal for solar cells. "What used to take us 9 months now takes us 5 days," Berlinguette says.

Other material scientists also reported successes with such "closed loop" systems that combine the latest advances in automation with AI that directs how the experiments should proceed on the fly. Drug developers, geneticists, and investigators in other fields had already melded AIs and robots to design and do experiments, but materials scientists had lagged behind. DNA synthesizers can be programmed to assemble any combination of DNA letters, but there's no single way to synthesize, process, or characterize materials, making it exponentially more complicated to develop an automated system that can be guided by an AI. Materials scientists are finally bringing such systems online. "It's a

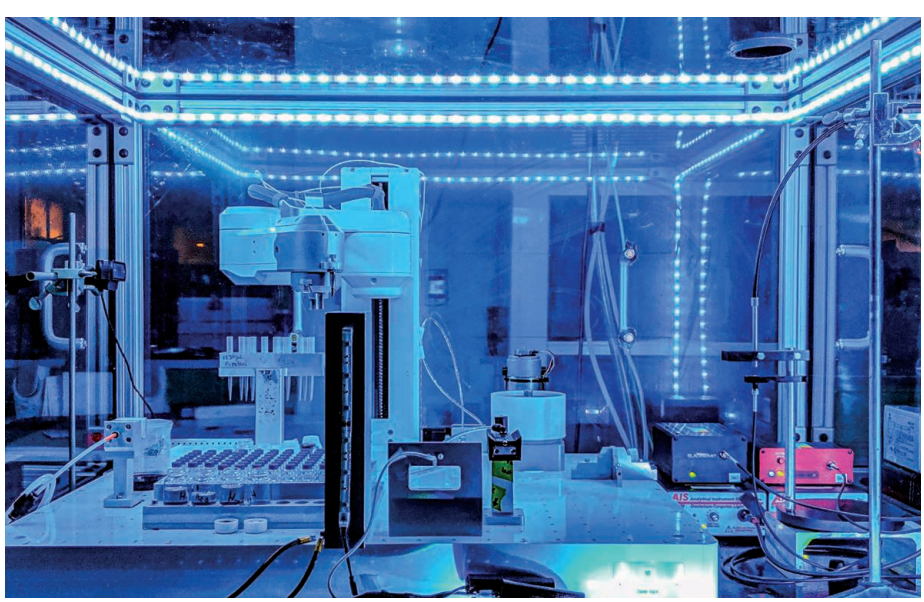


PHOTO: FRASER PARLANE

Ada, an AI-driven robot, searches for new solar cell designs at the University of British Columbia.



Published by AAAS

superexciting area," says Benji Maruyama, a materials scientist with the U.S. Air Force Research Laboratory east of Dayton, Ohio. "The closed loop is what is really going to make progress in materials research go orders of magnitude faster."

With more than 100 elements in the periodic table and the ability to combine them in virtually limitless ways, the number of possible materials is daunting. "The good news is there are millions to billions of undiscovered materials out there," says Apurva Mehta, a materials physicist at the Stanford Synchrotron Radiation Lightsource in Menlo Park, California. The bad news, he says, is that most are unremarkable, making the challenge of finding gems a needle-in-the-haystack problem.

Robots have already helped. They are now commonly used to mix dozens of slightly different recipes for a material, deposit them on single wafers or other platforms, and then process and test them simultaneously. But simply plodding through recipe after recipe is a slow route to a breakthrough, Maruyama says. "High throughput is a way to do lots of experiments, but not a lot of innovation."

> **"The good news is there are millions to billions of undiscovered materials out there."**
>
> **Apurva Mehta**, Stanford Synchrotron Radiation Lightsource

To speed the process, many teams have added in computer modeling to predict the formula of likely gems. "We're seeing an avalanche of exciting materials coming from prediction," says Kristin Persson of Lawrence Berkeley National Laboratory (LBNL) in California, who runs a large-scale prediction enterprise known as the Materials Project. But those systems still typically rely on graduate students or experienced scientists to evaluate the results of experiments and determine how to proceed. Yet, "People still need to do things like sleep and eat," says Keith Brown, a mechanical engineer at Boston University (BU).

So, like Berlinguette, Brown and his colleagues built an AI-driven robotics system. Their goal was to find the toughest possible 3D-printed structures. Toughness comes from a blend of high strength and ductility, and it varies depending on the details of a structure, even if the material itself doesn't change. Predicting which shape will be toughest isn't feasible, Brown says. "You have to do the experiment."

As a test case, the BU team set out to make salt shaker–size, barrel-shaped structures from a plastic. They varied the number of struts that make up the outer wall of the barrel and details of each strut's shape and orientation. Testing all possible combinations, about a half-million, wasn't realistic. So, they initially had their robots fabricate 600 structures that sampled the full array of options. A kind of vise then squeezed each one until it gave way.

The group then added an AI decision-making algorithm that calculated the most likely next best design after each test. The program spots trends in attributes that confer toughness, such as the thickness and radius of each strut, in order to predict even sturdier structures. "We basically turned on the machine and walked out the door," Brown says. After 24 hours and just over 60 designs, the AI-driven system had come up with a tougher barrel than any of the original designs.

Many more closed loop efforts were showcased at the MRS meeting. Researchers at the Massachusetts Institute of Technology in Cambridge and LBNL have independently developed autonomous systems to find better perovskite photovoltaics—cheap, lightweight materials that are poised to revolutionize solar energy. A team at Carnegie Mellon University in Pittsburgh, Pennsylvania, reported using another AI system to find safer charge-carrying electrolytes for lithium-ion batteries, which are now prone to catching fire. And researchers at the University of Liverpool in the United Kingdom have developed a suite of AI-driven robots to discover novel catalysts for generating hydrogen gas, a potential carbon-free fuel, from water.

Few of these projects have turned up blockbuster results, researchers acknowledge. However, Maruyama says, "It's still early days." One challenge is that materials scientists themselves often don't agree on how best to relate a material's conductivity or other testable properties to its structure, says John Gregoire, a physicist at the California Institute of Technology in Pasadena. "If we haven't figured out how to break that down in the community, it's hard to imagine how we will teach a computer to do it," he says.

Another issue is that each team must design its own robotics and software systems, as standards have yet to take shape. "Everyone is exploring different ways to do this," says Joshua Schrier, a computational chemist at Fordham University in New York City. Eventually, the materials community may coalesce around a handful of systems that can be used by a wide swath of researchers, Schrier says. "Over the next year or two I think we'll begin to see things converge." ■

ANCIENT CLIMATE

# DNA from Arctic lakes traces past climate impacts

## Ancient vegetation shifts offer clues to the impacts of future warming

*By* **Paul Voosen**, *in San Francisco, California*

High in the Canadian Arctic on Baffin Island, beneath 10 meters of water and many more of mud, sits a refrigerated archive of Earth's past life. The deep sediments in a small lake called CF8 hold ancient pollen and plant fossils. But it now appears that the mud harbors something else: ancient DNA from as far back as the Eemian, a period 125,000 years ago when the Arctic was warmer than today, left by vegetation that otherwise would have vanished without a trace.

"We feel confident that we are getting authentic results," says Sarah Crump, a paleoclimatologist at the University of Colorado in Boulder who is presenting the work here this week at the annual meeting of the American Geophysical Union. She acknowledges the finding needs to be confirmed. But if it holds up, it could open a window on the ecosystems that flourished in the high Arctic at a time when temperatures were a few degrees warmer than today. It would also attest to the power of sedimentary DNA, as it's called, to show how Arctic plants responded to past climate shifts—hinting at how they might respond in the future. "We are now at the point that this is a really useful signal for reconstructing biodiversity," says Ulrike Herzschuh, a paleoecologist at the Alfred Wegener Institute in Potsdam, Germany, who uses the technique to study how the larch forests of Siberia in Russia reacted after the end of the last ice age some 12,000 years ago.

Although the search for bits of DNA preserved in sediments began 2 decades ago, it has taken off in the past few years as the cost of genetic sequencing has plummeted. Because cold temperatures help preserve DNA, the Arctic has been a prime hunt-



Published by AAAS

CF8, a 40-hectare lake on Baffin Island in Canada, holds a DNA record of the surrounding ecosystem that may go back 125,000 years.

ing ground, drawing geneticists and geoscientists to sample permafrost, cave soils, and other settings, looking for molecular clues to the giant ice age animals that once roamed there (*Science*, 24 July 2015, p. 367).

Lately, Arctic lakes have emerged as the premier archive for sedimentary ancient DNA, because they collect clues to entire ecosystems. Leaves, flowers, dung—some part of every organism that lives around a lake ends up in the water. "You've got an incredibly complicated mixture of DNA there," says Peter Heintzman, a molecular paleobiologist at the Arctic University of Norway (UiT) in Tromsø. DNA is sprung from its cells by decay, then attaches to mineral grains or organic compounds, which provide protection from ultraviolet radiation and oxidation. Temperatures at the lake bottom hover just above freezing, keeping the DNA stable. And year after year, the sediment keeps accumulating, its layers allowing clear dating of the DNA's deposition.

Traditionally scientists have used pollen grains from lakebed cores to study past plant communities. But most plants in the Arctic are pollinated by insects, not by wind, so that little pollen ends up in the soil—and what's there could have blown in from a distance. DNA in lake sediments more accurately reflects the plants and wildlife found nearby. For example, in a study published last month in *Global Change Biology*, Crump and others used ancient DNA to show that, after the planet warmed following the last ice age, dwarf birch arrived on Baffin Island some 2000 years later than researchers had concluded from pollen. "The pollen records are a bit biased," Crump says, "leading us to believe migration was rapid, when it was in fact slower and inhibited by migration barriers."

It's not easy to extract the DNA. Scientists must be careful to collect a sample—typically 1 gram of mud—without contaminating it with modern DNA. Then, in a clean lab, the DNA must be extracted in a painstaking process of trial and error. "There's not the most efficient way of pulling this stuff out," says Beth Shapiro, an evolutionary biologist at the University of California, Santa Cruz. Organic-rich soils seem to be particularly problematic; they are ripe with molecules like humic acid, which behaves like DNA and can foul later sequencing efforts.

To identify plant species in the resulting mélange of DNA—most of which comes from microbes—researchers often turn to metabarcoding, a technique that targets and amplifies a section of DNA that's nearly universal in plants but is bracketed by sections that are distinctive to species. Other researchers simply sequence all the DNA in a sample and sift through the result for genetic gold or use genetic probes that can capture even short strands of plant DNA. But existing databases of plant sequences aren't always reliable enough to identify the species in the sediment. "There were a lot of erroneous sequences uploaded," says Inger Alsos, a paleobotanist at UiT.

To get around that problem, Alsos and her peers have almost finished creating a full genome reference library for Arctic plants, called PhyloNorway, that contains some 2000 species. (A similar effort for the Alps describes 4000 species.) Armed with these improved databases, Alsos and her collaborators are sampling more than three dozen lakes in northern Norway and the Alps, studying how shrubs spread across what had been tundra as the last ice age ended. Another European project, called Future Arctic Ecosystems, will examine a dozen lake cores from the continents ringing the Arctic, studying the dual roles played by climate and large herbivores, such as mammoths, in shaping past plant populations. Other researchers are sampling lakes in southeastern Tibet in China or the Tetons in Wyoming—anywhere cold where DNA could endure.

Early results challenge the simple notion that climate change triggers a wholesale turnover of plant species. "This is very often nonsense," Herzschuh says. Take the Siberian larch forests she studies. Herzschuh found that certain larch species did not shift northward as the planet warmed after the last ice age, despite preferring a colder climate. Instead, she suggests, warming allowed the forest to grow denser, which favored a cold-loving larch. Alsos is studying a core from Lake Bolshoye Shchuchye, at the northern end of Russia's Ural Mountains, that tells a related story. The DNA suggests taller plants invaded as the region warmed: first shrubs, then trees. But Arctic flowers persisted, perhaps by retreating uphill.

Future climate warming, however, may dislodge—or extirpate—these holdouts. For clues to the fate of ecosystems in an even warmer world, researchers want to find other ancient DNA records that stretch back to the Eemian. Herzschuh is hopeful: Her oldest Siberian record goes back 70,000 years so far, and DNA at its bottom is as well-preserved as at the top. "We do not see a clear decaying signal somewhere," she says.

And as the lakes yield even older records, the field is inching toward studying not only the changes in plant diversity and abundance, but also how individual species adapted to climate change, Herzschuh adds. "We are not so far from tracing evolution through time." ■

PHOTO: ZACH MONTES/ORIJIN MEDIA



Published by AAAS

BIOMEDICINE

# Doubts persist for claimed Alzheimer's drug

## Once declared a failure, Biogen's antibody drug to be submitted for U.S. approval in 2020

*By* **Kelly Servick**

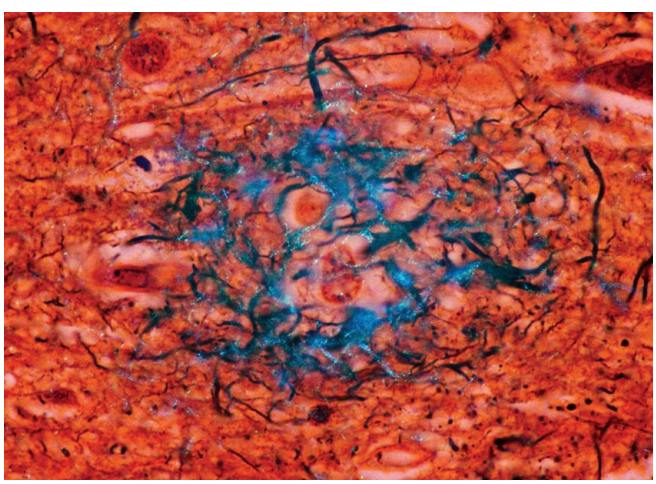
Last week, with trading in the company's stock halted for the widely anticipated event, Biogen gave its first scientific presentation in defense of its startling claim to have developed the first drug that can change the devastating course of Alzheimer's disease. But some scientists and analysts had hoped for more detail, and the community remains divided over whether the drug is a turning point in the quest for an Alzheimer's treatment—or a false hope.

At the Clinical Trials on Alzheimer's Disease congress in San Diego, California, Samantha Budd Haeberlein, Biogen's head of clinical development, tried to clarify what had emboldened the Cambridge, Massachusetts–based biotech to say in October it would soon ask the U.S. Food and Drug Administration (FDA) to approve its drug, aducanumab. That announcement was a striking turnaround for a drug the company had publicly abandoned in March after a discouraging preliminary analysis. But after examining more patient data, she explained, investigators found in one trial that the higher of two tested doses led to 22% less cognitive decline than a placebo after 78 weeks. However, the small margins separating the treated and placebo groups and the failure of a second trial leaves some researchers with a grim outlook.

"I surely don't think that it should be given market approval on the basis of these data," says Robert Howard, a psychiatrist at University College London who has run clinical trials of potential Alzheimer's treatments. By touting positive results from a subset of patients who weren't preselected at the trial's launch and remaining mum about other details of the data, he says, "[Biogen has] broken all the rules, really, about how you analyze data and report it."

Yet some welcome the glimmers of clinical benefits Biogen reported. In a panel discussion at the meeting, Sharon Cohen, a behavioral neurologist at the Toronto Memory Program in Canada and a principal investigator in the Biogen-funded trial, pointed to ratings on a scale of daily activities that suggest aducanumab helped trial participants retain some independence. "Those of us who know this disease well know what it means to lose yourself, slice by slice, and anything you can hang on to and do well is a triumph."

Aducanumab is among the last potential drugs standing that targets beta amyloid, the protein fragment that forms sticky plaques around neurons in the brains of people with Alzheimer's. The failure of several antiamyloid drugs in large clinical trials have suggested that plaque buildup, though a hallmark of Alzheimer's, might be the wrong target for stopping disease progression once people show symptoms.



A new drug is designed to stop buildup of beta amyloid (blue) in the brains of people with Alzheimer's disease, but its clinical benefits remain murky.

Still Biogen and its partner, Tokyo-based Eisai, had high hopes for aducanumab, a monoclonal antibody that binds to and eliminates forms of brain amyloid that are thought to be particularly noxious. In 2015, the companies launched two parallel clinical trials, known as ENGAGE and EMERGE, each testing two different doses of aducanumab against a placebo in more than 1600 people with early-stage Alzheimer's and confirmed amyloid buildup in the brain.

In March, Biogen and Eisai halted the trials after a planned "futility analysis" in a subset of participants suggested the drug wasn't helping. Then came the October surprise: the announcement that there were signs of benefit after all. In both trials, aducanumab had reduced amyloid buildup, Biogen reported, and in the EMERGE study, the high-dose group showed less cognitive decline than the placebo group, based on a standard dementia rating scale. But in ENGAGE, the high-dose group declined slightly more than the placebo group.

Budd Haeberlein tried last week to explain the conflicting results. A major factor, she said, was how the trials treated participants who had a genetic variant called *APOE4*. Those participants had an increased risk of brain swelling—a side effect of antiamyloid antibodies that occurred in about one-third of people getting the high dose and can cause symptoms such as headache, dizziness, and nausea. Patients in both arms with *APOE4* received a reduced amount of the antibody at first, as a precaution, but in 2017, the researchers decided they could safely ramp up.

ENGAGE started about 1 month before EMERGE and had more participants already enrolled at the time of the change. As a result, a smaller proportion of its participants got a full, uninterrupted course of the maximum dose—15% versus 21% in EMERGE, which might account for the lack of overall benefit. When Biogen analyzed the subsets of EMERGE and ENGAGE participants who consented to the change in protocol early enough to get a full maximum-dose course, they saw about a 30% reduction in cognitive decline compared with placebo in both trials.

The need for high, sustained dose of the antibody is "a reasonable explanation," for the results, says Zaven Khachaturian, editor-in-chief of *Alzheimer's & Dementia*. But, "The final answer will come [when] studies with higher doses for longer periods replicate these findings."

Whether Biogen should be allowed to market aducanumab in the meantime, perhaps through a conditional approval that requires a follow-up trial, is hotly debated. Biogen's Chief Medical Officer Al Sandrock drew criticism last month for suggesting that if FDA requires another trial before any approval, it would consign many people to dementia. Agency officials can start that knotty calculus for themselves when Biogen's data hit their desks early next year. ■

PHOTO: CECIL FOX/SCIENCE SOURCE



Published by AAAS

A portion of a narrative scene painted on an Indonesian cave wall shows tiny hunters corralling a dwarf buffalo with ropes or spears.

PHOTO: RATNO SARDI

ARCHAEOLOGY

# Cave painting suggests ancient origin of modern mind

## Half-human, half-animal hunters in mythical scene show early artists in Indonesia had sophisticated imaginations

*By* **Michael Price**

Some 44,000 years ago, an artist climbed high onto a cave ledge on an Indonesian island, paintbrush in hand. Perhaps inspired by spiritual visions, the artist sketched a dynamic scene featuring tiny, animal-headed hunters armed with spears cornering formidable wild hogs and small buffaloes. In a new study, researchers argue that the scene's visionary storytelling—which they claim represents the oldest known figurative art made by modern humans—shows that people already had imaginations much like our own at the time of the cave painting, and likely much earlier.

"We think of the ability for humans to make a story, a narrative scene, as one of the last steps of human cognition," says the study's lead author, Maxime Aubert, an archaeologist at Griffith University in Nathan, Australia. "This is the oldest rock art in the world and all of the key aspects of modern cognition are there."

For the past 5 years, Aubert and colleagues have been exploring dozens of caves on the Indonesian island of Sulawesi and have turned up hundreds of hand stencils, cave paintings, red pigment crayons, and carved figurines. Archaeological data suggest the artists came with an early wave of modern humans some 50,000 years ago. (Modern Sulawesians hail from successive waves of Australasian populations that began to arrive much later, between 3500 and 4000 years ago.)

In 2017, co-author Pak Hamrullah, an Indonesian archaeologist and caver, noticed a small opening in the ceiling of a previously explored limestone cave. Scrambling up a fig tree vine, he found his way into a small grotto. Its far wall bore a panel, painted with a red ocher pigment. When Aubert saw it, he was astounded.

"I thought, 'Wow, it's like a whole scene,'" he says. "You've got humans, or maybe half-human half-animals, hunting or capturing these animals … it was just amazing."

The hunted animals appear to be the Sulawesi warty pig and a small horned bovine called an anoa, or dwarf buffalo, both of which still live on the island. But it was the animallike features of the eight hunters, armed with spears or ropes, that captivated Aubert. Several appear to have elongated muzzles or snouts. One seems to possess a tail, while another's mouth resembles a bird beak.

The features could depict masks or other camouflage, but the researchers argue that dressing like small animals would be a poor disguise for hunters. More likely, the figures represent mythical animal-human hybrids, Aubert says. Such hybrids feature in several instances of early artwork, including a 35,000-year-old ivory figurine of a lion-man found in the German Alps.

To date the Sulawesi cave painting, Aubert carefully pried out a few centimeter-wide shards from the painted cave wall—avoiding the figures and trying to do as little damage as possible—and brought the shards back to his lab. Over the years, as rainwater trickled through the cave's porous limestone and seeped down its walls, it left small mineral deposits called cave popcorn on top of the paint. The popcorn holds trace amounts of uranium, which over time decays into thorium at a fixed rate. By analyzing the ratio of uranium to thorium in the mineral layer directly on top of the pigment, the researchers calculated the painting's minimum age: 44,000 years old, they report this week in *Nature*.

That would make the cave scene at least 4000 years older than other instances of figurative ancient rock art found in Indonesia and Europe, and some 20,000 years older than the oldest depictions of hunting scenes in Europe. In 2018, scientists dated some examples of disks and abstract designs from caves in Spain to 65,000 years ago, but these were attributed to Neanderthals, and some scientists have challenged the dating.

The ability to imagine beings that don't exist is a critical cognitive milestone, Aubert says, and forms the roots of religion and spirituality. Seeing this ability fully formed 44,000 years ago in Sulawesi suggests it was probably already present in the early modern humans who left Africa and populated the rest of the world.

Nicholas Conard, an archaeologist at the University of Tübingen in Germany who wasn't involved in the work, says that scenario makes sense given that every modern human society has its own creative and mythic traditions. "These depictions underline the great antiquity of narratives and storytelling," he says. "It is encouraging to find concrete evidence for narrative depictions at this early date."

The findings should also help dispel the outdated and mistaken notion that humanity first became fully modern in Europe, adds April Nowell, an archaeologist at the University of Victoria in Canada. "We have long known this view is no longer tenable, and the richness of [this and other recent findings] continues to underscore … the importance of the record outside Europe." ■



Published by AAAS

FEATURES

# EPILEPSY'S NEXT FRONTIER

## Almost one in every three patients suffers from seizures that are resistant to drugs. Scientists are looking for new ways to help them

By **Jennifer Couzin-Frankel**

Andrea VonMarkle arrived in Madison by helicopter ambulance 2 years ago, her life hanging in the balance. One month earlier she'd been a healthy 21-year-old juggling community college, waitressing, and weightlifting at a local gym. But after several weeks of feeling vaguely ill and forgetful, she was struck by a terrifying crisis.

On New Year's Eve, VonMarkle's aunt had returned to the home they shared in northern Michigan to find her niece in trouble. "The door was open, and our dog was running down the street," VonMarkle says of the scene that greeted her aunt. "I just kept saying, 'I don't know what's going on, and I don't know why I don't know.'" Then she started seizing.

The seizures, which VonMarkle had never experienced before, didn't stop. Doctors at a local hospital were unable to quell her brain's electrical storm with powerful antiseizure medications. Because unremitting seizures can destroy brain tissue and damage other organs, the doctors put her into a medically induced coma.

"They didn't know what to do with me," she says, "so they flew me to Madison," where the University of Wisconsin hospital had more resources and staff. VonMarkle, unconscious for weeks, wouldn't find out until much later what a stroke of luck that was. She became one of the first people whose sudden-onset, life-threatening epilepsy would be treated in a whole new way: not with standard antiseizure medications, but by disabling the deeper roots of the disease. For her, that meant a drug normally used for arthritis that seemed to soothe the inflammation powering her seizures.

Although most cases of epilepsy don't carry such high stakes, struggles to find a workable treatment play out daily. Afflicting about one in 26 people, epilepsy is defined by its most visible symptom: seizures, caused by abnormal electrical activity in the brain. About two dozen medications are commonly prescribed for epilepsy, but roughly 30% of cases are classified as drug resistant because even with treatment, patients continue to have seizures. That percentage hasn't budged in decades despite many new drugs—suggesting a pivot in tactics is desperately needed.

"People assume that since there's so many medications for epilepsy out there, it's taken care of," says Vicky Whittemore, a neuroscientist and program director at the National Institute of Neurological Disorders and Stroke (NINDS) in Bethesda, Maryland. "It's not."

After years of frustration, epilepsy researchers are shifting from targeting the seizures to seeking their cause, sometimes one patient at a time. Much about the condition remains a mystery, including why anti-

ILLUSTRATION: M. ATAROD/*SCIENCE*



Published by AAAS

seizure medications fail so many people, among them more than 1 million in the United States. Yet leaps in gene sequencing and better animal models are changing how scientists and doctors understand, study, and—sometimes—treat the disease. The patients who stand to benefit include newborns with genetic syndromes who start to seize hours after birth and go on to suffer ruinous developmental delays, 10-year-olds who wake up wondering whether seizures will strike at school and battle difficulties with memory and learning, and adults whose head injuries put them at risk for epilepsy years later (see sidebar, p. 1304).

In San Francisco, California, one lab boasts rows of petri dishes filled with eyelash-size zebrafish larvae; they are genetically engineered to match some of the dozens of epilepsies that strike young children and are used to test new treatments. In Belgium, industry scientists fashioned a medication for one of the most common epilepsies, focal epilepsy, on the basis of experiments in drug and disease biology. And some researchers hope inflammation—the key to VonMarkle's condition—will help them crack other treatment-resistant cases.

"It used to be that people considered epilepsy one big bucket," Whittemore says. "What we now know is there are likely so many different causes, different ways you can get to a seizure." Converting that awareness to targeted treatments is the next frontier.

Andrea VonMarkle was felled by a rare epilepsy syndrome until an arthritis drug subdued brain inflammation, the apparent cause of her seizures.

PHOTO: ADAM JOHNSON/BROCKIT

**EPILEPSY STRIKES NEWBORNS**, the elderly, and everyone in between. Some cases are genetic—researchers have linked more than 100 genes to the disease. Other forms are caused by an infection such as encephalitis or by structural malformations in the brain, whether congenital or due to trauma or a stroke. Often, the cause is unknown. Despite the varied paths leading to epilepsy, all share one endpoint: disruption in electrical missives between neurons, which manifests as seizures.

For about two-thirds of patients, drugs that modify the function of ion channels—proteins in the membranes of neurons and other brain cells that transmit electrical signals—fully restrain seizures. For the rest, drug-resistant epilepsy can be crippling. "Through the day, I would think, 'What if I have a seizure right now, what if I have one in this moment?'" says Kristen Grip, who was almost 14 when she was diagnosed with focal epilepsy, which affects discrete regions of the brain. About 60% of epilepsy patients share this form of the disease. Medications initially quieted her seizures. But when treatment stopped working, seizures—up to 12 per month—punctuated her high school classes and basketball practice. While cycling through medications in a vain effort to find something that worked, she battled drug side effects that included personality changes and exhaustion. "I felt like I wasn't living," she says.

At 16, Grip underwent surgery in which part of her brain was excised—an option because her seizures were concentrated in a region that she could live without. Now 23, she lives in Boston and is seizure-free.

The dilemma for many patients with tough-to-treat epilepsy is that for the most part, "The new drugs do what the old drugs do," says Amy Brooks-Kayal, a pediatric neurologist at Children's Hospital Colorado and the University of Colorado in Denver. "We play whack-a-mole—we've got a bunch of different epilepsies popping out of different holes caused by different things, and we try to whack all of them with the same hammer."

Investigators are searching for better hammers in a gleaming glass and brick building in Salt Lake City, where hundreds



Published by AAAS

of cages line the walls. This is the Epilepsy Therapy Screening Program at the University of Utah, an NINDS effort that since 1975 has used rodents to test more than 32,000 potential antiseizure drugs, including many on the market today.

Like other researchers around the world, the Utah team traditionally studied animals with healthy brains, inducing seizures one at a time with chemicals or electroshock. But as mail carriers dropped off vials of experimental drugs designed by companies and academic labs, Karen Wilcox, the pharmacologist who runs the Utah program, fretted that healthy animals might not help struggling patients. "If we never used brains that had epilepsy, we might never fix this refractory problem," she says.

So about 4 years ago, following recommendations from an NINDS working group, the program was overhauled to include animals that had chronic or spontaneous seizures. Some rodents have infection-induced epilepsy, whereas others have experienced a prolonged seizure, which leaves their brains prone to later spontaneous seizures.

These more sophisticated rodents cost five to seven times as much, because using them is more laborious and time-consuming than traditional animal models. Staff must monitor the animals around the clock and review reams of computer data chronicling their brain waves. "You don't know when they're going to have their seizures," Wilcox says. "They might have two a day, three a day, skip days—like people." Staff test only the most promising drugs on them; scientists are exploring new compounds these animals have helped identify.

**THE REVAMPED RODENT MODELS** reflect growing insight into epilepsy's varied roots. For every patient, "Something happened to this brain that led to this," says Annapurna Poduri, a pediatric neurologist who also runs a lab at Boston Children's Hospital. Poduri witnesses heartbreak all too often in her ninth-floor clinic: severe epilepsies in very young children that respond poorly to treatment. In about 40% of these children, genetic testing reveals a culprit, and in scattered cases, physicians are reaching for unconventional treatments. A toddler in Poduri's clinic received the Alzheimer's drug memantine because in lab studies, it partially corrected the effects of the genetic defect he carried. A little girl went from "seizing, seizing, to walking, running, jumping," Poduri says, after genetic testing revealed a rare mutation whose biochemical effects pointed to the dietary supplement uridine as a potential treatment. Still, such arresting stories are "the extreme," she cautions. "Maybe one in 100 you find one of these."

To move beyond anecdotes, neuroscientist Scott Baraban at the University of California, San Francisco, is hatching thousands of genetically engineered zebrafish. Members of each clutch carry an epilepsy-causing mutation. Baraban is especially interested in one genetic epilepsy, Dravet syndrome. It triggers prolonged seizures, developmental delays, and a host of other problems, and it is profoundly resistant to antiseizure treatments. Baraban was startled to find that an old antihistamine called clemizole blunted seizures in a fish model of Dravet, and he and his colleagues found that it worked by binding to serotonin receptors that mediate neuron excitability.

> **"What we now know is there are likely so many ... different ways you can get to a seizure."**
>
> **Vicky Whittemore**, National Institute of Neurological Disorders and Stroke

Clemizole was discontinued in the United States and elsewhere in favor of newer antihistamines, but a drug for weight loss, lorcaserin, binds to the same receptors and had the same helpful effect on Baraban's fish. The discovery surfed straight "from aquarium to bedside," Baraban says, as physicians offered the drug to five patients. Seizures were considerably diminished in four of them. A company Baraban co-founded is redeveloping a pediatric version of clemizole and recently finished a small clinical trial of the drug in Dravet.

Dravet is also the target of an eagerly awaited therapy, slated for a trial opening next year. The treatment, a form of RNA called antisense, was developed by Stoke Therapeutics in Bedford, Massachusetts, and scientists at the University of Michigan in Ann Arbor. Most children with Dravet have one normal copy of the culprit gene and one defective copy; in mice, the therapy binds to RNA. In doing so, it increases production of RNA and the healthy protein, compensating for the underlying defect.

"We're just praying these treatments have good results, they're safe, and they're available in time for Anna," says Kim Odlaug, an advertising executive in Chicago, Illinois, whose 2-year-old daughter has Dravet. In children with Dravet, developmental delays often surge between the ages of 2 and 5, and Odlaug worries about what's to come for Anna, who performs acrobatics on the family's furniture and loves assembling puzzles. Because most of Anna's more than 30 seizures—three of which left her on a ventilator—trace to excitement or illness, her parents isolate her from other children and activities that she loves, such as music class. Odlaug hopes that if the therapy appears safe, Anna can receive it in a clinical trial, because approval is uncertain and could take years.

But what of the many children with epilepsy for whom gene variants are suspected but DNA panels come back clean? Poduri and others are studying brain tissue from some of those children, usually donated after surgery. Some samples harbor DNA changes that don't appear in the blood and must have happened spontaneously during development. Such changes are harder to identify but may prove powerful in steering treatment, especially if they're shared across patients. Surprise discoveries include mutations in tumor genes and in a gene that helps sugar molecules bind to proteins.

Recently, Poduri began to collaborate with colleagues in Melbourne, Australia, who are assessing whether such DNA changes are detectable in spinal fluid. "Precision medicine," she says, "has to start with precision diagnosis."

**THAT'S NOT ALWAYS POSSIBLE TODAY**. Many cases of epilepsy can't be traced to an easily treatable cause, leaving patients like Grip to try drug after drug. The inability to hit on a precision diagnosis for many patients has led some drug company scientists to try another tack: understanding why the drugs they make don't always work and figuring out how to improve on them.

In 1999, a medication called levetiracetam (Keppra) was approved for some cases of focal epilepsy. But the drug doesn't help everyone, and scientists at the company that makes it, UCB in Braine-l'Alleud, Belgium, wanted to understand why. Experiments revealed that levetiracetam selectively binds to a protein called SV2A, which belongs to a family of proteins called SV2; the drug's binding affects release of neurotransmitters in the brain, which in turn can stanch seizures. Collaborating with scientists at the University of Liège in Belgium, the UCB team learned that another protein in the family, SV2C, is markedly elevated in some brain regions of patients with drug-resistant focal epilepsy. Months of study yielded another clue: Although levetiracetam prevents errant release of neurotransmitters, the drug was more potent when combined with compounds that target certain neurotransmitters after they're pumped out.

With that knowledge in hand, UCB sci-



Published by AAAS

entists created padsevonil. The experimental drug acts both before and after neurotransmitters are released between nerve cells, and it targets the effects of all three members of the SV2 protein family. The company is now running two phase III trials, involving more than 1000 adults with drug-resistant focal epilepsy. "I'm not saying it's the solution to drug resistance," says Henrik Klitgaard, who consults for UCB, where he was a vice president and led epilepsy drug research for several years. But it's a "rational approach."

Other leads come from evidence that inflammation can drive some forms of the disease. Powerful anti-inflammatories such as steroids occasionally ease seizures in drug-resistant epilepsy. The ketogenic diet, an extremely high-fat, low-carbohydrate diet, can sometimes help, as well. Some researchers theorize that the diet works in part by subduing inflammation.

Two scientists pushing the epilepsy-inflammation connection forward are Eleonora Aronica and Annamaria Vezzani. In the late 1990s, Aronica, a neuropathologist at the University of Amsterdam, and Vezzani, a neuropharmacologist at the Mario Negri Institute for Pharmacological Research in Milan, Italy, grew intrigued by the presence of inflammatory molecules in brain cells in affected animals and people. Brains of fetuses that had died and had a genetic disease called tuberous sclerosis complex, which carries about an 80% chance of epilepsy, showed inflammation. So did the brains of animals with induced seizures that Vezzani was studying. When inflammatory molecules persisted in an animal's brain, they made future seizures more likely. After years of work, the pair learned that the inflammatory molecules increase neurons' excitability.

Could those links to inflammation inspire new treatments? Four years ago, a chance encounter allowed Vezzani to test the idea. At a conference, she met a doctor from the Mayo Clinic in Rochester, Minnesota. The hospital was then trying to save a toddler whose disease looked awfully like that of Vezzani's mice—the little girl had unyielding seizures that began after a fever and infection. The child's diagnosis was an extremely rare but catastrophic form of epilepsy called febrile infection-related epilepsy syndrome (FIRES), which strikes about one in 1 million people.

When asked, Vezzani recommended that doctors try an anti-inflammatory drug called anakinra, which is approved for rheumatoid arthritis. In the rodents she'd studied with brain inflammation, anakinra minimized seizures.

It worked in the little girl, too. "We had an outcome that we've never seen before," says Eric Payne, a Mayo pediatric neurologist who treated the child. She turned 7 in September, and a mild language delay and occasional seizures are the only signs of a brain that nearly destroyed itself.

Two years later came VonMarkle. "I heard about this woman in the adult neurocritical care unit who was seizing, seizing, and wouldn't respond," says David Hsu, a pediatric neurologist at the University of Wisconsin. He had trained with Payne and knew about the girl. Hsu snagged VonMarkle's medical chart, which revealed that her tale began with headaches and a fever, and she, too, was diagnosed with FIRES. At Hsu's urging, VonMarkle was started on anakinra. She had been in an induced coma for more than 1 month, but within 24 hours of starting the drug, her seizures stopped. Within another day she was weaned out of the coma.

**A THERAPY SUCH AS ANAKINRA** embodies epilepsy's next chapter: treatments that, though not always fully understood, seem to

## The faces of epilepsy

Epilepsy is defined by its most visible symptom—seizures—but varies in its causes and manifestations. Almost one-third of patients continue to suffer seizures despite taking medication.

### Varying causes

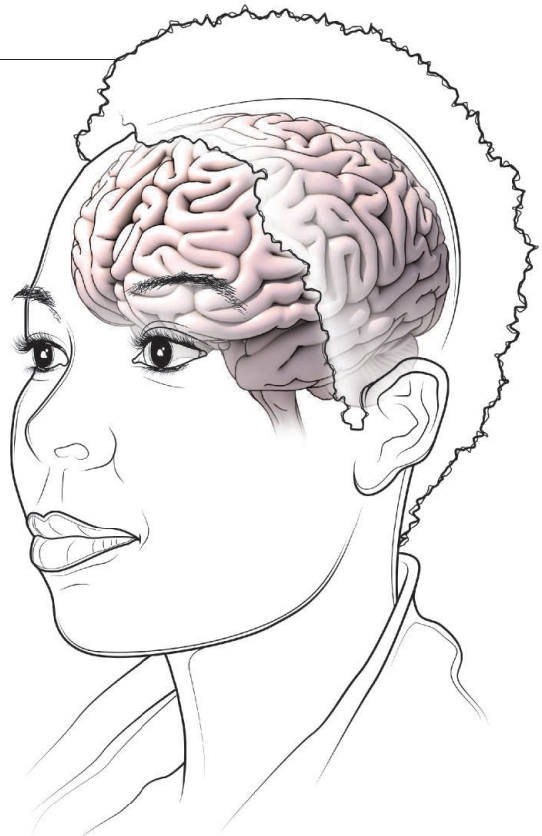
Scientists are working hard to understand epilepsy's many mechanisms in hopes of developing more targeted treatments.

1. **Genetics**
More than 100 genes have been linked to epilepsy, and some variants cause devastating symptoms in children.

2. **Infection**
Infections such as encephalitis can sometimes trigger epilepsy, perhaps by inducing brain inflammation.

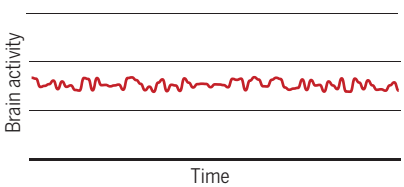
3. **Injury**
Traumatic brain injury and other forms of brain damage, such as stroke, can cause epilepsy.

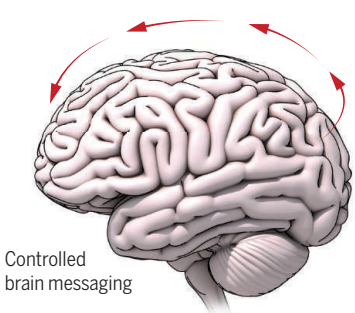
4. **The unknown**
In about half of cases, the cause is a mystery.



### Normal firing

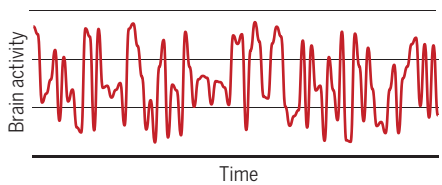
Neurons relay messages to one another when electrical impulses trigger neurotransmitter release. Neurotransmitters encourage other neurons to fire or hinder their firing, keeping brain activity in check.

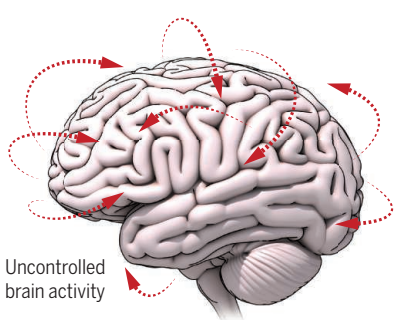



### Misfiring

An imbalance in neurotransmitter signaling causes neurons to become overexcited or lose their inhibitory control, causing abnormal bursts of electrical activity, resulting in a seizure.




GRAPHIC: C. BICKEL/*SCIENCE*



*Published by AAAS*

# Foiling epilepsy in a brain at risk

*By* **Jennifer Couzin-Frankel**

When Greg Parks was found unconscious 5 years ago with his bicycle lying nearby, the victim of a likely hit and run, his wife had an inkling of the arduous road ahead. "His brain had sheared away from his skull," says Kathleen Pullen-Norris, a neurotrauma nurse in the intensive care unit where her husband was treated, at the University of California, Los Angeles (UCLA). A helmet had saved the 43-year-old's life, but barely.

Parks, an engineer and former triathlete, spent weeks in the hospital learning to how to eat, walk, and talk as he had before the accident. As he healed, a shadow hovered: the risk of seizures. About 30% of people who suffer a severe traumatic brain injury (TBI) like Parks's—and a smaller percentage of people with a milder one—later develop epilepsy.

TBIs can cause "a rewiring of the brain, going from normal to short-circuited," says Paul Vespa, a neurologist and neurocritical care specialist at UCLA who cared for Parks. But the details are hazy. There's "the brain injury, something happens in the middle, and then 20 years later, seizures," Vespa says. "Nobody understands what the 'middle' is."

A flurry of research is probing that arc to epilepsy, aiming to find ways to interrupt it. "TBI offers us a specific point in the patient's life that we can identify as the potential start of the disease," says Aristea Galanopoulou, a neurologist and neuroscientist at the Albert Einstein College of Medicine in New York City. With that goal in mind, 3 years ago, a $21 million project spanning five countries took the first steps toward prevention for brain-injured patients: searching for biomarkers that presage epilepsy and, in animals, drugs that prevent it.

Doctors already know that patients who have seizures in the days after their injury are at higher risk of epilepsy later. In a 2018 study in *Neurobiology of Disease*, researchers followed 90 people with moderate to severe TBIs for at least 2 years to quantify risk factors. Among the 21 people who developed epilepsy, 38% had seizures shortly after injury, sometimes detected only by electroencephalography (EEG) monitoring. Another factor also emerged: Eighty-six percent had hemorrhaged in their temporal lobe, which controls hearing and language.

Vespa worked on that study and is one of the leaders of the multinational consortium. His center is among 13 now enrolling 300 patients hours after a severe brain injury—with the consent of their families—for 2 years of follow-up in a hunt for epilepsy biomarkers. The scientists are using MRIs to probe structural brain damage, EEGs to look at brain wave changes, and blood proteins that indicate cell damage, inflammation, and other disruptions.

**Greg Parks (right), pictured with his wife Kathleen Pullen-Norris, sustained a traumatic brain injury, which raises the risk of epilepsy.**

Scientists in the consortium, co-led by Galanopoulou, are studying potential preventives, too. One target of interest is the tau protein. A buildup of tau has been linked to Alzheimer's disease as well as chronic traumatic encephalopathy, which has devastated the brains of football players. Inflammation—already implicated in some cases of epilepsy not linked to trauma—is also being investigated as a potential target. On the basis of animal studies, Vespa believes brain bleeds trigger inflammation, which can irritate nerve cells and leave them firing inappropriately.

In the long run, physicians treating TBI could follow the lead of the first epilepsy prevention trial, in babies with a rare genetic syndrome. Tuberous sclerosis complex causes benign tumors in the brain and elsewhere, and epilepsy in more than 80% of patients. Earlier work identified EEG anomalies that often presage full-blown seizures. In late 2016, researchers led by pediatric neurologist Martina Bebin at the University of Alabama in Birmingham began to enroll the first of 80 infants. Those that develop the EEG anomalies are randomly assigned to receive either the epilepsy drug vigabatrin or a placebo, to test whether the drug can delay or prevent epilepsy—and whether doing so improves children's neurodevelopment. The medicine is the gold standard for treating tuberous sclerosis patients who develop epilepsy.

The goal is to trace a similar path for TBI. "We need something that would give us a readout on biological processes," to predict epilepsy before it happens, says Henrik Klitgaard, who led epilepsy drug research for several years at UCB, a pharmaceutical firm in Braine-l'Alleud, Belgium, and now consults for the company. Then doctors could give a preventive only to patients who need it most.

Parks, who volunteered for a pilot study to help scientists plan the consortium, remains seizure-free. He focuses on managing lingering effects of the accident, such as sleep disturbances, not contemplating the unknown. For Pullen-Norris, the risk of epilepsy nags, but she knows she can't let it consume her. Instead, she considers Parks's remarkable recovery and hopes for healthy days ahead.

disable the deeper mechanisms of disease in narrow groups of patients.

For anakinra, mysteries endure. The little girl Payne cared for produced a defective version of an immune protein, which anakinra supplemented, he and his colleagues later reported. That finding might help explain the drug's potency for her. VonMarkle's doctors haven't determined whether she shares this aberration or harbors another—or whether the drug's lifesaving effects for her will linger unexplained. "I was ecstatic" the drug helped her, Hsu says, but "we don't know why." At least, not yet.

Worldwide, an unknown number of children have received anakinra, including at least 25 with FIRES who are part of an unpublished cohort and have experienced varied results. Payne is trying to understand whom anakinra can help and why, and he is hunting for neuroinflammation in other tough-to-treat cases of epilepsy. He and others caution that although inflammation is worth chasing, no single drug will be a cure-all.

VonMarkle was among the lucky, though it didn't feel that way at first. After waking up, she weighed just 43 kilograms and spent another 6 weeks in the hospital. Nearly 2 years later, her short-term memory is sluggish and she hasn't returned to college, but she is waitressing, exercising, and working as a makeup artist.

She still has occasional seizures but finds them manageable. "Just reassimilating to normal life was very difficult," VonMarkle says. But, she adds, "I'm very fortunate to be able to do what I'm doing." ■

PHOTO: KATHLEEN PULLEN-NORRIS



Published by AAAS

INGENUITY: NEXTGEN'S VISION

# Foods of the future

We asked young scientists to write an advertisement that answers this question: **How will food options, food availability, and individuals' food choices change in the future?** A selection of their suggested marketing campaigns is below. Follow NextGen Voices on Twitter with hashtag #NextGenSci. Read previous NextGen Voices survey results at https://science.sciencemag.org/collection/nextgen-voices. **—Jennifer Sills**

## Health food

Tired of managing your diet? Health Capsule provides non-invasive, cognitive control of your hunger, satiation, and weight. Made possible by deep brain stimulation and neuromodulation technology, this environmentally sustainable capsule will keep you healthy and fit while satisfying all your cravings. Put calculating your calories and carbon footprint behind you!

**Saima Naz**
Lahore, 54570, Pakistan.
Email: saimanaz85@gmail.com

Try our new Bacteri-Os: The Sustainable Microbial Sustenance™! We start with the freshest plasmids and gene fragments, harvested from diverse biospheres and raised organically and sustainably. Then, we splice and dice these into stable, non-infective carrier organisms, ensuring production of all secondary metabolites and metabolic fuels growing eukaryotes might need: yeasty vitamins, siderophore mineral sponges, and essential amino acids. A little sprinkle of HMG-CoA adds your fats and sterols, et voila! Buy your Bacteri-Os pellets wherever fine nutritional cakes and shakes are sold.

**Michael A. Tarselli**
SLAS, Oak Brook, IL 60523, USA.
Email: tarselli@gmail.com

## Personalized meal plans

Tired of your 3D-printed food all tasting the same? Make your fab food fabulous! Our FlavorPack™ inks will make your proteins meatier, desserts sweeter, and veggies crunchier. Compatible with any home fab unit!

**Mark Martin Jensen**
Massachusetts General Hospital, Department of Surgery, Harvard Medical School, Boston, MA 02114, USA. Email: MMJensen@MGH.Harvard.edu

Send us your DNA, and we will predict your food preferences! Receive your personalized food basket, with a day-by-day diet program. We will send you full meals and personalized smoothies based on your genetic taste predisposition. We know what you love; it's in your DNA.

**Ada Gabriela Blidner**
Laboratorio de Inmunopatología, IBYME-CONICET, 1428 Ciudad Autónoma de Buenos Aires, Argentina. Twitter: @adagbb

Introducing Broth-3D technologies, by machine-learning chefs! Broth recipes use boiled local and seasonal ingredients. Miss your grandmother's recipes and the beautiful plating of the past? We've uploaded images of meals created by 33,003 grandmothers and distilled their knowledge into 3D-printable dishes using broth solidifiers. Add nostalgic structure to your dish!

**Athanasia Nikolaou**
Department of Physics, Sapienza University of Rome, Rome, Italy.
Email: athanasia.nikolaou@protonmail.com

ILLUSTRATION: LEONARD DUPOND



Published by AAAS

ILLUSTRATION: ADAM SIMPSON/HEART AGENCY

## Meals on the go

Convert polluted ambient air into basic nutrients with Smart Straw! Sensors in this nano-technological device identify potentially nutritious particles suspended in the air around you. Just take a sip, press the Convert button, and imbibe basic nutrients through chemical reactions and filtering.

**Emre Ozan Polat**
ICFO–The Institute of Photonic Sciences, Castelldefels, Barcelona, 08860, Spain.
Email: emre-ozan.polat@icfo.eu

With intravenous (IV) nutrients, you no longer need to make time for trivial things like sitting down to a full meal! Plus, our targeted modulation of the hunger hormone ghrelin means no more hunger pangs. Don't be tied down by mundane human concerns like eating and weight gain—make time for more with IV nutrients!

**Mallory Peterson**
Penn State College of Medicine, Hershey, PA 17033, USA. Email: mpeterson1@pennstatehealth.psu.edu

## Meat

Beeph™ is the newest, tastiest insect patty! If you're old enough to remember Kobe, you're in for a treat! Made with pure, farm-raised Periplaneta and synthetic spice blend for that wholesome goodness your family loves. When you're hungry for burgers, think Beeph™—It's what's for dinner!

**Donnelly A. West**
Walnut Creek, CA 94597, USA.
Twitter: @catscausetypos

Sustainable Steaks provides the most scrumptious, environmentally sensitive beef. Our steaks are sourced from natural grasslands and prairies, where livestock now fills the ecological roles left vacant by the wild grazers that used to roam these plains. Using our open-source FoodFootprint supplementary feeding system, we enhance natural grazing with protein-rich fodder to optimize animal growth while minimizing methane production and water consumption. At only $219.00/kg (including carbon taxes and ecological levies), our steaks are affordable for the whole family.

**Falko Buschke**
Centre for Environmental Management, University of the Free State, Bloemfontein, 9300, South Africa. Email: falko.buschke@gmail.com

## Fresh fruit

As Earth's temperatures rose, temperate zones were lost. Do you miss sweet temperate fruits? Welcome to our Moon Farm. Our fruit trees are planted in hybri-soil and artificial air that simulate Earth's environment from 5000 years ago. Pick fruits with your family or ship to your doorstep with MoonEx. Freshness guaranteed.

**Yongsheng Ji**
Department of Human Parasitology, Anhui Medical University, Hefei, Anhui, 230032, China.
Email: jiyongshengkey@hotmail.com

West Siberian Citrus Company offers a bountiful array of lemons, oranges, grapefruits, and limes. Once a permafrosted no-man's land, Siberia is now a gorgeous Garden of Eden. With no deep freeze days during the growing season, we guarantee that our fruits have attained maximum size and flavor upon harvest. For a limited time only, get your dose of Vitamin "Sea" with our genetically engineered Florida-style oranges; they accurately replicate that 2023 pre-flood tang! Voted "Better Flavor" than both Alaskan and Nunavut Oranges.

**Kyle J. Isaacson**
Biomedical Engineering, University of Utah, Salt Lake City, UT 84112, USA. Twitter: @kjisaacson

Want to feel like a globalist glutton from the greedy years of centuries past? Cicada Steaks let you relive the joy and horror of eating fried animal flesh! The new recipe of textured Insekt™ protein, marbled with saturated fats, evokes steaks Grandma will remember from her childhood when people ate cows.

**Matilda Newton**
University of Colorado, Boulder, CO 20309, USA.
Email: matilda.newton@colorado.edu

Do you long for the days before the Great Climate Shift when meat wasn't only for special occasions? Are you fed up with paying outrageous beef and pork prices? Well don't have a cow! Why pig out when you can GRUB! out. GRUB! (Great Rations Using Bugs!) is a delicious, nutritious insect protein that is packed full of vitamins and minerals. Whether it's GRUB! burgers on the grill or GRUB! taco Tuesdays, make GRUB! part of your family's dinner plans.

**Eric Britt Moore**
Department of Agronomy, Iowa State University, Ames, IA 50011, USA. Email: ebm256@iastate.edu

## Carbs and starches

Using the precision breeding technology CRISPR, scientists have developed a high-yield, disease-resistant cassava. Cassava, a historically crucial staple for millions of people, is now an even more reliable food source, providing food security to the developing world where it is needed most.

**Syed Shan-e-Ali Zaidi**
Gembloux Agro-Bio Tech, University of Liège, Gembloux Namur, 5030, Belgium.
Email: shan.e.ali@outlook.com

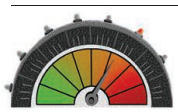
Our dedicated team of food engineers has pioneered a technique to convert the chemical bonds in wood, turning indigestible cellulose into a scrumptious starch. The result? A rustic delicacy ushering in the new era of staple wood-based foods. Try our signature, honey-coated wood chips at a Trader Joe's near you!

**Divyansh Agarwal**
Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA.
Twitter: @divyansh_aga

Introducing high-energy, nutrient-rich Wherice, a unique mix of wheat and rice! Just 100 g of this tasty, digestible food fulfills your energy requirements for the day, saving you time and money. Sustainable Wherice can be cultivated in high-temperature areas of the world three times a year.

**Sudhakar Srivastava**
Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, Uttar Pradesh, 221005, India.
Email: sudhakar.srivastava@gmail.com

10.1126/science.aba2238



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



*Published by AAAS*

PERSPECTIVES

NANOMATERIALS

# Two-dimensional polymers grow up

## Silicon wafer–sized single-layer films are synthesized with a new method

By **Oliver MacLean** and **Federico Rosei**

Creating tunable and highly conductive two-dimensional polymers (2DPs) using organic chemistry has been a goal since the isolation of graphene (*1*, *2*). The high surface area and well-defined pore sizes make 2DPs attractive for a range of applications, including sensors, electronics, catalysis, energy storage, and energy conversion. Conjugated 2DPs, which have alternating single and double bonds that enable electron delocalization, are promising materials for optoelectronics. They have high predicted carrier mobilities and semiconducting behavior that contrasts with the metal-like properties of graphene (*3*). Challenges in scaling up from the laboratory have prevented different types of 2DPs from making the leap to practical applications. On page 1379 of this issue, Zhong *et al.* (*4*) report a synthetic method that solves many difficulties related to isolating and processing 2DPs and fabricate a simple large-area electronic device.

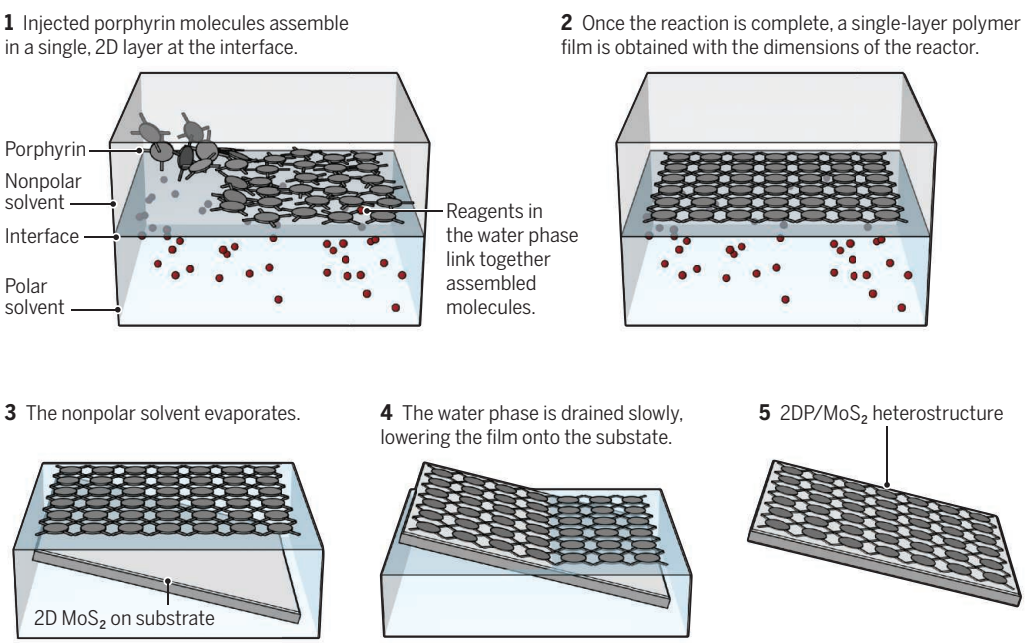
Several strategies have been pursued to produce single-layer 2DPs. One approach starts with a bulk 2DP crystal that is chemically exfoliated into polymer flakes, although this yields a range of thicknesses and reduced quality. The bulk crystal can be synthesized following either the covalent organic framework (COF) paradigm, which applies reversible chemistry to attain a highly ordered product (*5*), or by solid-state topochemical polymerization, where precursor molecules within a crystal are aligned to directly form ordered 2D layers (*6*). Alternatively, single-layer 2DPs can be produced directly by confining polymerization on a single-crystal surface. Conjugated polymers have been synthesized on catalytic metal surfaces in a vacuum (*7*), but usually with nanometer domain sizes. Recently, a micrometer-sized organometallic polymer was synthesized (*8*). Larger 2DPs with highly ordered hundred-nanometer–sized domains have been produced with the dynamic COF approach at liquid-solid interfaces. However, this has been accomplished for only a limited family of polymers (*9*). In addition, 2DP synthesis on crystalline surfaces generally requires a subsequent, nontrivial, step of transferring the polymer to an insulating substrate for device fabrication.

In a third category, polymers are formed at a liquid-liquid interface, where the films can grow uninterrupted to the size of inches. A sharp interface between the two liquid layers is formed by using two immiscible fluids. Adding planar amphiphilic molecules, with both hydrophilic and lipophilic groups, to one layer causes them to form a 2D assembly at the interface. This is linked into a polymer by a reagent soluble only in the other phase. The second key advantage of this approach is that the 2DP is easy to transfer to a substrate by simply removing the liquid slowly (see the figure).

Zhong *et al.* developed an interfacial synthesis method, laminar assembly polymerization (LAP), that produces uniform single-layer films at the inch scale. This scale is the same as that of silicon wafers, and the method is compatible with existing patterning and integration techniques. They synthesized porphyrin-based covalent and coordination polymers that are the 2D analogs of COFs and metal-organic frameworks, respectively. The LAP technique injects amphiphilic precursors using microsyringe pumps to form 2D structures by self-assembly, so multiple pumps could produce larger films or stripes with tun-

### Synthesizing and scaling two-dimensional polymers

Centimeter-sized 2D polymers (2DPs) are fabricated with a multiple-step process to fit neatly on substrates. The chemical reaction occurs at the interface of immiscible liquids, which are siphoned off later to deposit the film.




GRAPHIC: VERONICA FALCONIERI/*SCIENCE*

Centre Énergie, Matériaux et Télécommunications, Institut National de la Recherche Scientifique, 1650 Boul. Lionel Boulet, Varennes, Québec J3X 1S2, Canada.
Email: rosei@emt.inrs.ca



Published by AAAS

able widths. This method therefore offers improved scalability and patterning over the air-liquid polymerization previously used to achieve a wafer-scale covalent 2DP (*10*). Beyond stripes, the film can be arbitrarily patterned using a laser to selectively remove material. The authors used scanning electron microscopy and atomic force microscopy to establish the uniformity and single-layer nature of the polycrystalline films. The films were sufficiently robust to be suspended across 2-μm holes in a metallic grid.

The intense interest in 2D materials stems from their remarkable intrinsic properties and the potential to layer them into tailor-made structures. These so-called van der Waals heterostructures may give rise to intriguing phenomena that are not present in each individual material (*11*). Zhong *et al.* produced heterostructures by transferring 2DP films onto stacked $MoS_2$ layers. These films were used to fabricate a series of capacitors, demonstrating the potential of their approach to realize new active materials for electronic devices.

The scalability and technical compatibility of the LAP method promises exciting new developments in 2DPs. Porphyrin-based 2DPs have numerous possible applications (*12*), yet the LAP method could be extended to other planar, amphiphilic precursors and different linkage chemistries. One target would be to achieve fully conjugated 2DPs with high charge mobilities. Better optimizing the crystallinity of the films will be important for improving the distinctive properties of van der Waals heterostructures and more generally for electronics and optoelectronics applications. The advent of wafer-scale 2DPs opens up an array of opportunities in photonics and other emerging technologies, promising to continue the revolution sparked by graphene. ■


### REFERENCES AND NOTES

1. D. F. Perepichka, F. Rosei, *Science* **323**, 216 (2009).
2. J. Sakamoto, J. van Heijst, O. Lukin, A. D. Schlüter, *Angew. Chem. Int. Ed.* **48**, 1030 (2009).
3. M. Ebrahimi, F. Rosei, *Nature* **542**, 423 (2017).
4. Y. Zhong *et al.*, *Science* **366**, 1379 (2019).
5. A. P. Côté *et al.*, *Science* **310**, 1166 (2005).
6. W. Liu *et al.*, *Nat. Chem.* **9**, 563 (2017).
7. L. Grill *et al.*, *Nat. Nanotechnol.* **2**, 687 (2007).
8. Y.-Q. Zhang *et al.*, *J. Am. Chem. Soc.* **141**, 5087 (2019).
9. R. Dong, T. Zhang, X. Feng, *Chem. Rev.* **118**, 6189 (2018).
10. H. Sahabudeen *et al.*, *Nat. Commun.* **7**, 13461 (2016).
11. K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. Castro Neto, *Science* **353**, aac9439 (2016).
12. J. Tian, W. Zhang, *Prog. Polym. Sci.* **95**, 65 (2019).

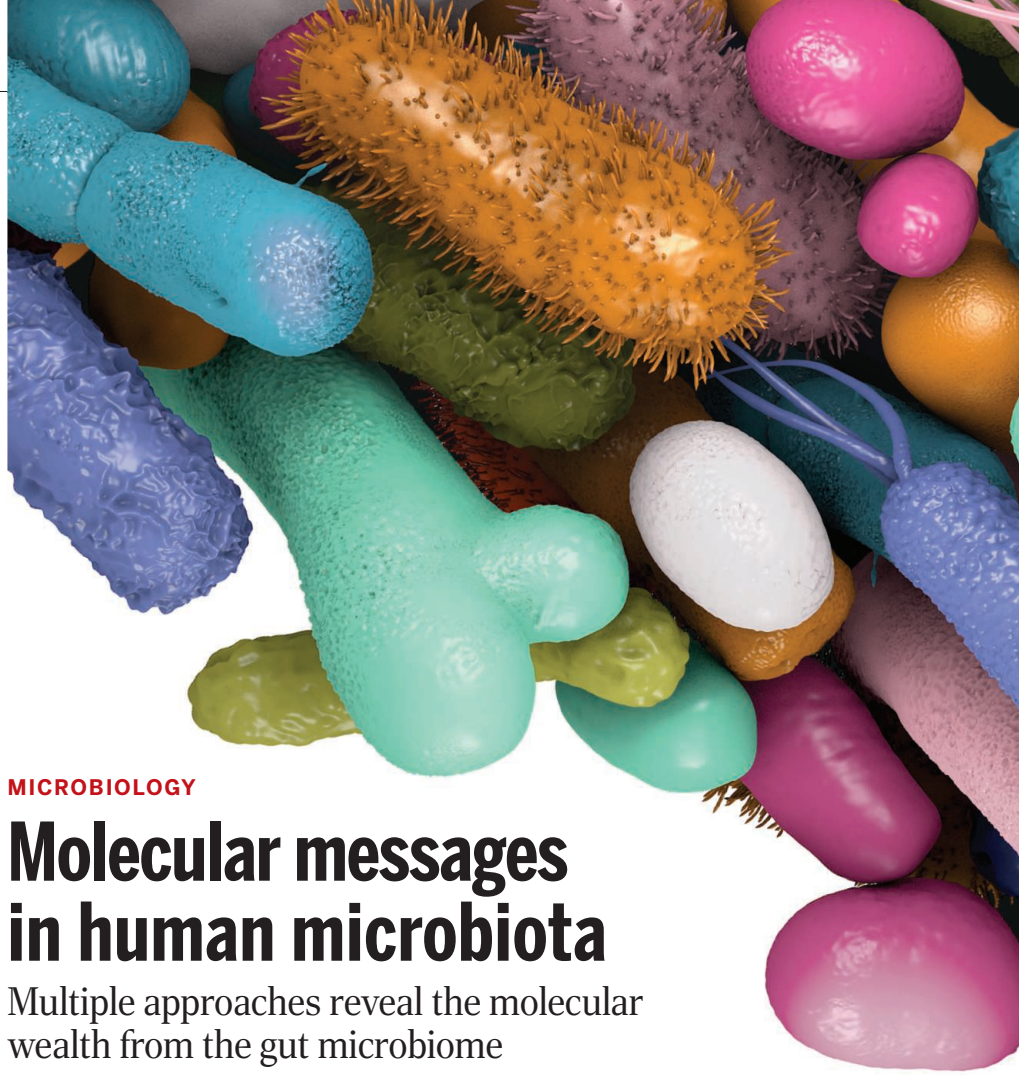


### ACKNOWLEDGMENTS

F.R. thanks the Canada Research Chairs Program for partial salary support.

10.1126/science.aaz9326


ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*



MICROBIOLOGY

# Molecular messages in human microbiota

## Multiple approaches reveal the molecular wealth from the gut microbiome

*By* **Matthew T. Henke** *and* **Jon Clardy**

The roughly 1 trillion bacteria typically living in and on a human sense and respond to their environment with diffusible small molecules. Some of these molecules also affect their human hosts or neighboring microbes. There are numerous studies correlating bacterial populations, or changes in bacterial populations, with health or disease, but only a few of the molecules and mechanisms that underlie these correlations have been identified (*1*–*3*). Two systematic approaches to identify these bacterially produced molecules and their associated functions are described in this issue. On page 1332, Sugimoto *et al.* (*4*) describe how bioinformatic analysis of metagenomes—the collective genes of a microbial community—can reveal previously unknown bacterial metabolites that may function as antibiotics. On page 1331, Guo *et al.* (*5*) describe a gene-deletion approach that discovered the ability of known bacterial metabolites to affect human immune responses.


Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Cambridge, MA, USA. Email: jon_clardy@hms.harvard.edu


Researchers have access to truly staggering amounts of DNA sequence data from bacterial genomes and, increasingly, metagenomes. These data can reveal the biosynthetic potential of the gut microbiota (*6*). In this sense, human microbiome research is roughly at the stage reached by the human genome project in 2003—there is a good grasp of the genes involved. But there are a hundred times more genes distributed over hundreds of different kinds of bacteria found in or on a typical human. Unscrambling which bacteria are making what molecular messages, and to whom they are addressed, is a massively complex task. These bacterially produced molecules can be addressed to the host or to other microbes in the environment. Both studies investigated this challenge in different ways. Sugimoto *et al.* describe molecular messages, likely between bacteria, and discovered these metabolites by searching for newly described biosynthetic gene clusters related to known antibiotics. Guo *et al.* sought to assign host functions to known microbiota-produced metabolites.

Both of these approaches required the ability to manipulate bacterial genes, and the single greatest impediment to




*Published by AAAS*

mechanistic microbiome research is the lack of broadly applicable genetic tools for virtually all members of the human gut microbiota. The ability to delete a gene, to reintroduce a gene, or to confer a gene's proposed role to another organism is essential for rigorous mechanistic studies. Several groups, including Guo *et al.*, have developed tools that are applicable to a subset of the microbiota; the method of Guo *et al.* is promising for the genus *Clostridium*, which includes pathogens such as *C. difficile* and *C. botulinum*. Although developing a tool kit for each individual or bacterial group of interest (*7*) seems both tedious and daunting, it might be the best path forward, because developing broadly applicable genetic tools has been difficult. Another important obstacle to mechanistic microbiota research is the complexity of the host-microbiota interactions. Differences in human genetics, the spatial organization and temporal dynamics of an individual's microbiota, the variability of microbiomes between people and between an individual in different contexts, and functional redundancy within microbiomes all obfuscate the roles that individual molecules may play.

Sugimoto *et al.* extended a previous study (*8*), which parsed complete genomes in databases to identify the distribution of biosynthetic genes into various classes, to use metagenomic data that have not been sorted into individual bacterial genomes. To illustrate the technique, they focused on an easily identifiable feature of a widely distributed pathway—an essential domain found in type II polyketide synthase gene clusters, which produce drug-like molecules that share an evolutionary origin with fatty acids. When metagenomic data were found to contain this feature, the authors could selectively assemble related sequences into a complete pathway. In this manner, they found 13 previously undescribed polyketide biosynthetic gene clusters. They found that almost all people harbor bacteria with at least one of these gene clusters, that different gene clusters are found in different body sites (that is, skin, gut, mouth, or vagina), and that these gene clusters are found at varying frequencies in different groups of people. When two of these assembled pathways were expressed recombinantly, one from an oral microbe and one from a gut microbe, they produced molecules similar to clinically used antibiotics and anticancer drugs. They suggested that these molecules might play a role in interspecies competition in the gut and oral cavity and that their influence on community structure could affect host biology. Crucially, soil microbes have been an invaluable source of clinically used small-molecule drugs; the findings of Sugimoto *et al.* underscore that the human microbiota are an underexplored resource for drug-like molecules.

Guo *et al.* developed a CRISPR-Cas9 gene-editing system for *C. sporogenes*, a microbe that widely colonizes the guts of many animals. *C. sporogenes* synthesizes high amounts of several molecules that are produced only by the microbiota. Many of these molecules are found at high amounts in people, suggesting that they could affect human biology. Using their genetic system, Guo *et al.* interrupted genes in eight biosynthetic pathways that produce 10 microbiome-derived molecules. Satisfyingly, interrupting (knocking out) each of these pathways abolished production of the corresponding metabolites, whereas the other metabolites remained unperturbed. Next, Guo *et al.* introduced these wild-type and knockout strains of *C. sporogenes* into germ-free mice, which lack any other microbe. This allowed the assessment of the physiological contribution of a single molecule from a single microbe while keeping other factors constant. As expected, mice colonized with *C. sporogenes* mutants had reduced amounts of the corresponding metabolites. In addition, mice harboring mutant *C. sporogenes* defective in the metabolism of the branched-chain amino acids (valine, leucine, and isoleucine) to branched short-chain fatty acids had an increase in immunoglobulin A (IgA)–expressing cells as well as an increase in immune cells that express the IgA receptor. IgA is the predominant form of antibodies secreted into the gut, and although the exact role that IgA plays in gut homeostasis is not clear, dysregulation of IgA production and recognition is likely to affect how the host immune system responds to both commensal (symbiotic) bacteria and pathogens.

**"...human microbiota are an underexplored resource for drug-like molecules."**

The studies of Sugimoto *et al.* and Guo *et al.* represent important steps forward, but it is important to note some caveats. Although the study of Sugimoto *et al.* shows that the gut and oral microbiota produce molecules similar to clinically used drugs, they focused on type II polyketides, which are a small subset of the drug-like molecules that bacteria are already known to make. Furthermore, because they searched for homology to known, conserved domains, they necessarily discovered molecules very similar to those that have already been discovered. It will be interesting to see if further application of their analytical pipeline of metagenomic sequence data to direct pathway assembly outperforms assembling metagenomes and then searching for clusters using tools, such as the antibiotics and secondary metabolite analysis shell (antiSMASH), which analyzes microbial genomes for small-molecule drug-like biosynthetic potential (*9*). There are a few paths for future research. One would be based on the recognition that molecules with similar structures may not have similar functions, and screening libraries of molecular analogs could turn up some surprises. Another is to alter the algorithm so that pathways containing the target gene that are similar to known pathways are rejected, not selected.

From a tool perspective, Guo *et al.* provide a CRISPR-Cas9 system that is applicable to *C. sporogenes*, and probably other Clostridia species, which have different metabolic capacities than *C. sporogenes*. Their observation that branched short-chain fatty acids likely have a suppressive role in IgA production in mice monocolonized with *C. sporogenes* is interesting, but further research to delve into the mechanism by which these molecules exhibit these effects is needed. In particular, studies to determine whether these same molecules function identically in mice colonized with a more conventional microbiota ecosystem and whether they are relevant to human as well as mouse immune cells are warranted.

Human microbiology research faces the challenge of deconvoluting the enormous complexity of the microbiota and its associated molecules into mechanistic insights relevant to human health and disease. These two studies describe strategies for efficiently parsing the complexity. Perhaps of greater importance, they highlight that meaningful contributions to microbiota research will increasingly be dependent on collaborations between geneticists, microbiologists, chemists, immunologists, and physicians. ■

## REFERENCES AND NOTES


1. L. J. Cohen *et al.*, *Nature* **549**, 48 (2017).
2. M. R. Wilson *et al.*, *Science* **363**, eaar7785 (2019).
3. M. Xue *et al.*, *Science* **365**, eaax2685 (2019).
4. Y. Sugimoto *et al.*, *Science* **366**, eaax9176 (2019).
5. C.-J. Guo *et al.*, *Science* **366**, eaav1282 (2019).
6. M. Poyet *et al.*, *Nat. Med.* **25**, 1442 (2019).
7. L. García-Bayona, L. E. Comstock, *mBio* **10**, e01762-19 (2019).
8. M. S. Donia *et al.*, *Cell* **158**, 1402 (2014).
9. K. Blin *et al.*, *Nucleic Acids Res.* **47**, W81 (2019).


## ACKNOWLEDGMENTS

The authors acknowledge support from the NIH: GM R01AT009708 to J.C. and F32GM126650 to M.T.H.

10.1126/science.aaz4164



Published by AAAS

NEUROSCIENCE

# Cell types exposed by social scent

## A function-guided expression profiling strategy uncovers how pheromones are detected

*By* **Sabine L. Renninger**

Social interactions are an integral part of our lives, but sometimes they are as difficult to read as the activity in our brains. Animals use pheromones as a means of social communication to convey information about gender, age, individuality, and physiological state to other members of their species (conspecifics) (*1*). How such multifaceted information is detected and transformed into meaningful signals is an open question. On page 1384 of this issue, Lee *et al.* (*2*) present an optical approach called physiological optical tagging sequencing (PhOTseq) to disentangle sensory neuron cell types, which are specialized detectors and gateways to the brain. The authors combined functional and molecular profiling of cellular properties to expose a molecular logic of pheromone sensation.

Mice detect pheromone molecules with specialized cells in the vomeronasal organ, which is located at the base of the nasal cavity above the mouth. Vomeronasal sensory neurons recognize only one or a few molecules through specific receptor-ligand interactions (see the figure). Expression of vomeronasal receptor (VR) proteins, V1Rs and V2Rs, can vary between individuals and substantially between species (*3*). More than 300 potential functional VR genes have been identified in the mouse genome (*4*, *5*). Regulatory mechanisms largely limit the expression of VRs to one gene per cell (*6*).

Lee *et al.* hypothesized that a comprehensive response profile of vomeronasal sensory neurons to a defined set of pheromone molecules can uncover distinct chemoreceptive fields and allow them to reliably identify functional cell classes and molecular identities. Using state-of-the-art imaging technologies, the authors detected 20 different neuronal activity patterns in response to a set of steroid pheromones. They then optically tagged five functional types for isolation and single-cell gene expression profiling (see the figure). In the densely packed epithelium of the mouse vomeronasal organ, they identified and enriched even rare cell types that constitute less than 0.3% of the entire population.

Focusing on the VR gene family, Lee *et al.* correlated the chemoreceptive fields of neuronal types with molecular properties. For each functional class, they identified the most abundantly expressed VR, thereby confining putative V1R genes that mediate responses to specific steroid hormones. Moreover, comprehensive single-cell gene expression profiling exposed a systematic deviation from the one receptor–one neuron rule. The authors found coexpression of several VR gene pairs located in close chromosomal proximity. Whether this coexpression is pivotal for sensory tuning or a result of ongoing subfunctionalization after gene duplication remains to be explored.

What can mapping of a closed sensory space tell scientists about the functional diversity of the VR family? And can the protein sequence of a receptor provide information about its specificity? Lee *et al.* found that evolutionary distances between protein sequences did not accurately capture their functional relationship. However, when the authors mathematically transformed all sequence similarities into a two-dimensional representation, VRs clustered better with respect to their ligand specificity. With this analysis, the authors demonstrated that the primary receptor sequence may have predictive power for functional specificity.

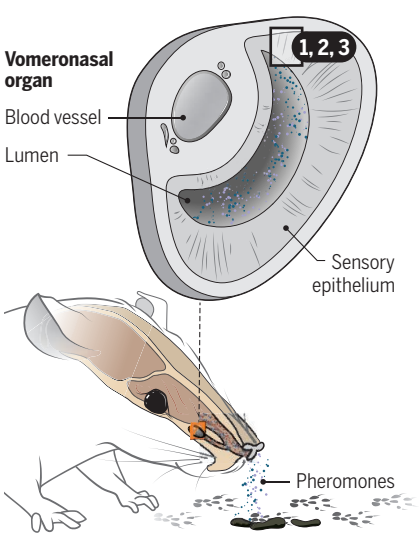
Being able to predict the functions of VRs is key for understanding how pheromones are perceived and how they serve social communication to guide interactions between individuals. Lee *et al.* described multiple sensory neuron cell types and receptors with overlapping specificity for estrogen-like pheromones, but did not investigate their impact on social interactions. It remains unknown how the individual VRs or how sensory integration across cell types signal behavioral or physiological changes. Specific interactions between pheromones and members of the V2R family control aggressive and sexual behavior in a gender-specific manner (*1*). V1Rs have so far been shown to function in a cooperative manner to influence behavior (*7*). Some sulfated estrogens work in the context of other gender cues to induce mounting behavior in male mice (*8*). Effects on more subtle changes in the physiology of an animal are barely explored.

Pheromones are naturally encountered as a blend of molecules. Integration of such multifaceted information directs innate responses and ensures fitness of an animal.

Champalimaud Research, Champalimaud Centre for the Unknown, 1400–038 Lisbon, Portugal.
Email: sabine.renninger@neuro.fchampalimaud.org

### Detecting social cues

Vomeronasal sensory neurons detect chemical cues, such as pheromones, through specific ligand-vomeronasal receptor interactions.

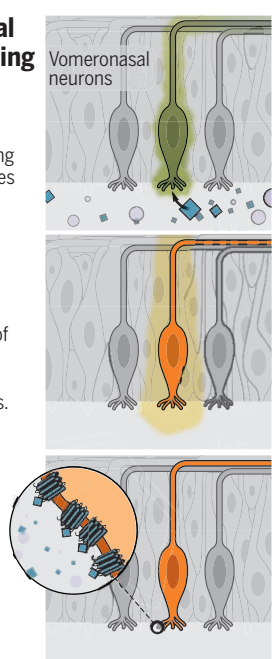



### Physiological optical tagging sequencing (PhOTseq)

**1** Calcium imaging uncovers cell types with specific chemoreceptive fields.

**2** Targeted photoactivation of a fluorescent marker labels specific cell types.

**3** Single-cell RNA sequencing of tagged cells reveals ligand-receptor specificity.




GRAPHIC: V. ALTOUNIAN/*SCIENCE*



Published by AAAS

Pheromone sensitivity can be modulated in a state-dependent or experience-dependent manner (*9*, *10*). For example, female mice are temporarily blind to male pheromones according to the phase of their estrus cycle. This sensory silencing, exclusive to conspecific peptide pheromones, is mediated by the hormone progesterone. Sensitivity to equivalent peptides from other species is preserved to secure detection of putative danger. Lee *et al.* investigated chemosensation in the context of receptor-ligand specificity. However, PhOTseq provides rich information on the gene expression profiles of individual cells and opens the door to systematically exploring the molecular details that define and distinguish neuronal function.

Combined in-depth analysis of functional and molecular properties will help refine the boundaries of specific cell types. Identification of functional classes of neurons is often ambiguous and becomes increasingly difficult with the computational power of a brain area. Primary sensory areas already display complex responses, rendering incoming sensory information in the context of ongoing behavior, current goals, and expectations (*11*). Molecular markers are crucial to disentangle such complex neuronal response dynamics. The method established by Lee *at al.* links distinct activity patterns with molecular fingerprints and allows the identification of specific neuronal markers. Applying this optical approach to awake and behaving animals will be a big step forward in the design of more advanced transgenic tools to specifically target neural circuits and uncover their computational principles. ■

**"...functional and molecular profiling...expose a molecular logic of pheromone sensation."**


### REFERENCES AND NOTES

1. S. D. Liberles, *Annu. Rev. Physiol.* **76**, 151 (2014).
2. D. Lee, M. Kume, T. E. Holy, *Science* **366**, 1384 (2019).
3. L. Silva, A. Antunes, *Annu. Rev. Anim. Biosci.* **5**, 353 (2017).
4. J. M. Young, H. F. Massa, L. Hsu, B. J. Trask, *Genome Res.* **20**, 10 (2010).
5. J. M. Young, B. J. Trask, *Trends Genet.* **23**, 212 (2007).
6. L. Capello, D. Roppolo, V. P. Jungo, P. Feinstein, I. Rodriguez, *Eur. J. Neurosci.* **29**, 671 (2009).
7. K. Del Punta *et al.*, *Nature* **419**, 70 (2002).
8. S. Haga-Yamanaka *et al.*, *eLife* **3**, e03025 (2014).
9. S. Dey *et al.*, *Cell* **161**, 1334 (2015).
10. P. S. Xu, D. Lee, T. E. Holy, *Neuron* **91**, 878 (2016).
11. A. G. Khan, S. B. Hofer, *Curr. Opin. Neurobiol.* **52**, 131 (2018).



### ACKNOWLEDGMENTS

Thanks to G. Costa for designing the figure.

10.1126/science.aaz8969


HIGH PRESSURE

# Extreme diamond-based quantum sensors

## Nitrogen vacancies make for superlative sensors of material properties at high pressures


*By* **James J. Hamlin**[1] *and* **Brian B. Zhou**[2]


We spend our entire lives at pressures near 1 atm. But most of the matter in our planet exists at far higher pressures. Experiments conducted under applied pressure are crucial to understanding condensed matter. High-pressure experiments have provided data on the matter in planetary interiors that have improved our understanding of seismic events. Most notably, applied high pressures permit the synthesis and study of new materials with extraordinary properties, such as extreme hardness. Recent experiments on hydride materials compressed to greater than 1 million atm have revealed near-room-temperature superconductivity (*1*–*3*), finally pushing past record critical temperatures that had stagnated since the 1990s. On pp. 1349, 1359, and 1355 of this issue, Hsieh *et al.* (*4*), Lesik *et al.* (*5*), and Yip *et al.* (*6*), respectively, report on a comprehensive set of experiments that demonstrate that quantum sensors based on so-called nitrogen vacancy (NV) centers offer powerful new tools for probing matter at extreme pressures.

The current record holder for the highest sustained static pressure (>400 GPa or 4 million atm) is a device called the diamond anvil cell (DAC) (see the figure). The DAC achieves these extraordinary pressures by squeezing the sample between the flattened tips of two diamonds. The small diameter of the diamond flats means that large pressures can be reached with only small applied forces but also means that the sample is typically very small (on the order of tens or hundreds of micrometers across). One key challenge associated with performing measurements under high applied pressures is that the sample must be contained by a bulky pressure vessel that severely constrains the types of measurements that can be performed. For example, sensitive magnetization measurements in a DAC are difficult because the magnetic background of the DAC is typically thousands of times larger than that of the comparatively tiny sample. However, a substantial benefit of using diamond is that it is transparent over a wide range of wavelengths. An ideal sensor for measurements in a DAC would take advantage of this property and feature an optical readout. In addition, other desirable qualities include high sensitivity, small size for spatial resolution, and operability in a broad range of temperature and pressure.

Fortuitously, just such a probe may have existed inside the diamond anvil cell all along. The spin associated with an atomic-scale defect in diamond known as the NV center has emerged as a superlative sensor. When the NV center is irradiated by a microwave field, narrow transitions between the NV spin states can be resolved optically because of the NV center's spin-dependent fluorescence intensity (see the figure). This permits spectroscopy of the spin energies, which depend sensitively on the external magnetic field or other stimuli such as temperature and stress. Placing NV centers in close proximity to a sample enables high-sensitivity, nanoscale-resolved mapping of the magnetization or electric current.

In 2014, pioneering work by Doherty *et al.* (*7*) demonstrated the feasibility of performing optically detected magnetic resonance on ensembles of NV centers in diamond samples placed inside DACs. Steele *et al.* (*8*) adopted a similar approach but used a "designer" diamond anvil with an integrated microwave waveguide. These initial works did not interface NV centers with a sample material under high pressure but envisioned the imaging of the magnetic profile on the surface of a pressurized superconductor. The three groups of authors (*4*–*6*) now realized this concept, using NV centers in anvil cells to spatially resolve bulk magnetic phenomena in condensed-matter systems across a temperature-pressure phase space reaching ~50 GPa and cryogenic temperatures.

All three works used ensembles of NV


[1]Department of Physics, University of Florida, Gainesville, FL 32611, USA. [2]Department of Physics, Boston College, Chestnut Hill, MA 02467, USA. Email: jhamlin@ufl.edu, brian.zhou@bc.edu




Published by AAAS

centers, simultaneously detecting the transition frequencies of four inequivalent orientations of NV centers in the lattice to deduce the full components of the local magnetic field vector or strain tensor. Although the broad concept is the same, the three works are distinct in their implementation of NV magnetometry inside anvil cells along with the physical phenomena probed. This variety demonstrates the versatility of the technique. Yip *et al.* (*6*) coated diamond particles onto the iron-pnictide superconductor $BaFe_2(As_{0.59}P_{0.41})_2$, placing them together inside an anvil cell. By focusing on particles at specific locations, the authors sampled the magnetic field profile adjacent to this type II superconductor and detected the onset of both the vortex and Meissner states.

Alternatively, both Hsieh *et al.* (*4*) and Lesik *et al.* (*5*) created a dense layer of NV centers directly on the surface of the diamond anvil culet using ion implantation. This enables continuous, diffraction-limited spatial resolution by either scanning a focused laser beam in a confocal microscope (*4*) or by use of widefield illumination and imaging with a camera (*5*). The authors demonstrate these capabilities by imaging the vector magnetic field profile of an iron pellet undergoing the $\alpha$-$\varepsilon$ magnetic phase transition and an $MgB_2$ sample across its superconducting transition. In addition, by leveraging the sensitivity of NV spin energies to lattice compression, Hsieh *et al.* (*4*) directly imaged the different components of the stress field inside the DAC.

Several important revolutions in high-pressure science have hinged on the development of new ways to use light to extract information from a compressed sample. After the invention of the DAC in 1959 (*9*), many of the early experiments involved simple visual observation of the sample under a microscope as pressure was gradually increased. It was immediately clear that the device offered access to a vast array of new phenomena at previously inaccessible pressures. Despite this, for more than a decade the use of the DAC was reported in only a small number of publications. A part of the problem was that there was no quick, simple method with which to measure the magnitude of the pressure inside the cell. This changed in 1972, when Forman *et al.* developed an optical pressure gauge based on the fluorescence spectrum of a small chip of ruby located inside the pressure cell (*10*). Use of the DAC expanded rapidly after the development of this convenient means of determining pressure. Ruby fluorescence remains the most common method of pressure determination inside a DAC. Light has continued to play a key role in the evolving use of the DAC with the development of several third-generation synchrotron facilities. In the past few decades, such facilities, which feature high x-ray flux and tight spatial resolution, have permitted x-ray structure determination at previously unthinkable pressures. Moreover, synchrotron-based inelastic x-ray scattering techniques can probe other properties at high pressure, including phonon spectra, local electronic structure, and element-specific magnetic environments.

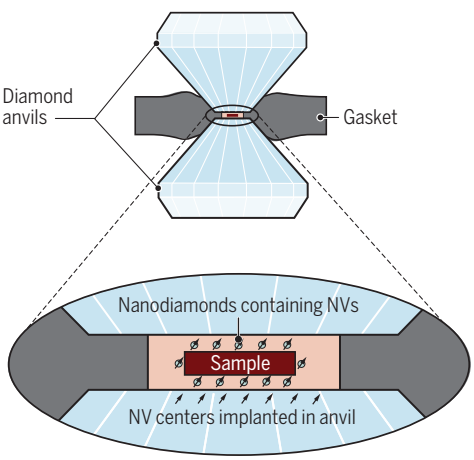
Adding defect-based quantum sensors to the toolkit for probing materials at high pressure opens up several opportunities. NV measurements can be built around a simple tabletop fluorescence microscope. Beyond static magnetic and stress fields, NV sensors can be used to investigate magnetization dynamics, spin noise spectroscopy, and nuclear magnetic resonance of materials inside DACs. Moreover, the sensitivity to temperature (*11*) means that NV centers could form the basis for high-pressure calorimetric measurements. The magnetic-field resolution can be enhanced by improving the properties of near-surface NVs and using more sophisticated sensing sequences. The latter relies on coherent manipulation of NV superposition states inside DACs (*12*). To investigate phenomena at even higher pressures, overcoming the reduction in the spin readout contrast of the diamond NV center is a substantial challenge because contrast is greatly diminished by 60 GPa (*7*). However, NVs are not the only optically addressable defect that can be explored, and continued progress may take quantum sensing to the ultimate pressure limits of the diamond anvil cell. ■

## A sensor embedded in diamond

Nitrogen atoms can replace carbon and combine with an adjacent vacancy in the diamond lattice. The fluorescence signal from the nitrogen vacancy (NV) centers can be used to determine the properties of a nearby sample.

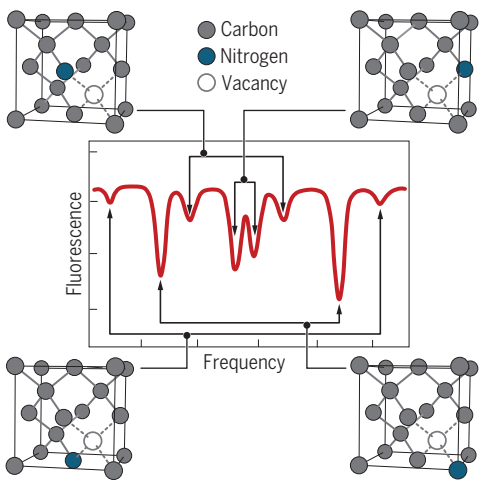
**High-pressure applications**

Opposing diamond anvils apply extreme pressure on a sample. NVs come into sample proximity either by embedding them into the diamond anvils or by depositing diamond particles with NVs onto the sample.




**Sensing with NV centers**

The fluorescence signal from four NV orientations display magnetic resonances that are sensitive to the local environment. Measuring through a transparent anvil allows determination of sample properties under applied pressure and external magnetic field.




GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

### REFERENCES AND NOTES


1. A. P. Drozdov, M. I. Eremets, I. A. Troyan, V. Ksenofontov, S. I. Shylin, *Nature* **525**, 73 (2015).
2. M. Somayazulu *et al.*, *Phys. Rev. Lett.* **122**, 027001 (2019).
3. A. P. Drozdov *et al.*, *Nature* **569**, 528 (2019).
4. S. Hseih *et al.*, *Science* **366**, 1349 (2019).
5. M. Lesik *et al.*, *Science* **366**, 1359 (2019).
6. K.Y. Yip *et al.*, *Science* **366**, 1355 (2019)
7. M. W. Doherty *et al.*, *Phys. Rev. Lett.* **112**, 047601 (2014).
8. L. G. Steele *et al.*, *Appl. Phys. Lett.* **111**, 221903 (2017).
9. C. E. Weir, E. R. Lippincott, A. Van Valkenburg, E. N. Bunting, *J. Res. Natl. Bur. Stand. Sect. Phys. Chem.* **63A**, 55 (1959).
10. R. A. Forman, G. J. Piermarini, J. D. Barnett, S. Block, *Science* **176**, 284 (1972).
11. G. Kucsko *et al.*, *Nature* **500**, 54 (2013).
12. Y.-X. Shang *et al.*, *Chin. Phys. Lett.* **36**, 086201 (2019).


10.1126/science.aaz4982



*Published by AAAS*

POLYMERS

# Making stronger carbon-fiber precursors

## Cross-linking multiple strands improves strength and toughness of polyacrylonitrile fibers

*By* **Bronwyn Fox**

Carbon fiber–reinforced polymers are used to create lightweight, high-strength structures such as airframes. More than 96% of commercial carbon fiber is derived from high-temperature heating (up to 1700°C) and stretching of a precursor fiber made from polyacrylonitrile (PAN) (*1*). In this process, PAN cyclizes to create a ladder structure, which is then carbonized to create the final turbostratic (disordered) structure. The high strength and stiffness of the carbon fibers are highly dependent on the composition of the precursor fiber and are limited by any defects such as voids or foreign particles. Generally, there is a trade-off between tensile strength and toughness, but on page 1376 of this issue, Liao *et al.* (*2*) report increasing both properties of PAN-based fibers. Their approach draws on an ancient understanding that multiple filaments wound together are stronger than individual filaments—namely, that "a cord of three strands is not easily broken" (Eccles. 4:12).

Both chemical and physical approaches have been undertaken to try to improve the properties of PAN-based precursor fibers. The industrial polymerization of acrylonitrile and efficient conversion to PAN precursors is a convoluted and mostly batch process that accounts for about half the cost of the resulting carbon fiber, so alternative systems for PAN manufacture have potential cost savings. The PAN homopolymer has limited solubility and a poor ability to be spun into fibers, and when heat treated, it undergoes an uncontrollable exothermic chemical reaction that makes it unsuitable for conversion to carbon fiber (*3*).

For these reasons, small amounts (2 to 5%) of one or more comonomers are added during PAN polymerization to facilitate processability. Although it is common to add two comonomers, such as itaconic acid and methyl acrylate, it is synthetically complicated to distribute two different comonomers evenly along the backbone of the polymer chain. An alternative approach developed by Moscowitz, Wiggins, and co-workers (see the figure, top) uses a *N*-isopropylacrylamide comonomer that can be distributed evenly. This approach also facilitates the cyclization reaction to create well-controlled ladder structures (*4*, *5*).

An alternative approach to increasing the mechanical properties of PAN-based precursor fibers has been to disperse additives such as carbon nanoparticles. Chae *et al.* (*6*) reported an increase in fiber modulus to 20 GPa through the addition of carbon nanotubes to gel-spun PAN. Sayyar *et al.* (*7*) reported an increase of 28 and 20% in tensile strength and Young's modulus of carbonized fibers, respectively, with the dual strategy of adding graphene (0.5 weight %) and creating a multifibrillar yarn.

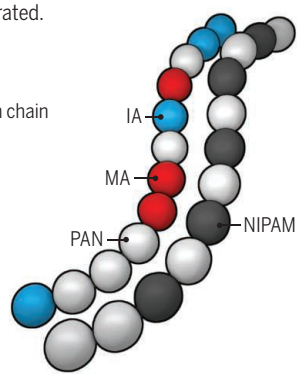
Much progress toward understanding the effect of molecular weight (the length of the polymer chain) and the molecular weight distribution or polydispersity (consistency of the chain lengths) on the properties of carbon fibers has recently been made. Cai *et al.* (*8*) developed a new type of precursor using reversible addition–fragmentation chain-transfer (RAFT) polymerization, which allows control of the PAN polymer chain length (see the figure, top). This precursor has a controlled molecular weight and maintains a low polydispersity, which improved the mechanical properties of both precursor and carbon fiber compared with control polymers of similar molecular weights. These RAFT PAN polymers also had much lower viscosity profiles, which facilitated new and different processing routes (*9*).

The effect of physical defects such as voids, chain entanglements, and foreign particles in precursor fibers was highlighted by Chae and Kumar (*10*) more than a decade ago. Since then, substantial efforts have focused on understanding the coagulation of fibers in the wet-spinning process. In this first step of the manufacturing process, the polymer solution is extruded into a water bath. This stage effectively locks in the physical structure of the fiber. Newcomb *et al.* (*11*) highlighted the importance of the temperature of the

Manufacturing Futures Research Institute, Swinburne University of Technology, Hawthorn, VIC 3122, Australia. Email: blfox@swin.edu.au

### Improving PAN fibers

Polyacrylonitrile (PAN) fibers are used to create carbon fibers for composite materials. Some of the strategies used to improve their properties are illustrated.

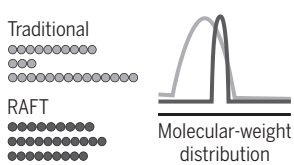



#### Chemical strategies

Added comonomers and more uniform chain lengths can improve PAN processing.

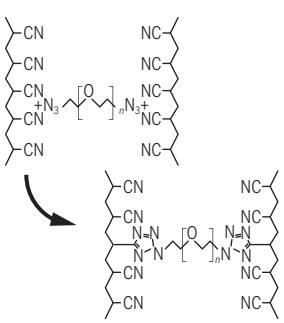
**Monomer control**
Added monomers such as itaconic acid (IA) and methyl acrylate (MA) can distribute unevenly, but *N*-isopropylacrylamide (NIPAM) adds more uniformly and provides cross-linking (*4*, *5*).

**Chain-length control**
Reversible addition–fragmentation chain-transfer (RAFT) polymerization can create more uniform chain lengths than traditional routes (*8*).

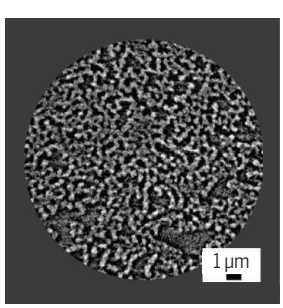



#### Combining cross-linking and crystallization

Liao *et al.* used cross-linking and processing to create stronger, tougher PAN fibers.




**Copolymer cross-links**
Poly(ethylene glycol) bisazide (PEG-BA) cross-links with the nitrile groups (CN) on PAN of PAN-*co*-MA polymers to create ladder polymers.




**Processing yarns**
Yarn electrospinning, followed by stretching and annealing of the ladder polymers, creates highly crystalline, uniform fibers.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*



Published by AAAS

bath in creating nonideal internal porosity and kidney bean–shaped fibers. Other researchers have contributed to understanding the role of added solvents in this bath in controlling fiber density, diameter, and morphology (*12*).

Liao *et al.* combined both a chemical and physical approach and report the mechanical properties of multifibrillar polyacrylonitrile-*co*-methyl acrylate fibers, which were modified with a small amount of poly(ethylene glycol) bisazide (PEG-BA) (see the figure, bottom). The PEG-BA fraction was varied from 0 to 6 weight % relative to PAN. The properties achieved represent an important breakthrough in overcoming the trade-off between specific strength (tensile strength divided by density) and toughness in fibers. For example, a recent study of electrospun PAN fibers (*13*) reported a specific strength of 0.153 GPa $g^{-1}$ $cm^{-3}$ and a toughness of 16 J $g^{-1}$, whereas the fibers of Liao *et al.* had a specific strength of 0.76 to 1.27 GPa $g^{-1}$ $cm^{-3}$ and a toughness of 118 to 166 J $g^{-1}$. The authors attributed this improvement to the click-chemistry cross-linking reactions between fibrils during spinning as well as the annealing under tension that increased fiber crystallinity from 56.9 to 92.4%, as measured by wide-angle x-ray scattering.

The PAN fibers produced by Liao *et al.* have properties approaching those of dragline spider silk. The next challenge is to carbonize these electrospun multifibrillar yarns and measure the resulting carbon fiber properties. Combining a number of the aforementioned strategies could create a new generation of high-performance carbon fibers with unprecedented properties. ■


### REFERENCES AND NOTES

1. B. A. Newcomb, *Compos. Part A* **91**, 262 (2016).
2. X. Liao *et al.*, *Science* **366**, 1376 (2019).
3. H. Liu *et al.*, *Polymers (Basel)* **11**, 1150 (2019).
4. J. D. Moskowitz, J. S. Wiggins, *Polymer (Guildf.)* **84**, 311 (2016).
5. K. D. Cordell, R. J. Dias, D. C. Walker, J. S. Wiggins, in *CAMX Conference Proceedings: The Composites and Advanced Materials Expo*, V. Bailey *et al.*, Eds. (Society for the Advancement of Material and Process Engineering, 2018), pp. 1–9.
6. H. G. Chae *et al.*, *Carbon* **93**, 81 (2015).
7. S. Sayyar, J. Moskowitz, B. Fox, J. Wiggins, G. Wallace, *J. Appl. Polym. Sci.* **136**, 47932 (2019).
8. J. Y. Cai *et al.*, *Mater. Today Commun.* **9**, 22 (2016).
9. J. Kaur, K. Milllington, J. Y. Cai, *J. Appl. Polym. Sci.* **133**, 44273 (2016).
10. H. G. Chae, S. Kumar, *Science* **319**, 908 (2008).
11. B. A. Newcomb *et al.*, *Polym. Eng. Sci.* **55**, 2603 (2015).
12. E. A. Morris, M. C. Weisenberger, G. W. Rice, *Fibers (Basel)* **3**, 560 (2015).
13. G. Duan, S. Liu. H. Hou, *e-Polymers* **18**, 569 (2018).



### ACKNOWLEDGMENTS

I thank S. Smith, J. Wiggins, and J. Mardel for their helpful insights.


GRAPHIC: A. KITTERMAN/*SCIENCE*

10.1126/science.aaz7928

CANCER

# Surprising regulation of cell cycle entry

## Structures of heterotrimeric cyclin-dependent kinase 4 complexes reveal the mechanism of drug inhibition

*By* **Charles J. Sherr**

The activities of cyclin-dependent kinases (CDKs), regulated primarily by the periodic expression of their cyclin binding partners, temporally order sequential cell cycle transitions through $G_1$, S phase, $G_2$, and mitosis. In mammalian cells, regulators of $G_1$ transit include three D-type cyclins, as well as CDK4 and CDK6, and the CDK inhibitory proteins, p21 and p27 (*1*). In response to mitogens, individual D-type cyclins assemble with CDK4 or CDK6 and, paradoxically, with the p21 or p27 "inhibitors" to yield active higher-order holoenzymes that drive $G_1$ progression and prime cells to enter S phase and begin DNA replication (*2*). On page 1330 of this issue, Guiley *et al.* (*3*) report the crystal structures of trimeric complexes containing cyclin D1, CDK4, and either p27 or p21 and reveal that active trimers containing tyrosine-phosphorylated p27 are surprisingly refractory to the U. S. Food and Drug Administration (FDA)–approved CDK4/6 inhibitors that are used to treat hormone-dependent breast cancer.

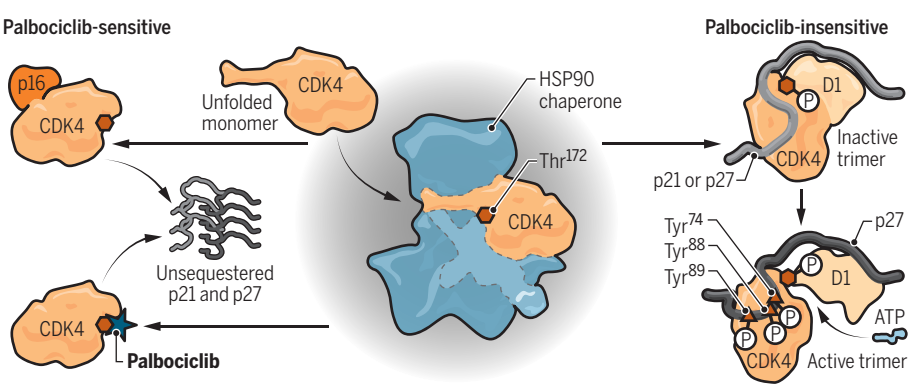
St. Jude Children's Research Hospital, Memphis, TN, USA.
Email: sherr@stjude.org

Cyclin D holoenzymes preferentially phosphorylate the tumor suppressor retinoblastoma (RB) protein, and a few other substrates, to cancel the antiproliferative activity of RB. By contrast, cyclin E–CDK2 (which acts at the $G_1$/S transition) and cyclin A– and B–driven CDK2 and CDK1 complexes sequentially phosphorylate many hundreds of substrates (including RB) during the remainder of the cycle. Through their simultaneous binding to both cyclin and CDK subunits, p21 and p27 can inhibit CDKs 1, 2, 4, and 6 to enforce cell cycle arrest in response to mitogen withdrawal, antiproliferative cytokines, cell contact inhibition, or cellular stress. Unexpectedly, all cyclin D1–CDK4-dependent kinase activity toward RB requires higher-order complexes containing p27 or p21, which are required for holoenzyme assembly (*2*). Phosphorylation of bound p27, but not p21, by nonreceptor tyrosine kinases converts the inactive cyclin D1–CDK4-p27 complex to an active RB kinase (*4*) (see the figure). But it remains unknown how CDK4 holoenzyme assembly and activity are mediated by p27 and whether p21 is subject to similar regulatory controls.

Guiley *et al.* find that both p21 and p27 bind cyclin D1 and CDK4 and deform the

### Regulation of cyclin-dependent kinase 4 complexes

Newly synthesized CDK4, folded by the HSP90-chaperone system, can be sequestered by palbociclib or by p16, which increases unsequestered p27 and p21 and inhibits CDK2 complexes to block the $G_1$/S transition. Extraction of CDK4 by p27 or p21 and cyclin D1 yields an inactive trimer capable of undergoing CAK-mediated $Thr^{172}$ phosphorylation on its reconfigured activation segment. Activation of the trimer requires phosphorylation of tyrosine residues on p27.




ATP, adenosine triphosphate; CAK, CDK-activating kinase; CDK4, cyclin-dependent kinase 4; HSP90, heat shock protein 90; P, phosphorylation.



Published by AAAS

adenosine triphosphate (ATP) binding pocket of CDK4, which is required for substrate phosphorylation. Yet p27 induces additional structural changes that independently shape the pocket to "prime" it for catalysis. Concordant release of an activation segment from the CDK4 ATP binding pocket exposes $Thr^{172}$ that is phosphorylated by CDK-activating kinase (CAK, which comprises cyclin H and CDK7) to stimulate subsequent enzyme activity (*5*). Although none of these structural changes are enough for CDK4 activation, phosphorylation of p27 on $Tyr^{74}$ weakens its association with CDK4 to elicit recombinant trimer activity; p21 contains a phenylalanine at the analogous position and remains tightly bound and inhibitory. However, there is conflicting evidence that p21, like p27, might function as a CDK4 activator under some circumstances (*2*), perhaps through other mechanisms. Phosphorylation of p27 $Tyr^{88}$ and $Tyr^{89}$, not visualized in the crystal structures of Guiley *et al.*, likely makes an additional contribution to holoenzyme activity (*3*, *4*).

A distinct group of inhibitors—called the INK4 proteins—specifically target CDK4 and CDK6 to arrest proliferation in $G_1$ in an RB-dependent manner (*1*, *2*). In particular, a stress-induced INK4 protein, p16, like RB, is a tumor suppressor, whereas CDK4 and CDK6 have proto-oncogenic activity. The finding that deregulation of the p16-CDK4/6-RB pathway aberrantly drives cancer cell proliferation spurred pharmaceutical development of specific CDK4/6 inhibitory drugs (palbociclib, ribociclib, abemaciclib), which, when used together with inhibitors of estrogen receptor (ER) signaling, were FDA-approved for the treatment of ER-positive breast cancer (*6*). Combinatorial therapy with CDK4/6 and ER inhibitors significantly increases progression-free survival in women with advanced, metastatic breast cancer (*7*–*9*), and clinical trials of CDK4/6 inhibitors are under way for other malignancies (*6*, *10*).

Counterintuitively, Guiley *et al.* find that palbociclib does not inhibit the active phosphorylated p27-CDK4–cyclin D1 trimer and instead targets monomeric CDK4. Newly synthesized CDK4 requires the heat shock protein 90 (HSP90)–containing chaperone system for proper folding, and palbociclib disrupts this interaction to increase the relative abundance of the CDK4 monomer at the expense of trimer assembly. Therefore, palbociclib mimics p16, which also sequesters monomeric CDK4 to prevent its assembly with cyclin D1 and p27 or p21. The resulting increase in untethered p27 and p21 sets an elevated threshold to be overcome for activation of cyclin E–CDK2 and cyclin A–CDK2 at the $G_1$/S transition (*2*). Thus, inhibition of both CDK4 and, indirectly, CDK2 by palbociclib or p16 orchestrates $G_1$ arrest of cancer cells.

These findings advance the mechanistic understanding of p27-CDK4–cyclin D1 activation after more than two decades since the component molecules were discovered. The development of palbociclib and related drugs relied on inhibition of recombinant cyclin D and CDK4 produced using baculovirus vectors in insect cells, which express the HSP90-chaperone and CAK. This reconstituted dimeric enzyme retains activity, although vanishingly small quantities of such dimers are detected in mammalian cells. In continuously cycling cells, the rate of progression through $G_1$ into S phase is determined by events in the previous cell cycle. Notably, cyclin D1 is preferentially degraded in S phase and expressed again in $G_2$, raising the possibility that palbociclib "traps" monomeric CDK4 late in the cell cycle, preventing its assembly with cyclin D1 and leading to RB-dependent $G_1$ arrest. It is unknown what factors determine the equilibrium between palbociclib-sensitive monomeric CDK4 and the drug-resistant p27-CDK4–cyclin D1 trimer, or whether p27 and p21 compete in forming active and inactive trimers. Moreover, proliferating cells bifurcate into two subpopulations after the anaphase stage of mitosis based on the amounts of replicative DNA damage and induction of p21 by the tumor suppressor p53 in the previous cycle (*11*, *12*). Cells inheriting high p21 (and low CDK2) expression exhibit a greater dependency on renewed RB phosphorylation to progress through $G_1$, implying that nonuniform responses to CDK4 inhibitors depend on variable inheritance of key regulators (e.g., p21, p27, cyclin D1).

Unlike CDK4, CDK6 is a weak client of the HSP90-chaperone and is generally dependent on cyclins D2 or D3 for activity (*13*). Robust CDK6 activity can confer palbociclib resistance in CDK4-driven cancers (*14*, *15*), possibly because CDK6 assembles more efficiently than CDK4 into palbociclib-resistant trimers. Structures of CDK4 and CDK6 in complexes with the D-type cyclins are needed to understand differences in regulatory mechanisms and to improve drug development. ■

## REFERENCES AND NOTES


1. M. Malumbres, *Genome Biol.* **15**, 122 (2014).
2. C. J. Sherr, J. M. Roberts, *Genes Dev.* **13**, 1501 (1999).
3. K. Z. Guiley *et al.*, *Science* **366**, eaaw2106 (2019).
4. P. Patel *et al.*, *Mol. Cell. Biol.* **35**, 1506 (2015).
5. M. M. Schachter *et al.*, *Mol. Cell* **50**, 250 (2013).
6. U. Asghar *et al.*, *Nat. Drug Discov.* **14**, 130 (2015).
7. R. S. Finn *et al.*, *N. Engl. J. Med.* **375**, 1925 (2016).
8. A. Marra, G. Curigliano, *NPJ Breast Cancer* **5**, 27 (2019).
9. T. K. Eggersmann *et al.*, *BioDrugs* **33**, 125 (2019).
10. C. J. Sherr *et al.*, *Cancer Discov.* **6**, 353 (2016).
11. S. L. Spenser *et al.*, *Cell* **155**, 369 (2013).
12. J. Moser *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8219 (2018).
13. S. T. Hallett *et al.*, *Cell Rep.* **21**, 1386 (2017).
14. Z. Li *et al.*, *Cancer Cell* **34**, 893 (2018).
15. E. Haines *et al.*, *Oncotarget* **9**, 31572 (2018).


10.1126/science.aaz4043

MARINE BIOLOGY

# The biology of big

## Whales became the world's largest animals thanks to giant gulps of "bite-size" prey

*By* **Terrie M. Williams**

With so many recent scientific advances focused on the mini, micro, nano, and molecular scales, there is a tendency to overlook the titanic biology of giants that share the Earth. The sheer magnitude of a scientific undertaking to study an oceanic, 25-m-long, 95,000-kg wild blue whale (*Balaenoptera musculus*)—the largest vertebrate in the animal kingdom—has long left researchers with little more than brief glimpses of their presence when the leviathan surfaces to breathe. On page 1367 of this issue, Goldbogen *et al.* (*1*) describe how they took advantage of developments in microprocessor technology to design submersible wildlife tags. The authors used the new tools to measure the feeding performance and prey choices of the largest mammals in the seas. These data revealed the ecological and evolutionary factors that drive the biology of being, not just big, but the biggest ever—perhaps the biggest possible.

Not surprisingly, growing large depends on food. Physiologists and ecologists have long known that growth, reproduction, and survival of individuals and populations require a balance between energy expended in daily living and energy gained through the acquisition of food (*2*). Ultimately, "calories in" must equal or exceed "calories out." For carnivorous mammals, this balance involves decisions on whether to invest the energy required for nibbling on many small prey over time or killing and gorging on a single, large prey item. Body size figures prominently in this decision; a 300-g weasel would be hard-pressed to bring down a 300-kg buffalo, whereas a 200-kg African lion has a wider array of options. On land, the energetic tipping point for choosing between large or small prey revolves around a body mass of 21 to 25 kg, with smaller carnivores able to sat-

Department of Ecology and Evolutionary Biology, 130 McAllister Way, University of California–Santa Cruz, CA 95060, USA. Email: williams@biology.ucsc.edu



Published by AAAS

isfy energetic demands by feeding on many small prey and bigger carnivores foraging more efficiently on individual large vertebrate prey (*3*).

This size differential between predator and prey fits the common model of expectation until we examine marine-living predators. As described eloquently by Goldbogen *et al.* and others (*4*), the largest predators in the oceans paradoxically feed on the smallest prey. Here, the energetic breakpoint occurs at a body mass of ~10,000 kg, which separates smaller toothed whales (suborder: Odontoceti) that chase and ingest individual large prey from the enormous filter-feeding baleen whales (suborder: Mysticeti) that engulf swarms of tiny (<4 cm long) crustaceans.

When viewed together, the relationship between preferred prey size and mammalian body mass approaches a bell-shaped distribution spanning mammalian predators from weasels to whales (see the figure). At the extremes are the smallest and largest hunting mammals, which subsist on diminutive vertebrates and invertebrates. This leaves mid-sized to large marine and terrestrial mammals, weighing from 25 to 10,000 kg, to hunt big vertebrate prey that might exceed their own size. Feeding mechanisms employed by the species within each group support the effective capture and processing of these preferred prey. However, as shown by Goldbogen *et al.*, it is the hyperefficiency of the filter-feeding apparatus used by baleen whales that permitted mammalian gigantism to evolve in the oceans. Thus, baleen whales—named for keratinized baleen plates that filter plankton, krill, and small fish and replaced the teeth of ancient whale ancestors—ultimately became the world's largest vertebrates.

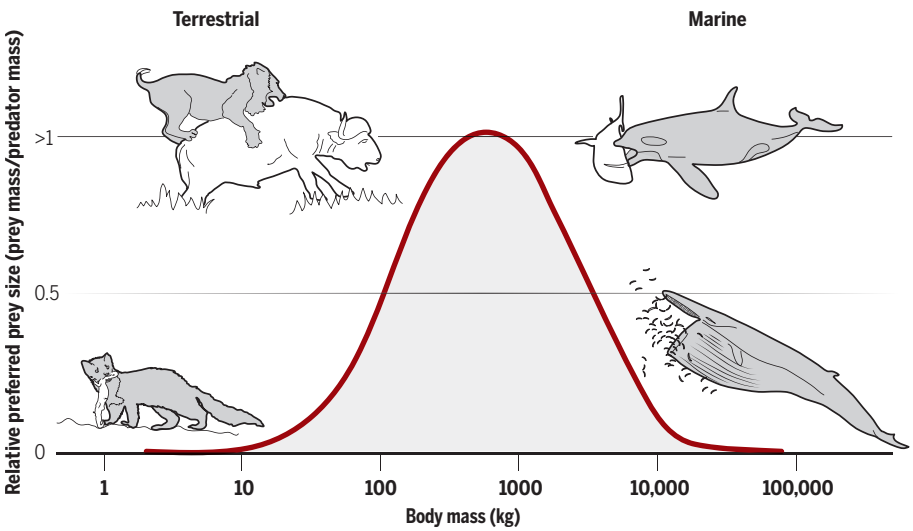
The anatomy of the largest lunge filter-feeders—including minke, humpback, fin, and blue whales examined by Goldbogen *et al.*—leaves little doubt that this group is designed to eat. Nearly one-third to one-half of the baleen whale's body comprises a head and pleated throat used for engulfing large volumes of seawater containing aggregations of small prey. Considered "the greatest biomechanical action in the animal kingdom" (*5*), the act of lunge feeding is mechanically costly but energetically rewarding as a result of the quantity and quality of densely aggregated prey ingested per endeavor (*4*, *6*, *7*). Processing all of this food also requires an extraordinary long intestinal tract, which in the fin whale reaches 100 m in length and is proportionately twice the length predicted for a theoretical land carnivore of the same size (*8*). Such morphological specialization contributes to an evolutionary release for growth in the great whales.

Why don't whales grow even bigger? As the new study suggests, it is in part because of limitations in the spatial and temporal availability of high-quality prey. Another likely factor is associated with physiological limits of the organs involved in prey acquisition and assimilation. For mammals, from humans (*9*) to sea otters, pinnipeds, dolphins, and whales (*10*), the alimentary tract sets the capacity for sustained metabolic performance. In general, there are only so many calories that can be physically processed from prey in 1 day, which ultimately sets a limit on the energy available for maintenance functions and growth.

By expanding these limits through behavior, morphology, and physiology, the great whales have maintained size dominance in the oceans for millions of years (*11*). As giant apex predators, their appetites, as well as their absence (caused by industrial whaling during the 20th century), have had fittingly enormous impacts on the ecosystems in which they live (*12*, *13*). In addition to removing astonishing amounts of prey in a single gulp, waste products from their food processing fertilize the oceans, thereby enhancing the primary production of prey that they and other marine animals rely upon (*14*). Even in death, their enormous bodies feed members of marine ecosystems at nearly every trophic level, from killer whales that hunt them to highly mobile scavengers and microbial assemblages that devour their carcasses (*15*). The critical, multifaceted role of whales within worldwide marine ecosystems warrants increased efforts to preserve these giants by deciphering and meeting their basic biological needs. Such efforts begin with appreciating the biology of big. ■

## Preferred prey mass depends on anatomy of mammalian carnivores

The red line demonstrates how the size of prey taken by predators changes across the mammalian size continuum. At the extremes, the smallest and largest mammalian carnivores consume large numbers of small prey. Mid-sized mammals hunt and consume larger individual prey items.




GRAPHIC: N. DESAI/*SCIENCE*

### REFERENCES AND NOTES


1. J. A. Goldbogen *et al.*, *Science* **366**, 1367 (2019).
2. D. W. Stephens, J. R. Krebs, *Foraging Theory* (Princeton Univ. Press, Princeton, NJ, 1986).
3. C. Carbone, G. M. Mace, S. C. Roberts, D. W. Macdonald, *Nature* **402**, 286 (1999).
4. A. Acevedo-Gutiérrez, D. A. Croll, B. R. Tershy, *J. Exp. Biol.* **205**, 1747 (2002).
5. P. Brodie, *Can. J. Zool.* **71**, 2546 (1993).
6. J. A. Goldbogen *et al.*, *Funct. Ecol.* **26**, 216 (2012).
7. D. E. Cade, A. S. Friedlaender, J. Calambokidis, J. A. Goldbogen, *Curr. Biol.* **26**, 2617 (2016).
8. C. E. Stevens, I. D. Hume, *Comparative Physiology of the Vertebrate Digestive System* (Cambridge Univ. Press, Cambridge, UK, 1995).
9. C. Thurber *et al.*, *Sci. Adv.* **5**, eaaw0341 (2019).
10. T. M. Williams, J. Haun, R. W. Davis, L. A. Fuiman, S. Kohin, *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **129**, 785 (2001).
11. G. J. Slater, J. A. Goldbogen, N. D. Pyenson, *Proc. Biol. Sci.* **284**, 20170546 (2017).
12. J. A. Estes *et al.*, *Science* **333**, 301 (2011).
13. A. M. Springer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12223 (2003).
14. J. Roman *et al.*, *Front. Ecol. Environ.* **12**, 377 (2014).
15. C. R. Smith, in *Whales, Whaling, and Ocean Ecosystems*, J. A. Estes, D. P. DeMaster, D. F. Doak, T. M. Williams, R. L. Brownell Jr., Eds. (Univ. of California Press, Berkeley, CA, 2006), p. 286.


### ACKNOWLEDGMENTS


T.M.W. is supported by grants from the Office of Naval Research (N00014-17-1-2737), National Science Foundation (NSF) Division of Polar Programs (1341401), and NSF Division of Biological Infrastructure (1255913-015).

10.1126/science.aba1128





*Published by AAAS*

POLICY FORUM

PUBLIC HEALTH

# Evidence, alarm, and the debate over e-cigarettes

## Prohibitionist measures threaten public health

*By* **Amy Fairchild[1], Cheryl Healton[2], James Curran[3], David Abrams[2], Ronald Bayer[4]**

This is a moment for legitimate alarm at the intersection of two distressing but distinct epidemiological patterns involving e-cigarettes ("vaping"): an increase in vaping among youth and a sudden outbreak of acute lung injuries and deaths in the United States, associated most strongly with vaping tetrahydrocannabinol (THC), the main psychoactive compound in cannabis. Discussions of vaping, however, often neglect distinctions between nicotine and THC; between adults and youth; and between products obtained through the retail and black markets. As we move to confront these challenges, we face the danger that justifiable alarm will turn alarmist, short-circuiting careful analysis of the full range of evidence and focusing attention on the most frightening, thus enhancing the prospect of adopting counterproductive policy. We suggest that the evidence warns against prohibitionist measures. Restricting access and appeal among less harmful vaping products out of an abundance of caution while leaving deadly combustible products on the market does not protect public health. It threatens to derail a trend that could hasten the demise of cigarettes, poised to take a billion lives this century.

For years, some leaders in the public health community observed with concern the rise in the number of U.S. adolescents who are vaping flavored liquids. Debate centered on whether vaping devices, which do not burn tobacco but rather heat liquids containing some combination of flavors and/or nicotine and/or THC, should be viewed as an intolerable threat to nonsmoking youth, or whether they can be managed with reasonable regulations that give adult smokers access to less hazardous alternatives to deadly combustible cigarettes (*1*, *2*).

This summer, the tone of the debate shifted markedly. The U.S. Centers for Disease Control and Prevention (CDC) reported sudden clusters of serious and sometimes fatal respiratory injuries. As of 4 December 2019, the CDC reported 2291 cases and 48 deaths. The CDC and the U.S. Food and Drug Administration (FDA) warned consumers not to vape THC or any liquids obtained off the streets or from unknown sources (*3*). In the ongoing investigation to determine the causes of illness and death, the CDC identified vitamin E acetate, a THC product additive, as "a chemical of concern." Nicotine or flavored vaping liquids have not yet been implicated.

Even as the investigation was under way, interest in prohibitionist measures swelled. Massachusetts banned retail and online sale of all nicotine vaping products until January 2020. San Francisco will ban all nicotine vaping products in early 2020. Michigan has banned all flavors (except tobacco flavor, meant to make a vaping liquid taste like a traditional combustible cigarette) for nicotine products, but not for THC. Ohio may ban menthol and mint flavors in e-cigarettes while still allowing them in more harmful cigarettes and little cigars. The American Medical Association called for a total ban on all vaping products. New York City is poised to become the largest jurisdiction to ban all flavored nicotine vapes including menthol, although it chose not to enact a menthol ban for combustible products. As a measure of the magnitude of the urge to weigh in, even the White House has offered policy pronouncements, including a potential ban on all nicotine vape flavors except tobacco, from which it recently backed off. Such sweeping measures have been adopted in the face of continuing debate over the public health impact of such interventions.

### THE CONTESTED HISTORY OF HARM REDUCTION

There has been a long debate in the United States about harm reduction as an evidence-based approach to reduce harms associated with life-threatening behaviors by providing safer, although not totally safe, alternatives. For example, despite strong, albeit always imperfect, evidence demonstrating the effectiveness of needle exchange programs in countering the spread of HIV infection, it took decades to overcome reluctance to providing sterile needles to people who inject drugs. Battles over needle exchange persisted long after the evidence was clear, and programs suffered under on-again, off-again federal funding.

For those addicted to combustible tobacco, harm reduction is a pragmatic approach (*4*). From the 1950s through the early 1980s, public health officials saw great promise in so-called "safer" combusted tobacco. Interest died with the emergence of damning evidence about mass deception on the part of the tobacco industry over light and low-tar cigarettes. In the 1990s, nico-

Although not safe, vaping nicotine represents a safer alternative to combustible cigarettes for adult smokers who cannot or will not quit smoking.

PHOTO: LADYWRITER55/ISTOCKPHOTO.COM

[1]College of Public Health, The Ohio State University, Columbus, OH, USA. [2]College of Global Public Health, New York University, New York, NY, USA. [3]Rollins School of Public Health, Emory University, Atlanta, GA, USA. [4]Mailman School of Public Health, Columbia University, New York, NY, USA. Email: fairchild.139@osu.edu



Published by AAAS

tine replacement therapies (NRTs) became widely available over the counter. NRTs were framed both as medicinal treatment and as a harm reduction approach, and the public health and medical communities were prepared to tolerate lifelong use of nicotine if necessary (*5*).

When vaped nicotine products first came on the scene, the harm reduction debate reignited. In the early years, the scientific evidence for this new technology was sparse (*4*). Some saw in them promise for adults who smoked, but most assumed a precautionary posture, arguing that we first required certainty about safety and efficacy.

Even as scientific studies emerged, many who were skeptical of vaping nicotine wanted stronger proof that such products were safe, effective, would not lead to greater net population harm than benefit, and would not renormalize smoking, in advance of allowing them to be sold. As time went on, although uncertainties remained, those open to harm reduction became more willing to act to provide alternatives. They believed the emerging science was sufficiently strong and the global toll of preventable smoking deaths remained so massive and urgent that benefits outweighed harms (*4*, *6*).

Vaping nicotine grew in popularity as products delivered nicotine in ways that were more appealing (e.g., flavored e-liquids) or more efficient (e.g., nicotine absorption began to mimic the effect of combustible products). In parallel, systematic reviews of the science accumulated, demonstrating that although such products were not safe, they were safer than combustible products. They became more popular and more effective than medicinal NRTs at helping smokers quit.

The U.S. FDA and the United Kingdom's (UK) Public Health England (PHE), which carefully tracked the evidence, showed openness to harm reduction, describing the importance of recognizing a continuum of risk, with combustible products at the far end of that continuum. Even some early skeptics of nicotine vaping changed their minds and began to attend to the scientific evidence that increasingly addressed the uncertainties when it came to a harm reduction approach. Yet there was also a wide spectrum of what counted as harm reduction. For example, some proposals suggested taxing nicotine vaped products at the same rate as combustible products (described by some as harm reduction in name only). Others suggested leaving vape products untaxed while doubling the tax on combustibles to incentivize smokers to switch. Nonetheless, with different degrees of enthusiasm and intent, harm reduction was the new lingua franca by 2017 (*7*).

## VEXING TENSIONS

In the United States, broad comfort with using harm reduction as a kind of policy language evaporated in the face of evidence that both the promise and peril of safer nicotine delivery alternatives to combusted tobacco that we have imagined for more than half a century materialized at the intersection of the twin phenomena of increased youth vaping and acute lung injuries and deaths.

In the case of youth who would not likely otherwise smoke or use nicotine in any form, vaping offers no benefits and introduces potential harms of nicotine dependence and a possible transition to combustible products (*1*, *4*, *8*–*10*). U.S. surveys confirm a large increase in the proportion of high schoolers who reported any vaping in the past 30 days, from 11.7% in 2017 to 27.5% in 2019. As expected, the majority of vaping is infrequent (experimental) and there is a positive association between vaping and smoking, but the public health impact remains unknown, nor is there a consensus on whether such an association constitutes a causal pathway (*4*, *6*, *8*–*10*). The majority who vape (about 60% of those who experiment, about 89% of regular vapers) are also smokers or former smokers. The calculus is complex: In a policy landscape in which, in virtually all locales, vaped nicotine products and all types of tobacco (smokeless and combusted) can be legally purchased at age 18, large numbers of 12th graders who vape do so legally (*1*, *2*, *4*, *8*, *9*). Contemporaneously, population youth smoking rates dropped much faster in the years vaping surged the most (2013–2019) than in prior years, reaching record lows during that same period (*2*, *9*), which suggests that nicotine vape use may be replacing smoking more than promoting it (*1*, *4*, *6*, *8*, *9*).

**"...the evidence warns against prohibitionist measures."**

We share strong concern about the large surge in youth vaping (some call it an epidemic and point to studies of a possible but unproven causal gateway into smoking) and we promote harm minimization and management. Yet we suggest that careful analysis of all the data in context indicates that the net benefits of vaped nicotine products outweigh the feared harms to youth (*4*, *6*, *8*).

Complicating the question of the relative harms of youth vaping are data showing that some youth only vape flavored products (without nicotine or THC), and 41.8% of vaping youth report vaping THC. Some U.S. states have legalized cannabis for adults. In others it remains illegal. The impact of the black market, a source of contaminated THC-based oils, has been devastating. Although the carrier liquids and additives in THC oils are different and more hazardous than those used safely for a decade in commercial nicotine products or flavors alone, there are unknown risks from unregulated and illegally obtained products. Addressing the rapid rise in teen use of vaping is imperative. But public health measures must not neglect distinctions between nicotine and THC as well as between products obtained through the retail and black markets. The illegal marketplace provides access to both THC and nicotine-based products at costs lower than over the counter. Users can obtain products not available over the counter. In a black market, age is no barrier to access. Further, there is evidence that in the face of bans, adolescent vapers switch to smoking (*11*). Regulation inevitably produces the possibility of the black market, but the approach to regulation can make the black market more or less attractive and more or less harmful to users, be they youth or adults.

In the case of adult smokers, there is solid scientific evidence that vaping nicotine is much safer than smoking. In a 2018 report by the U.S. National Academies of Sciences, Engineering, and Medicine (NASEM), commissioned by the FDA, an expert panel systematically reviewed the scientific evidence. It determined, "There is conclusive evidence that completely substituting vaping nicotine for combustible tobacco cigarettes reduces users' exposure to numerous toxicants and carcinogens present in combustible tobacco cigarettes" (*10*), consistent with other major evidence and systematic reviews (*4*, *6*, *12*).

But what this NASEM determination of less harm meant for overall public health impact was another question. Concerned with uncertainties, particularly involving presumed high risks to youth, some of the NASEM report authors made clear that their findings should not be construed as blanket support for harm reduction. A contemporaneous systematic review conducted by PHE came to a different conclusion about the promise of nicotine vaping as a safer alternative to smoking by weighing concerns about youth differently (*4*, *6*, *8*–*10*).

Although it may be decades before we fully understand the long-term consequences of vaping nicotine without smoke, many argue that we know enough and stress that too many smokers die every day we delay taking reasonable and rational action based on the science to date. Evidence from multiple strong observational studies and randomized trials suggests that vaping nicotine is more appealing and more effective than NRT at displacing smoking (*4*, *6*, *8*, *13*). Vaping flavors with or without nicotine may appeal to youth, but flavors also appeal to adult smok-



Published by AAAS

ers and help them switch. Evidence suggests that the vast majority of smokers who successfully switch completely from smoking combustible products to vaping do so—after weeks, months, or years of dual use—by transitioning from vaping tobacco, or menthol-flavored liquids, to other flavors and often to lower nicotine concentrations or even to no nicotine in order to reduce the triggers that remind them of their prior smoking product (*4*, *6*, *13*, *14*).

## THE WAY FORWARD

Making policy in the absence of evidentiary certainty must involve trade-offs. Policy action has consequences for those who have never smoked, especially youth. It also has implications for current and future smokers. It is estimated that more than 1 billion smokers will die prematurely across the globe in the 21st century. We believe that the complex calculus of pros and cons warrants finding an optimal balance (*4*, *6*, *8*), thereby making fully regulated nicotine vaping products available to smokers while adopting forceful measures to limit the risks to and use by youth as much as possible. The UK, which embraced nicotine vaping harm reduction as a safer alternative to combustible products, has been able to accomplish appropriate regulation that has managed both youth nicotine uptake and helping adult smokers to quit. The UK, through the Medicines and Healthcare Products Regulatory Agency (MHRA), has a notification system that requires assurance by the maker about the safety and quality of any product on the market. The UK also prohibits sale of THC products. In addition to a system for reporting adverse events, the MHRA maintains a website so users can determine whether products are being legally sold. UK measures reflect regulatory requirements in the European Union.

Although the U.S. health care, advertising, and regulatory systems are different from those in the UK, there are measures we can take short of outright bans. The United States needs a regulatory infrastructure to ensure that products on the market are as safe as possible. The FDA should implement a product monitoring system or require manufacturers to conduct product monitoring under Section 915 of the Tobacco Control Act. Prudent product standards (neither overly burdensome nor too lax) should rapidly be promulgated by the FDA Center for Tobacco Products for vaping nicotine products as a class.

Menthol is the single most critical flavor when it comes to both adult and youth smoking. Despite two FDA-derived reports that recommended a ban on menthol in combustibles, there has been policy paralysis in the face of appalling evidence: 52% of all youth and more than 90% of African American youth initiate smoking with menthol. If we are going to take policy action on flavors, menthol in combustible products must be the first target.

The challenges of nicotine vaping also demand a rigorous system of surveillance that might detect an unanticipated harm early, similar to pharmaceutical post-market surveillance for adverse events. Although the CDC and FDA, through searching and rigorous analyses, are pinpointing the source of the sudden serious and deadly lung injuries (*3*), there is the risk that additional complications will emerge. Vitamin E acetate may not be the only chemical of concern. Ongoing surveillance is the best means of detecting harm in a situation where we may never have absolute certainty about safety.

No youth (for policy purposes, traditionally less than age 21) should use nicotine in any form (regardless of whether it is vaping or more harmful smoking) or use any form of THC. Current U.S. laws restrict purchase of alcohol to those 21 or older. The U.S. Institute of Medicine issued a 2015 report indicating that age 21 restrictions on tobacco sales would reduce teen tobacco uptake and save lives. Failure to promulgate age 21 purchase laws across the United States and enforce restrictions is unacceptable. Taxation has also proved an effective means of pricing products out of the hands of youth. Setting vaping nicotine taxes lower than those on combustible products can help keep products out of the hands of teens but still provide an incentive for adult smokers to switch. Communicating accurately the absolute and relative harms for vaping nicotine compared with smoking is critical so smokers can make informed decisions. Finally, predatory marketing to youth should be prohibited.

But appropriate regulation and strong limits on youth access will only address part of the U.S. problem. Although they are not nicotine or tobacco products, THC vaping products must be addressed nationwide.

## THREADING THE NEEDLE

Every day, more than 2500 U.S. teens start smoking, and about 1300 U.S. adults who cannot stop smoking cigarettes die prematurely; 5.6 million U.S. youth alive today will die from smoking, feeding the pipeline of 35 million U.S. adult smokers, more than half of whom will die prematurely, despite efforts to prevent all youth uptake. Sixteen million people in the United States suffer with smoking-related illnesses such as cancer, emphysema, chronic obstructive lung disease, and other debilitating chronic diseases.

The most conservative estimates suggest that were vaping nicotine to replace most smoking over the next 10 years, 1.6 million premature deaths would be avoided and 20.8 million quality adjusted years of life would be saved in the United States alone. The greatest gains would be among younger cohorts (*15*). Across the globe, more than 8 million smokers will die prematurely from smoking cigarettes, not from nicotine itself, in 2019 alone. The potential benefit of appropriately regulated, innovative, noncombusted nicotine modes of delivery could have a tremendous impact globally.

The mounting numbers of acute lung injuries and deaths linked to vaping illicit THC cartridges have understandably fueled a policy impulse to do something. Although blanket bans on all devices, all types of liquids (with or without nicotine or THC), or flavors other than tobacco may provide immediate relief to our collective sense of urgency when it comes to protecting youth, the landscape has changed over the past decade. The calculus is no longer limited to nicotine vaping. Proposed solutions that conflate vaping THC oils with nicotine or with flavors, and that may lose sight of population-wide issues while focusing on subsets of the population (*4*, *6*, *8*, *9*), may do more harm than good.

Whatever specific regulations and policies are promulgated at local, state, national, or international levels, policies or regulations must be risk-proportionate. The most harmful products on the nicotine-harm continuum, combustible products, should be much more aggressively and stringently regulated than less harmful noncombusted nicotine products. Policies that fail to differentiate will fail public health. ■

### REFERENCES AND NOTES


1. M. J. Jarvis, R West, J. Brown, *Qeios* 10.32388/745076.2 (2019).
2. A. S. Gentzke *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **68**, 157 (2019).
3. U.S. FDA, "Vaping Illness Update: FDA warns public to stop using Tetrahydrocannabinol (THC) products and any vaping products obtained off the street" (4 October 2019); www.fda.gov/consumers/consumer-updates/vaping-illness-update-fda-warns-public-stop-using-tetrahydrocannabinol-thc-containing-vaping.
4. D. B. Abrams *et al.*, *Prev. Med.* **117**, 88 (2018).
5. A. Fairchild, J. Colgrove, *Am. J. Public Health* **94**, 192 (2004).
6. A. McNeill, L. S. Brose, R. Calder, L. Bauld, D. Robson, *Evidence review of e-cigarettes and heated tobacco products 2018* (Public Health England, 2018).
7. A. L. Fairchild, J. S. Lee, R. Bayer, J. Curran, *N. Engl. J. Med.* **378**, 216 (2018).
8. K. E. Warner, D. Mendez, *Nicotine Tob. Res.* **21**, 41 (2019).
9. D. T. Levy *et al.*, *Tob. Control* **28**, 629 (2019).
10. National Academies of Sciences, Engineering, and Medicine, *Public Health Consequences of E-Cigarettes* (National Academies Press, 2018).
11. A. S. Friedman, *J. Health Econ.* **44**, 300 (2015).
12. M. L. Goniewicz *et al.*, *JAMA Netw. Open* **1**, e185937 (2018).
13. P. Hajek *et al.*, *Health Technol. Assess.* **23**, 1 (2019).
14. C. Russell, N. McKeganey, T. Dickson, M. Nides, *Harm Reduct. J.* **15**, 33 (2018).
15. D. T. Levy *et al.*, *Tob. Control* **27**, 18 (2018).


10.1126/science.aba0032



Published by AAAS

Sandbags surround Goldman Sachs the day after Superstorm Sandy hit New York City in 2012.

PHOTO: VICTOR J. BLUE/BLOOMBERG/GETTY IMAGES

CLIMATE POLICY

# Planning for a warmer world

## A pair of Obama advisers advocate for proactive climate solutions

*By* **Daniel A. Reifsnyder**

In *Building a Resilient Tomorrow*, Alice Hill and Leonardo Martinez-Diaz have put together a superb primer on responding to the impacts of climate change. The writing is simple and tight and should appeal to anyone interested in how to prepare for the future of our physical world.

Hill is a senior fellow for climate change policy at the Council on Foreign Relations, while Martinez-Diaz is the global director of the Sustainable Finance Center at the World Resources Institute. Both played senior roles in the Obama administration.

The book's first section—"Systems for Large-Scale Change"—addresses topics such as how to rebuild in response to natural disasters (where, when, and to what standards); legal liability when losses occur; and market approaches to help improve resilience, including disclosure requirements, capturing climate risk in bond and real estate markets, and strengthening standards to obtain or maintain insurance. Here, the authors point to "no more" moments that can bring about large-scale, decisive action, but they acknowledge that "the rest of the time we largely fail to learn from our experience." Take, for example, their stunning observation that 8 of the 18 most hurricane-prone states in the union have no mandatory, statewide building code.

Why does this happen? In their view, the reasons are many and include prohibitive construction costs, the potential for lost tax revenue, a fear of litigation, optimism bias (our tendency to believe that situations will resolve favorably), and inertia.

The second part of the book deals with finding ways to pay for resilience, gathering data and making it more usable, and getting around human biases that make us reluctant to act. Here, they describe the staggering costs borne by the federal government for disaster relief ($130 billion between 2005 and 2008, for example, spent mostly in response to Hurricanes Katrina, Rita, and Wilma). The authors cite former Treasury Secretary Henry Paulson's argument that if Republicans really care about limited government, they should care about controlling climate change before it results in never-ending climate bailouts.

The book's final section—"The Upenders"—deals with strengthening the healthcare system, addressing inequality and improving community, relocating entire communities and immigration, and reconceiving national security. Here, Hill and Martinez-Diaz deal sensitively with issues of inequality. "[M]obility in the age of climate change is not just about escaping an imminently approaching storm or wildfire. It's also about having the option to change your permanent address as climate conditions change," they write.

Each chapter is packed with specific examples, and the authors include hopeful innovations and approaches as well. In chapter 1, for example, they describe the "Dome of a Home" in Pensacola, Florida, that incorporated resilient design features, enabling it to withstand Hurricane Ivan in 2004 largely intact.

Particularly gripping is chapter 9, which focuses on relocating people in harm's way. For years, the issue of displacement and relocation was something of a taboo subject in international climate debates, both because it is so sensitive and because solutions are not readily apparent. Hill and Martinez-Diaz tackle this issue head-on, dealing with it in a sensitive but clear-eyed manner.

"Of all the hard lessons in this book, managing climate migration may be the hardest," they argue, acknowledging that even the Obama White House dropped "managed retreat" for more politically palatable alternatives. "Our political leaders will need to begin a national conversation on this sensitive topic, and the sooner the better," urge Hill and Martinez-Diaz, arguing that "[t]he earlier we start, the easier, and less costly, and less traumatic building resilience will be."

The end of each chapter includes a short section featuring policy recommendations. Surprisingly, these are sometimes less pointed than the suggestions put forward in the chapters themselves. In chapter 6, for example, they discuss the idea of "institutionalizing imagination," arguing that, "Part of the answer lies in making the exercise of imagination a regular, even mandatory practice … It's not about predicting the future, but about considering what different futures might look like regardless of how likely or unlikely they may seem." The granularity and urgency of this point seem lost in the subsequent policy recommendation: "Federal, state and local governments, as well as businesses should integrate regular climate-risk scenario analysis into key strategy processes."

Still, Hill and Martinez-Diaz have organized complex material across multiple disciplines in a way that makes their narrative accessible and stimulating. This book is a great entry point and a very welcome addition to the resilience literature. ■

**Building a Resilient Tomorrow**
*Alice C. Hill and Leonardo Martinez-Diaz*
Oxford University Press, 2019. 264 pp.

The reviewer is at the Frank Batten School of Leadership and Public Policy, University of Virginia, Charlottesville, VA 22904, USA, and is the former deputy assistant secretary for environment, U.S. State Department, Washington, DC 20520, USA. Email:dar7q@virginia.edu

10.1126/science.aaz6838



Published by AAAS

CHEMISTRY

# Mastering the element of competition

## Players race to research the periodic table in an engaging new board game

*By* **Jennifer V. Hines**

This year marks the 150th anniversary of Mendeleev's publication on the periodic table. *Periodic: A Game of the Elements* provides a fun way to celebrate the science behind it in an engaging, play-based format.

Cheerful colors create an inviting game setup, and pieces shaped like microscopes, test tubes, and flasks playfully underline the science theme. The game board displays the periodic table (without the lanthanides and actinides) and is color coded by element group and classification type. Lively visual references to the tools chemists use and the elements that compose selected materials are incorporated throughout the game, and an accompanying booklet connects the game with the science and history of the periodic table.

Players move around the board according to periodic trends related to atomic number, ionization energy, atomic radii, and atomic mass in order to "research" (land their marker on) specific elements. The direction in which one can move is defined by the trends but also summarized by arrows located along the bottom of the game board, so prior knowledge about the periodic trends is not required. To move around the table, players use "energy tokens" to activate the desired trend. "Lab tokens" in the shape of test tubes and round-bottom and Erlenmeyer flasks help players keep track of the points they have earned from researching elements.

Players can earn and maximize points using different strategies. "Goal cards" (arranged into four piles of increasing difficulty) provide groups of elements—related by a common use or other fact—that players can choose to research. Carbon, oxygen, and calcium are grouped on one card, for example, on the basis of the role all three play in the calcium carbonate in chalk, limestone, and seashells. The first player to successfully research all the elements on a goal card receives that card's points and an "award tile," which can be redeemed for additional movement around the game board. Players also select one "agenda card," which is kept hidden from the other players. Bonus points can be earned by achieving the objective stated on this card.

**Periodic: A Game of the Elements**
*John J. Coveyou and Paul Salomon*
Illustrator: *Tomasz Bogusz*
Genius Games, 2019.

Microscope markers track each player's progress along an outer course consisting of "element group cards." Progression around this track can only occur when a player ends a turn on one of the elements present in the next element group card, and points are tallied by moving a marker along the "academic track" on the game board. Each successively higher square is worth more points, but the three highest squares can only be occupied by a limited number of players. Once filled, no one else can advance to those squares. The limited availability of these higher positions on the academic track provides an extra layer of competition (one that might hit a bit too close to home for those in academia).

The game ends either when a goal card pile is depleted or the highest academic track spots are filled. The player who has earned the most points through a combination of researching goal card elements, advancing on the academic track, and achieving the agenda card objective(s) wins.

On its own, moving around the periodic table according to trends and winning points by researching elements is a good premise for a game. The extra layers of complexity resulting from the different point-acquisition modes create more opportunities for strategic game play. Experienced gamers will likely find this interesting, but it may disconcert some beginner players. Our group, consisting of two professional scientists and a college student, found the point acquisition for moving around the periodic table according to the goal cards the most interesting and intuitive part of the game.

While the rule book that comes in the box is useful, the online video instructions (accessed via QR code) are also worth watching. These can help with initial setup and offer players tips for developing gameplay strategies.

Overall, *Periodic* is an engaging game that leverages interesting aspects of the periodic table, without requiring prior knowledge for successful game play. The opportunities for strategic play provide a fun challenge, even for those who know the periodic table well. ■

The reviewer is at the Department of Chemistry and Biochemistry, Ohio University, Athens, OH 45701, USA. Email: hinesj@ohio.edu

Points are earned for researching elements, advancing on the academic track, and achieving secret objectives.

PHOTO: GENIUS GAMES, LLC

10.1126/science.aba2746



Published by AAAS

# RESEARCH

Computer-generated view of Mars at dawn reveals the surface topography

## IN *SCIENCE* JOURNALS

*Edited by* **Stella Hurtley**

MARTIAN ATMOSPHERE

### Mapping winds in Mars' upper atmosphere

The atmospheric loss processes that stripped Mars of most of its ancient atmosphere are poorly understood. Benna *et al*. analyzed atmospheric measurements collected by the Mars Atmosphere and Volatile Evolution (MAVEN) spacecraft as it repeatedly dipped into the red planet's upper atmosphere. Combining multiple observing modes allowed the authors to derive wind speeds and map the global circulation of the atmosphere at altitudes of ~150 kilometers. In some locations, winds followed the slope of the surface topography far below. Such insights into the upper levels of planetary atmospheres are limited, even for Earth. —KTS

*Science*, this issue p. 1363

MOLECULAR BIOLOGY

### Cohesin extrudes DNA loops

DNA is folded into loops in eukaryotic cells by a process that depends on a ring-shaped adenosine triphosphatase complex called cohesin. Davidson *et al*. and Kim *et al*. now show that in the presence of the NIPBL-MAU2 protein complex, the human cohesin complex can function as a molecular motor that extrudes DNA loops with high speed in vitro. In contrast to how it mediates sister chromatid cohesion, cohesin does not appear to entrap DNA topologically during loop extrusion. The results provide direct evidence for the loop extrusion model of chromatin organization and suggest that genome architecture is highly dynamic. —SYM

*Science*, this issue p. 1338, p. 1345

ANCIENT ATMOSPHERE

### Stepping to an internal beat

What caused the stepwise nature of the rise of molecular oxygen in Earth's atmosphere since it appeared in large quantities more than 2 billion years ago? Alcott *et al*. argue that a set of internal feedbacks involving the global phosphorus, carbon, and oxygen cycles, not individual external forces, could be responsible. Their model, which depends only on a gradual shift from reducing to oxidizing surface conditions over time, produces the same three-step pattern observed in the geological record. —HJS

*Science*, this issue p. 1333

HIGH-PRESSURE PHYSICS

### Diamond-based sensors

Material properties can change dramatically under pressure. Typically, to achieve high-pressure conditions, researchers place their samples in diamond anvil cells (DACs). However, monitoring the properties of the sample inside a DAC is tricky (see the Perspective by Hamlin and Zhou). Hsieh *et al*., Lesik *et al*., and Yip *et al*. developed monitoring techniques based on nitrogen-vacancy (NV) centers in diamond. The NV centers can act as sensors because their energy levels and the associated spectra are sensitive to strain and magnetic fields. This enabled optical readout of a spatially resolved signal. —JS

*Science*, this issue p. 1349, p. 1359, p. 1355; see also p. 1312

POLYMERS

### Strong and tough fibers

Dragline spider silk is known for its combination of strength and toughness, but this combination has been hard to replicate in synthetic fibers. Liao *et al*. electrospun polyacrylonitrile-*co*-methyl acrylate fibers modified with a small amount of poly(ethylene glycol)

PHOTO: NASA/JPL



Published by AAAS

bisazide (PEG-BA) (see the Perspective by Fox). After collecting the electrospun yarn, it was annealed under tension that both aligned the small fibers and cross-linked them together via the PEG-BA. As hoped, the overall properties were comparable to those of spider silk. —MSL

*Science*, this issue p. 1376; see also p. 1314

NEUROSCIENCE

## Cell-type identification in neural circuits

In many cases, a single molecular marker is insufficient to define a specific cell type and may label a few, or a few hundred, physiologically distinguishable cell types. Lee *et al*. developed a high-throughput technique, called physiological optical tagging sequencing (PhOTseq), for identifying the expression profile of cells that exhibit a particular physiological profile (see the Perspective by Renninger). They used PhOTseq to identify genes encoding vomeronasal receptors in mice, which detect pheromones and subserve social communication. —PRS

*Science*, this issue p. 1384; see also p. 1311

NEUROSCIENCE

## Sleep disruption in cognitive decline

Dementia affects nearly 36 million people worldwide, a number that is expected to double within the next two decades. Sleep disruption is thought to underlie some aspects of cognitive decline and neurodegeneration. Kaneshwaran *et al*. assessed sleep disruption in two cohort studies of older persons—the Rush Memory and Aging Project and the Religious Orders Study. In the subjects, greater sleep fragmentation was associated with higher neocortical expression of genes characteristic of microglial aging and activation. This valuable human data may provide therapeutic pointers for the treatment of neurodegenerative disorders. —KSL

*Sci. Adv.* 10.1126/sciadv.aax7331 (2019).

PSYCHIATRIC DISEASES

## Dissecting the effects of ecstasy

Owing to its prosocial effects, MDMA (commonly known as ecstasy) is being evaluated for the treatment of psychiatric disorders. Unfortunately, MDMA's rewarding addictive properties hinder its therapeutic use. Heifets *et al*. investigated the mechanisms mediating the effects of MDMA in mice. They found that the prosocial and rewarding effects were mediated by independent mechanisms: The prosocial effects were mediated by the serotoninergic system, whereas the rewarding effects required dopaminergic signaling. The results offer hope for the development of more specific therapeutics with fewer side effects. —MM

*Sci. Transl. Med.* **11**, eaaw6435 (2019).

BODY SIZE

## It's the prey that matters

Although many people think of dinosaurs as being the largest creatures to have lived on Earth, the true largest known animal is still here today—the blue whale. How whales were able to become so large has long been of interest. Goldbogen *et al*. used field-collected data on feeding and diving events across different types of whales to calculate rates of energy gain (see the Perspective by Williams). They found that increased body size facilitates increased prey capture. Furthermore, body-size increase in the marine environment appears to be limited only by prey availability. —SNV

*Science*, this issue p. 1367; see also p. 1316

# IN OTHER JOURNALS

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

LONGEVITY

## Immune-mediated longevity

The naked mole-rat is a fascinating example of an animal with an extended lifespan relative to its body size. Hilton *et al*. investigated whether there are distinct attributes of the immune system of this social rodent that could account for their healthful aging. They mapped the immune system of naked mole-rats using single-cell RNA sequencing and found a high ratio of myeloid cells to lymphoid cells compared with the immune system of short-lived mice. They also identified a previously undescribed subset of granulocyte cells and found that naked mole-rats lack natural killer cells. Perhaps this altered immune cell composition helps prevent aging-associated disease, including cancer, which allows naked mole-rats to live for such a long time. —GKA

*PLOS Biol.* **17**, e3000528 (2019).

The naked mole-rat (*Heterocephalus glaber*) has distinctive physiological adaptations, including those that may contribute to longevity.

PHOTO: ZSSD/MINDEN PICTURES

DNA DAMAGE

## Genomic sensitivity unmasked by UV

Exposure to ultraviolet (UV) light can cause DNA damage and is a risk factor for cancer. For individuals, it would be valuable to be able to measure past UV exposure and thus future risk of melanoma. Premi *et al*. developed methods to sequence and analyze UV damage in the genome of human fibroblasts and melanocytes. In melanocytes, specific genetic loci are exquisitely sensitive to UV. These hyperhotspots of DNA damage associate with specific DNA motifs in sequences adjacent to protein coding genes and are expressed. This means that in every instance of UV-induced sunburn, about 20 cellular pathways affecting cellular growth and physiology are hit at least once. This is where the risk from UV may lie, rather than in rare oncogenic mutations. —LMZ

*Proc. Natl. Acad. Sci. U.S.A.* **116**, 24196 (2019).



Published by AAAS

FISHERIES

## Fisherman's friends

The decline of fisheries is a story not just about fish, aquatic ecosystems, and sustainable food supplies but also about the people who base their livelihoods on fishing. Severe limitations imposed in recent years to maintain a sustainable cod fishery in the Northeastern United States have brought the catch to historic lows. Recognizing that fisheries management does not include any measure of the social impact of policy changes, Scyphers *et al*. surveyed the effect of the fisheries decline on people and communities. The results unveil severe psychological stress among fishing boat captains, exacerbated by ongoing lawsuits. Although trust for other fishers and the fishing business remains, low trust prevails for government bodies, fisheries management, and environmental nongovernmental organizations. The most distressed individuals were those without alternative income sources and with dependents at home. —PJH

*Proc. Natl. Acad. Sci. U.S.A.* **116**, 22912 (2019).

**The decline in fisheries has brought considerable psychological stress to coastal communities.**

SUPERNOVAE

## Type II supernovae from binary stars

When an isolated star of 8 to 25 solar masses runs out of fuel in its core, it explodes in a hydrogen-rich (type II) supernova. However, stars are often in binary systems. Interaction with a binary companion transfers mass between the stars, changing the stellar evolution. Zapartas *et al*. performed population synthesis simulations of five binary interaction channels that can produce a type II supernova, including in the initially less massive star or as the result of a stellar merger. They concluded that between one-third and one-half of all observed type II supernovae are the outcome of binary interactions, which may explain the diversity in observed properties of type II supernovae. —KTS

*Astron. Astrophys.* **631**, A5 (2019).

NANOMATERIALS

## Improving nanocrystal contacts

Semiconductor nanocrystals can be used in photodetectors and transistors by taking advantage of their size-tunable band gaps. However, organic coatings or poor contacts between crystals can create barriers that lead to carrier hopping rather than band transport. Walravens *et al*. deposited a close-packed monolayer of 6.5-nanometer lead selenide nanocrystals stripped of their organic coatings onto silicon or glass substrates. After annealing this superlattice at a temperature between 75° and 150°C, which preserved nanocrystal size and shape, x-ray diffraction showed a decrease in point defects and edge dislocations attributed to fewer interface stacking faults. Carrier mobility was increased by a factor of 10, and the band edge red-shifted up to 0.75 milli–electron volts, in good agreement with theoretical predictions for carrier delocalization. —PDS

*ACS Nano* **13**, 12774 (2019).

QUANTUM MECHANICS

## Quenched tunneling in ammonia

Inversion between two energetically equivalent umbrella-spaced configurations in ammonia ($NH_3$) is a classic textbook example of tunneling in a symmetric double-minimum potential. Park *et al*. report on how a very strong direct-current electric field, as powerful as intermolecular interactions, can suppress this quantum mechanical phenomenon in the system. They observed that $NH_3$ molecules isolated in a solid argon matrix at 10 kelvin strongly orient along the external field direction and that, as the field is increased up to 200 million volts per meter, their energy states shift, eventually quenching the tunneling from the normal state to the inverted one. The observed changes are reversible and open up opportunities for manipulating tunneling dynamics in many other molecules. —YS

*Proc. Natl. Acad. Sci. U.S.A.* **116**, 23444 (2019).

STEM CELLS

## Lubricating cell fate

The generation of hematopoietic stem cells (HSCs) is crucial for the treatment of diverse blood disorders. However, HSC turnover is high, and they are easily damaged, which means a risk for malignancy. Nevertheless, hematologic malignancies are rare, so Xie *et al*. investigated how human HSCs are maintained during stress and self-renewal. The authors found that the sphingolipid enzyme DEGS1, by modulating sphingolipid composition, mediates the shift from quiescent to activated HSCs and then regulates the differentiation of HSC lineages. If DEGS1 is inhibited, stress pathways are activated that converge on the endoplasmic reticulum, which removes damaged proteins. Obesity associates with altered HSC function in mice, perhaps because this condition also reprograms lipid-related HSC stress responses. —BAP

*Cell Stem Cell* **25**, 639 (2019).

PHOTO: PORTLAND PRESS HERALD/GETTY IMAGES



Published by AAAS

## ALSO IN *SCIENCE* JOURNALS

*Edited by* **Stella Hurtley**

CLIMATE CHANGE

### Measuring mitigation and adaptation

As more and more carbon dioxide is emitted into the atmosphere, humans and the natural world are beset by the damaging consequences of a rapidly changing climate. Natural and seminatural ecosystems are likely to be the best starting place for immediate adaptation and mitigation solutions. First, though, many natural environments need restoration to maximize their own resilience to climate change. In reviewing our options, Morecroft *et al*. point out that we can directly observe the success of mitigation strategies by quantifying atmospheric carbon dioxide. Successful adaptation is more challenging because it involves a range of social and biodiversity measures. However, we could make matters worse if we do not constantly monitor the effects of the interventions we devise and react flexibly as changing conditions unfold. —CA

*Science*, this issue p. 1329

GLOBAL CONSERVATION

### The time is now

For decades, scientists have been raising calls for societal changes that will reduce our impacts on nature. Though much conservation has occurred, our natural environment continues to decline under the weight of our consumption. Humanity depends directly on the output of nature; thus, this decline will affect us, just as it does the other species with which we share this world. Díaz *et al*. review the findings of the largest assessment of the state of nature conducted as of yet. They report that the state of nature, and the state of the equitable distribution of nature's support, is in serious decline. Only immediate transformation of global business-as-usual economies and operations will sustain nature as we know it, and us, into the future. —SNV

*Science*, this issue p. 1327

AGRICULTURE

### Mobile farming advice

Mobile phones are almost universally available, and the costs of information transmission are low. They are used by smallholder farmers in low-income countries, largely successfully, to optimize markets for their produce. Fabregas *et al*. review the potential for boosting mobile phone use with smartphones to deliver not only market information but also more sophisticated agricultural extension advice. GPS-linked smartphones could provide locally relevant weather and pest information and video-based farming advice. But how to support the financial requirements of such extension services is less obvious, given the unwieldiness of government agencies and the vested interests of commercial suppliers. —CA

*Science*, this issue p. 1328

STRUCTURAL BIOLOGY

### Revised view of anticancer drug mechanism

A crystal structure of the active form of cyclin-dependent kinase 4 (CDK4) provides insight into regulation of the cell cycle and the mechanism of action of a drug used for breast cancer therapy. The protein p27 has been thought to act as a CDK inhibitor. Guiley *et al*. performed a structural analysis of active complexes of CDK4 with cyclin D1 (CycD1) and p27 (see the Perspective by Sherr). The results showed that p27 actually remodels the active site of CDK4 to allow full activation when p27 is phosphorylated on tyrosine (phosp27). Furthermore, they found that the breast cancer drug palbociclib, a CDK4 inhibitor, doesn't actually interact with active phosp27-CDK4-CycD1 trimers. Instead, it appears that the drug, which shows promise in the clinic, binds to inactive CDK4 monomers and prevents interaction with p27. —LBR

*Science*, this issue p. 1330; see also p. 1315

MICROBIOTA

### Clostridial metabolite production

The clostridia are Firmicute bacterial commensals commonly found in the mammalian gut. Clostridia produce a range of metabolites that diffuse into the host's circulation and have been difficult to manipulate genetically, but Guo *et al*. successfully developed a CRISPR-Cas9 deletion system in *Clostridium sporogenes* (see the Perspective by Henke and Clardy). The authors used deletion mutants and mass spectrometry to elucidate clostridial synthesis of several different branched short-chain fatty acids (SCFAs), including isobutyrate, 2-methylbutyrate, and isovalerate. Germ-free mice colonized with mutants incapable of synthesizing SCFAs showed altered immunoglobulin A production. This finding potentially links bacterial SCFA production and host responses to the presence of the clostridia. —CA

*Science*, this issue p. 1331; see also p. 1309

MICROBIOTA

### Prospecting for drugs in the microbiome

The microbiome is an important source of natural products that can profoundly influence health and disease in the host. Sugimoto *et al*. constructed a modular, probabilistic strategy called *MetaBGC* to uncover biosynthetic gene clusters (BGCs) in human microbiome samples (see the Perspective by Henke and Clardy). The authors found geographic and strain-specific distributions of BGCs. By zeroing in on two type II aromatic polyketides, the native organisms were identified, the BGCs were reconstructed in *Streptomyces*, and the products were characterized. When expressed in *Bacillus subtilis*, the products resembled currently used anticancer drugs and antibiotics. These polyketides were not cytotoxic but had inhibitory activity against oral Gram-positive bacteria, which may reflect the niche and ecology of the originating organisms. —CA

*Science*, this issue p. 1332; see also p. 1309

GLYCOBIOLOGY

### A division of labor for glycosylation

Glycosylation is a ubiquitous modification of eukaryotic secreted proteins. Asparagine-linked chains of sugars are appended to many substrates as they are translocated into the endoplasmic reticulum. Ramírez *et al*. solved cryo–electron microscopy structures of two human oligosaccharyltransferase complexes, OST-A and OST-B. The catalytic subunits bind partner proteins that direct glycosylation of specific substrates either cotranslationally (OST-A) or on fully folded proteins (OST-B). High-resolution views of the active site and bound substrates in one of the complexes reveal important features of the human enzymes. —MAF

*Science*, this issue p. 1372

NANOMATERIALS

### Single-layer porphyrin polymerization

Two-dimensional polymers can be made as monolayer sheets through controlled synthesis at an interface. However, it is often difficult to create intact sheets over large areas that can



Published by AAAS

be transferred onto substrates. Zhong *et al*. polymerized derivatized porphyrin molecules during laminar flow at a sharp pentane-water interface to form sheets that are 5 centimeters in diameter (see the Perspective by MacLean and Rosei). The authors used electron microscopy and spectroscopy to confirm that they had produced intact monolayers. These films were then transferred onto monolayer sheets of molybdenum disulfide to form superlattices for use as capacitors. —PDS

*Science*, this issue p 1379; see also p. 1308

MUCOSAL IMMUNOLOGY

## Protecting intestinal stem cells

The intestinal epithelium is replaced every week. Thus, maintenance of this tissue requires rapid self-renewal driven by intestinal stem cells. This homeostasis is disrupted in a number of settings, including allogeneic bone marrow transplantation. After bone marrow transplantation, allogeneic T cells often attack and kill intestinal cells in an interferon-γ (IFN-γ)–dependent manner. Takashima *et al*. used in vivo transplant models and in vitro organoid systems to define the targets of IFN-γ in the mouse intestine. They found that IFN-γ directly targets intestinal stem cells. Furthermore, inhibitors of JAK-STAT signaling could be used to protect the intestinal stem cell compartment from T cell–mediated damage. —AB

*Sci. Immunol.* **4**, eaay8556 (2019).

PHYSIOLOGY

## Metabolic changes under pressure

Chronic hypertension causes irreversible damage to the kidneys. Rinschen *et al*. performed multiomics analyses of kidney tissue from rats that spontaneously develop hypertension when fed a high-salt diet. At early disease stages, kidney glomeruli showed metabolic changes. The alterations in kinase activity and metabolic enzyme abundance that occurred at later disease stages appeared to be triggered by the initial metabolic changes. Thus, metabolic interventions could potentially be useful in treating hypertension-induced kidney disease. —WW

*Sci. Signal.* **12**, eaax9760 (2019).



Published by AAAS

REVIEW SUMMARY

GLOBAL CONSERVATION

# Pervasive human-driven decline of life on Earth points to the need for transformative change

**Sandra Díaz*, Josef Settele, Eduardo S. Brondízio, Hien T. Ngo, John Agard, Almut Arneth, Patricia Balvanera, Kate A. Brauman, Stuart H. M. Butchart, Kai M. A. Chan, Lucas A. Garibaldi, Kazuhito Ichii, Jianguo Liu, Suneetha M. Subramanian, Guy F. Midgley, Patricia Miloslavich, Zsolt Molnár, David Obura, Alexander Pfaff, Stephen Polasky, Andy Purvis, Jona Razzaque, Belinda Reyers, Rinku Roy Chowdhury, Yunne-Jai Shin, Ingrid Visseren-Hamakers, Katherine J. Willis, Cynthia N. Zayas**

**BACKGROUND:** Human actions have long been known to drive declines in nature, and there is growing awareness of how globalization means that these drivers increasingly act at a distance (telecoupling). However, evidence from different disciplines has largely accumulated in parallel, and the global effects of telecouplings have never been addressed comprehensively. Now, the first integrated global-scale intergovernmental assessment of the status, trends, and future of the links between people and nature provides an unprecedented picture of the extent of our mutual dependence, the breadth and depth of the ongoing and impending crisis, and the interconnectedness among sectors and regions.

**ADVANCES:** Human impacts on life on Earth have increased sharply since the 1970s. The world is increasingly managed to maximize the flow of material contributions from nature to keep up with rising demands for food, energy, timber, and more, with global trade increasing the geographic separation between supply and demand. This unparalleled appropriation of nature is causing the fabric of life on which humanity depends to fray and unravel: Most indicators of the state of nature, whether monitored by natural and social scientists or by Indigenous Peoples and local communities, are declining. These include the number and population size of wild species, the number of local varieties of domesticated species, the distinctness of ecological communities, and the extent and integrity of many terrestrial and aquatic ecosystems. As a consequence, nature's capacity to provide crucial benefits has also declined, including environmental processes underpinning human health and nonmaterial contributions to human quality of life. The costs are distributed unequally, as are the benefits of an expanding global economy.

PHOTO CREDIT WWW.ESTEBANTAPELLA.COM

**Traditional diversity-rich human landscapes, and the livelihoods and identities that depend on them, face global threats.** Mosaics of crops, forest, and pasture have been maintained for millennia around the world. Now, they are under increasing threat from climate change and large-scale land use change to accommodate global demands for commodities. So are the livelihoods and cultural identity of the peoples that live in them, such as this woman collecting fodder for her flock in the Checacupe district, Perú.

These trends in nature and its contributions to people are projected to worsen in the coming decades—unevenly so among different regions—unless rapid and integrated action is taken to reduce the direct drivers responsible for most change over the past 50 years: land and sea use change, direct harvesting of many plants and animals, climate change (whose impacts are set to accelerate), pollution, and the spread of invasive alien species. Exploratory scenarios suggest that a world with increased regional barriers—resonating with recent geopolitical trends—will yield more negative global trends in nature, as well as the greatest disparity in trends across regions, greater than a world with liberal financial markets, and much greater than one that prioritizes and integrates actions toward sustainable development. Evidence from target-seeking scenarios and pathways indicates that a world that achieves many of the global biodiversity targets and sustainability goals related to food, energy, climate, and water is not—yet—beyond reach, but that no single action can get us there.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aax3100

**OUTLOOK:** Our comprehensive assessment of status, trends, and possible futures for nature and people suggests that action at the level of direct drivers of nature decline, although necessary, is not sufficient to prevent further deterioration of the fabric of life on Earth. Reversal of recent declines—and a sustainable global future—are only possible with urgent transformative change that tackles the root causes: the interconnected economic, sociocultural, demographic, political, institutional, and technological indirect drivers behind the direct drivers. As well as a pan-sectoral approach to conserving and restoring the nature that underpins many goals, this transformation will need innovative governance approaches that are adaptive; inclusive; informed by existing and new evidence; and integrative across systems, jurisdictions, and tools. Although the challenge is formidable, every delay will make the task even harder. Crucially, our analysis pinpoints five priority interventions ("levers") and eight leverage points for intervention in the indirect drivers of global social and economic systems where they can make the biggest difference. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: sandra.diaz@unc.edu.ar
Cite this article as S. Díaz *et al.*, *Science* **366**, eaax3100 (2019). DOI: 10.1126/science.aaw3100

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

REVIEW

GLOBAL CONSERVATION

# Pervasive human-driven decline of life on Earth points to the need for transformative change

Sandra Díaz[1,2]*, Josef Settele[3,4], Eduardo S. Brondízio[5], Hien T. Ngo[6], John Agard[7], Almut Arneth[8], Patricia Balvanera[9], Kate A. Brauman[10], Stuart H. M. Butchart[11,12], Kai M. A. Chan[13], Lucas A. Garibaldi[14], Kazuhito Ichii[15,16], Jianguo Liu[17], Suneetha M. Subramanian[18,19], Guy F. Midgley[20], Patricia Miloslavich[21,22], Zsolt Molnár[23], David Obura[24,25], Alexander Pfaff[26], Stephen Polasky[27,28], Andy Purvis[29,30], Jona Razzaque[31], Belinda Reyers[32,33], Rinku Roy Chowdhury[34], Yunne-Jai Shin[35,36], Ingrid Visseren-Hamakers[37,38], Katherine J. Willis[39,40], Cynthia N. Zayas[41]

**The human impact on life on Earth has increased sharply since the 1970s, driven by the demands of a growing population with rising average per capita income. Nature is currently supplying more materials than ever before, but this has come at the high cost of unprecedented global declines in the extent and integrity of ecosystems, distinctness of local ecological communities, abundance and number of wild species, and the number of local domesticated varieties. Such changes reduce vital benefits that people receive from nature and threaten the quality of life of future generations. Both the benefits of an expanding economy and the costs of reducing nature's benefits are unequally distributed. The fabric of life on which we all depend—nature and its contributions to people—is unravelling rapidly. Despite the severity of the threats and lack of enough progress in tackling them to date, opportunities exist to change future trajectories through transformative action. Such action must begin immediately, however, and address the root economic, social, and technological causes of nature's deterioration.**

Although previous large-scale environmental assessments have documented how human actions have been driving biodiversity loss and ecosystem deterioration, the recent Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services Global Assessment (*1*) has provided an unprecedentedly ambitious, interdisciplinary, and comprehensive synthesis of the evidence. It paints the clearest picture yet of how, despite humanity's profound dependence on nature, we are altering it at a truly planetary scale, with impacts that are distributed very unequally around the world and among sectors of society. This is the first comprehensive global assessment of nature that followed an intergovernmental process from start to end, covering not only the history of humanity's interactions with nature—with particular focus on the past 50 years—but also how these might change in the future. It was carried out by an independent, interdisciplinary team of experts from more than 50 countries within a framework that fully embraces the interdependence between people and nature (*2*–*4*), using a systematic approach that incorporates indigenous and local knowledge as well as the latest findings of natural and social sciences up to May 2018. Here, we distill the major findings of this report and augment them with more recent evidence.

## Taking stock of the fabric of life

The challenges of mitigating and adapting to climate change while achieving food, water, energy, and health security, and overcoming the unequal burdens of environmental deterioration and biodiversity loss, all rest on a common foundation: living nature. Specifically, we consider the fabric of life on Earth that has been "woven" by natural processes over many millions of years and in conjunction with people for many thousands of years. The vital contributions made by living nature to humanity, referred to as nature's contributions to people (*4*), affect virtually all aspects of human existence and contribute to achieving all the Sustainable Development Goals identified by the United Nations (*5*, *6*). These various contributions are now widely recognized in the scientific literature, but governmental policies and market transactions typically do not reflect their full value (*7*).

Human actions are causing the fabric of life to unravel, posing serious risks for the quality of life of people. Over the past 50 years, the capacity of nature to support quality of life has declined for 14 of the 18 categories of nature's contributions to people considered by the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) (Fig. 1). Nature's capacity to provide beneficial regulation of environmental processes—such as modulating air and water quality, sequestering carbon, building healthy soils, pollinating crops, and providing coastal protection from hazards such

[1]Consejo Nacional de investigaciones Científicas y Técnicas, Instituto Multidisciplinario de Biología Vegetal (IMBIV), Córdoba, Argentina. [2]Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba, Casilla de Correo 495, 5000, Córdoba, Argentina. [3]Department of Community Ecology, Helmholtz Centre for Environmental Research–UFZ, Halle, Germany. [4]German Centre for Integrative Biodiversity Research–iDiv, Leipzig, Germany. [5]Department of Anthropology, Indiana University, Bloomington, IN, USA. [6]Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) Secretariat, United Nations Campus, Platz der Vereinten Nationen 1, D-53113 Bonn, Germany. [7]Department of Life Sciences, University of the West Indies, St. Augustine Campus, Trinidad and Tobago. [8]Atmospheric Environmental Research, Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany. [9]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, CP 58190, Morelia, Michoacán, México. [10]Institute on the Environment, University of Minnesota, 325 Learning and Environmental Sciences, 1954 Buford Avenue, St. Paul, MN 55108, USA. [11]BirdLife International, David Attenborough Building, Pembroke Street, Cambridge CB2 3QZ, UK. [12]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. [13]Institute for Resources, Environment, and Sustainability, The University of British Columbia, Vancouver, Canada. [14]Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural, Universidad Nacional de Río Negro, Consejo Nacional de Investigaciones Científicas y Técnicas, Mitre 630, CP 8400, San Carlos de Bariloche, Río Negro, Argentina. [15]Center for Environmental Remote Sensing, Chiba University, 1-33,Yayoi-cho, Inage-ku, Chiba, 263-852, Japan. [16]Center for Global Environmental Research, National Institute for Environmental Studies, 16-2, Onogawa, Tsukuba, 305-0053, Japan. [17]Center for Systems Integration and Sustainability, Department of Fisheries and Wildlife, Michigan State University, 115 Manly Miles Building, East Lansing, MI 48823, USA. [18]United Nations University (UNU)–Institute for the Advanced Study of Sustainability, Tokyo, Japan. [19]UNU–International Institute for Global Health, Kuala Lumpur, Malaysia. [20]Global Change Biology Group, Department of Botany and Zoology, Stellenbosch University, P/Bag X1, Matieland 7602, South Africa. [21]Institute for Marine and Antarctic Studies, University of Tasmania, and Commonwealth Scientific and Industrial Research Organisation (CSIRO)–Oceans and Atmosphere, Hobart, Tasmania, Australia. [22]Departamento de Estudios Ambientales, Universidad Simón Bolívar, Caracas, Venezuela. [23]Centre for Ecological Research Institute of Ecology and Botany, Magyar Tudományos Akadémia, H-2163 Vácrátót, Hungary. [24]Coastal Oceans Research and Development–Indian Ocean (CORDIO) East Africa, Mombasa, Kenya. [25]Global Climate Institute, The University of Queensland, QLD 4072, Australia. [26]Sanford School of Public Policy, Duke University, Durham, NC 27708, USA. [27]Department of Applied Economics, University of Minnesota, 1994 Buford Avenue, St. Paul, MN 55108, USA. [28]Department of Ecology, Evolution, and Behavior, University of Minnesota, 1994 Buford Avenue, St. Paul, MN 55108, USA. [29]Department of Life Sciences, Natural History Museum, London SW7 5BD, UK. [30]Grand Challenges in Ecosystems and the Environment, Imperial College London, Ascot SL5 7PY, UK. [31]Department of Law, Faculty of Business and Law, University of the West of England, Bristol, Bristol, UK. [32]Stockholm Resilience Centre, Stockholm University, Sweden. [33]Department of Conservation Ecology, Stellenbosch University, Matieland, 7602, South Africa. [34]Graduate School of Geography, Clark University, Worcester, MA 01610, USA. [35]Marine Biodiversity, Exploitation and Conservation (MARBEC) Research Unit, Institut de Recherche pour le Développement (IRD), Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), Centre National de la Recherche Scientifique (CNRS), University of Montpellier, Montpellier, France. [36]Department of Biological Sciences, Marine Research Institute, University of Cape Town, 7701 Rondebosch, South Africa. [37]Department of Environmental Science and Policy, George Mason University, Fairfax, VA, USA. [38]Institute for Management Research, Radboud University, Nijmegen, the Netherlands. [39]Royal Botanic Gardens, Kew, Richmond, TW9 3AE, UK. [40]Long-Term Ecology Laboratory, Department of Zoology, University of Oxford, Oxford OX1 3SZ, UK. [41]Center for International Studies University of the Philippines, Diliman, Philippines.
*Corresponding author. Email: sandra.diaz@unc.edu.ar

**Fig. 1. Global trends in the capacity of nature to contribute to good quality of life from 1970 to the present.** Fourteen of the 18 categories of nature's contributions to people show a decline. Half of the 18 categories show consistent patterns globally, whereas the other half show declines in some regions and gains in others. For example, forest areas and the nature's contributions to people supported by forests have generally declined in tropical regions while increasing in some temperate areas over the past 50 years. Data supporting global trends and regional variations come from a systematic review of more than 2000 studies (*8*). Indicators were selected on the basis of availability of global data, prior use in assessments, and alignment with 18 categories. For many categories of nature's contributions, two indicators are included that show different aspects of nature's capacity to contribute to human quality of life within that category. Indicators are defined so that an increase in the indicator is associated with an improvement in nature's contributions. More details and illustrative examples of indicators and references are provided in in table S1. [Modified from (*1*).]

as storms and storm surges—has decreased globally, although for some benefits, trends vary by region (Fig. 1, third column) (*8*). For the 100 million to 300 million people who live in coastal areas below the 100-year flood level (*9*), the loss of coastal habitats has increased the risk of flooding and storm damage. Loss of animal pollinators affects more than 75% of global food crop types, risking US$235 billion to 577 billion of global crop output annually (*10*). The potential of nature to contribute in nonmaterial ways to human quality of life—through learning and inspiration, physical and psychological experiences, and supporting identities and sense of place—has also declined (Fig. 1). Exceptions to the downward trend of nature's contributions to people come from an increase in many of the material goods provided by nature, including food, energy, timber, and other materials (Fig. 1). For example, the harvest of commercial timber (*11*) and fish have both increased by almost 50% since 1970, and the value of agricultural crop production ($2.6 trillion in 2016) has increased approximately threefold. However, even within material contributions, some indicators show a strong decline, such as the abundance of marine fish stocks (*12*). In addition, the benefits of these increases have not been evenly distributed: Although the prevalence (percentage) of undernourishment has decreased globally in the past two decades, more than 800 million people still face chronic food deprivation (*11*).

The increase in the global production of consumer goods and the decline in almost all other contributions are directly related. The world is increasingly managed to accelerate the flow of material contributions from nature to keep up with rising demand. Since 1970, global population has doubled (*13*), per capita consumption has increased by 45%, the value of global economic activity as measured in gross domestic product (GDP) has increased by >300% (*14*), global trade has increased by ~900% (*15*), and the extraction of living materials from nature has increased by >200% (*16*).

The results of this unprecedented appropriation of nature can now be seen much more clearly than even 15 years ago (*17*), thanks to rapid advances in data and tools for observation, analysis, synthesis, and modeling of marine, freshwater, and especially terrestrial nature. These new data reveal that human actions have directly altered at least 70% of land surface (*18*, *19*); 66% of ocean surface is experiencing increasing cumulative impacts (*20*); around 85% of wetland area has been lost since the 1700s (*21*), and 77% of rivers longer than 1000 km no longer flow freely from source to sea (*22*). Coastal ecosystems show some of the largest and most rapid recent declines. Live coral cover on reefs has nearly halved in the past 150 years and is projected to virtually disappear this century unless there is strong climate change mitigation (*23*). Seagrass extent is decreasing by more than 10% per decade, while kelp forests have declined in 38% of their biogeographic regions (*24*).

The biomass of the world's vegetation (*25*) has halved over human history, and forests now span only 68% of their preindustrial extent

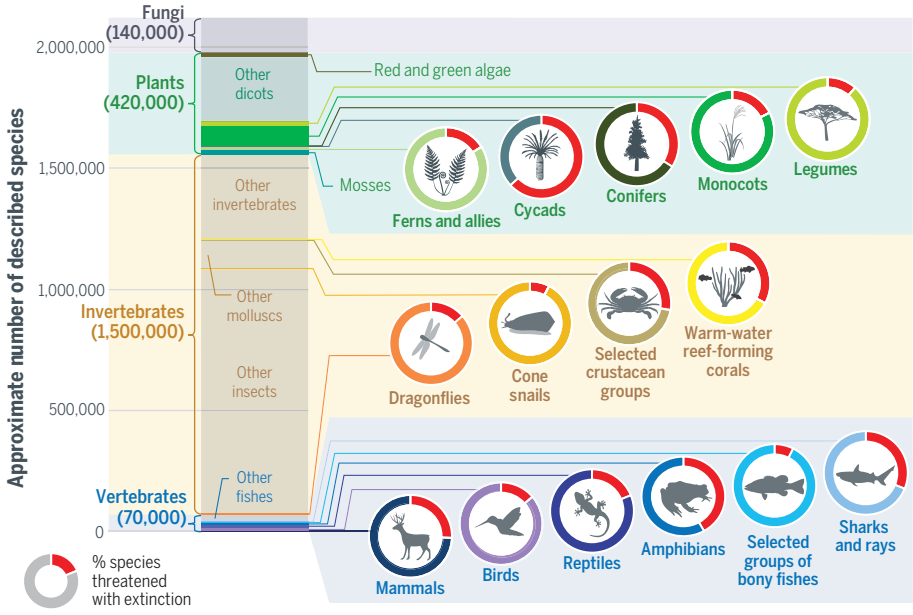



**Fig. 2. Extinction risk and diversity in different taxonomic groups.** Approximate number of described species of animals, plants, and fungi (bar) and the proportion of species that are threatened with extinction (pie charts) in groups that have been globally assessed for the IUCN Red List (*54*), either comprehensively or (for legumes, monocots, ferns and allies, dragonflies, and reptiles) through a sampled approach. Proportions assume that data deficient species are equally threatened as non–data deficient species. The proportions of data deficient species in each group are mammals, 15%; birds, 0.5%; reptiles, 21%; amphibians, 23%; bony fishes, 12%; sharks and rays, 42%; dragonflies, 35%; cone snails, 14%; crustaceans, 40%; corals, 17%; ferns, 0.4%; cycads, 1%; conifers, 1.2%; monocots, 12.1%; and legumes, 7.9%. The proportions of data deficient species in each realm are terrestrial, 10.7%; freshwater, 20.8%; and marine, 21.9%. [Sources: (*49*, *54*, *107*–*113*).]

(*26*). Although the rate of forest loss has slowed globally since the 1980s, it is still rapid in many tropical regions (*27*), and the increased extent of temperate and boreal forests (*28*) has been accompanied by increased fragmentation and changes in function [such as carbon storage (*29*)].

As a result of human impacts, terrestrial ecological communities worldwide are estimated to have lost more than 20% of their original biodiversity on average (*30*). In the ocean, animal populations and habitat extent have declined in the 20th century (*31*), with more than 20 described marine species having gone extinct (*32*). The status of marine fish populations has worsened globally, with one-third of the stocks being currently overfished (*12*). Illegal, unreported, and unregulated exploitation of fisheries undermine the effectiveness of stock management measures, especially in developing countries, contributing to conflicts (*33*, *34*). Many species that are large, slow-growing, habitat specialist, or carnivorous—such as large cats (*35*), large sharks (*36*), primates (*37*), reef-building corals (*38*), and woody plants (*39*)—are declining rapidly in, and being lost from, many places. On average, large terrestrial mammals have been extirpated from 75% of their natural ranges (*40*), while marine mammals have shown marked declines in abundance in recent centuries, many to near extinction (*41*, *42*). Endemic species have typically seen larger-than-average changes to their habitats and show faster-than-average population declines. By contrast, ecological generalists and disturbance-adapted species have tended to become more abundant, and some have spread quickly around the world (*43*). In 21 countries with detailed records, for example, the numbers of invasive alien species have risen by an average of 70% since 1970 (*44*). This combination of declining endemic species and the spread of already widespread species as humans purposefully or unwittingly transport species around the world drives "biotic homogenization" (*45*, *46*), a convergence of biological communities across regions that blurs the patterns on life's rich tapestry.

The number of species currently threatened with extinction is unprecedented in human history: an estimated 1 million species of animals and plants. This figure is derived from assessments for many terrestrial, freshwater, and marine vertebrate, invertebrate, and plant groups that have applied the International Union for the Conservation of Nature (IUCN) Red List categories and criteria (the global standard for assessing the relative extinction risk of each species) (*47*). On average across these groups, ~25% of species are threatened (classified as Critically Endangered, Endangered, or Vulnerable), although for insects—by far the most species-rich group—the proportion might be as low as 10% (Fig. 2) (*48*, *49*). Uncertainties in the numbers of animal and plant species and the percentage of them (particularly insects) that are threatened make any estimate necessarily approximate (*50*, *51*). Moving from species to populations, wild animal populations are continuing to decline on land and in the sea. The global biomass of wild mammals is now less than 25% of that before the late Pleistocene megafaunal extinction—and less than 10% that of the world's current human population (*52*). The global biomass of large predatory fish targeted by fisheries has fallen by two-thirds over the past 100 years (*53*). The outlook is worsening rapidly, with extinction risk increasing for all groups with known trends (*5*). A total of 705 vertebrate species are confirmed or presumed to have been driven extinct since 1500 (*54*), as have 571 plant species (*39*)—evidence that human actions have increased the global rate of species extinction by at least tens to hundreds of times over background rates before human intervention (*39*, *55*, *56*).

Domesticated species and varieties are also being lost. Fewer varieties of plants and animals are being maintained because of changes and standardization in farming practices, market preferences, large-scale trade, and loss of indigenous and local knowledge. Around 560 (~10%) of domesticated breeds of mammals had gone extinct by 2016, and at least 1000 more are threatened (*57*). Many hotspots of agrobiodiversity and of crop wild relatives are also under threat or lack formal protection (*58*), jeopardizing the pool of genetic variation that underpins the long-term resilience of agricultural production and food systems in the face of environmental change (*59*).

Declining trends are also documented in a worldwide evaluation of 321 indicators of nature important for quality of life developed by Indigenous Peoples and local communities. Although the decline in nature is lower in areas managed by Indigenous Peoples than in other lands (*60*), ~72% of the indicators assessed show deterioration (*51*).

In addition, rapid evolutionary responses to human drivers are now seen in all major taxonomic groups (*61*). This includes the evolution of resistance to pesticides and herbicides in insects and plants (*62*), smaller size and earlier maturation in marine fishes and invertebrates subject to fishing and global warming (*63*), changes in freeze tolerance in urban plants (*64*), and phenological shifts in a wide range of taxa in response to climate change (*65*).

### Direct and indirect drivers of change

The direct causes of changes observed in the fabric of life are (in decreasing order of relative impact worldwide) land and sea use change,

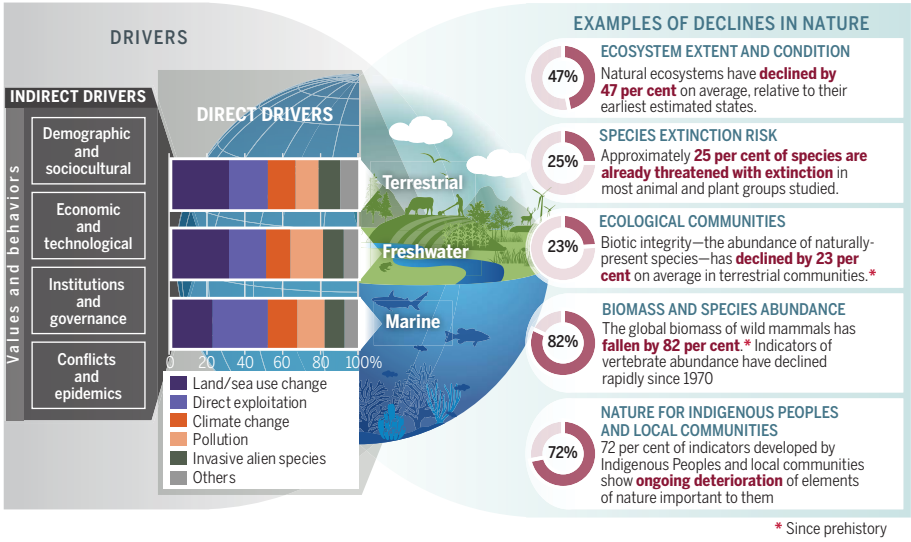



**Fig. 3. Examples of global declines in nature that have been and are being caused by direct and indirect drivers of change.** Each of the direct drivers of changes (land or sea use change; direct exploitation of organisms; climate change; pollution, including plastics, heavy metals, and direct effects of elevated $CO_2$ on, for example, terrestrial photosynthesis and seawater pH; and invasive alien species) represents the aggregation of many consequences from sectors such as crop production; animal husbandry; fishing; logging; hunting; mining for minerals, ores, and fossil fuels; development of cities and infrastructure for electricity and transport; and the transport of people and goods itself. The direct drivers result from an array of underlying societal causes. These causes can be demographic (for example, human population dynamics); sociocultural (for example, consumption patterns); economic (for example, trade); technological; or relating to institutions, governance, conflicts, and epidemics. These are called indirect drivers and are underpinned by societal values and behaviors (*3*, *114*). The color bands represent the relative global impact of direct drivers on (from top to bottom) terrestrial, freshwater, and marine nature as estimated from a global systematic review of studies published since 2005 (*51*). Land and sea use change and direct exploitation account for more than 50% of the global impact on land, in fresh water, and in the sea, but each driver is dominant in certain systems or places. The circles illustrate the magnitude of the negative human impacts on a diverse selection of aspects of nature over a range of different time scales, selected from a global synthesis of indicators; ecosystem extent, extinction risk, and biomass and species abundance include terrestrial, freshwater, and marine species and ecosystems, although most is known about life on land. Biotic integrity refers to the terrestrial realm only, and nature indicators for Indigenous Peoples and local communities are predominately terrestrial. [Reproduced from (*1*).]

exploitation of organisms, climate change, pollution, and invasive alien species (Fig. 3). Within terrestrial and freshwater ecosystems, the driver with the highest relative impact is land use change, mainly land conversion for cultivation, livestock raising, and plantations. The main driver in the ocean is direct exploitation through biomass extraction (mostly fishing). Although climate change is already a substantial driver of changes to nature and its contributions to people in many places (Fig. 3), even causing extinction in some cases [for example, (*66*)], it is not yet globally the most important.

The vast area of the world managed by Indigenous Peoples (at least 25 to 28% of land surface) (Fig. 4) under various property regimes is no exception to these trends. Because of their large extent, the fact that nature is overall better preserved within them (*60*), and because of the diverse stewardship practices carried within them around the world (Fig. 4, A to I), the fate of nature in these lands has important consequences for wider society as well as for local livelihoods, health, and knowledge transmission (*67*).

Indirect drivers of change—including demographic, economic, political, and institutional arrangements, and underpinned by societal values—underlie the observed direct drivers (Fig. 3). Indirect drivers interact with one another; for example, economic development choices could cause less deterioration in the presence of environmental policy, whereas the lack of publicly enforced rights could undermine resource management and conservation practices by Indigenous Peoples and local communities (*68*).

Over the past five decades, global socioeconomic trends have followed highly divergent pathways for countries with contrasting levels of income (Fig. 5) (*69*). With the dramatic increase in global trade, and more generally economic and social globalization, nature is ever more influenced by distant consumers (*70*). Trade has shifted where goods are produced and used, contributing to new economic opportunities but also generating or exacerbating inequities in both economic development and environmental burdens. The demand for material goods is predominantly from higher- and middle-income countries, and it is often satisfied by production in middle- and lower-income countries (Fig. 5, A, B, and C). For example, the European Union, the United States, and Japan together accounted for ~64% of the world imports of fish products in value, whereas developing countries accounted for 59% of the total volume of traded fish (*12*). These exchanges are often negotiated between actors and institutions of unequal power, which affects the distribution of the benefits and long-term social and ecological costs (Fig. 5F). A handful of transnational corporations control large (>50%) shares of supply chains in agriculture, fishing, logging, and mining (*71*, *72*), whereas funds channeled through tax havens support most illegal, unreported, and unregulated fishing (*71*, *73*), creating governance challenges. Many economic incentives are harmful to nature, including direct and indirect subsidies to fisheries (*74*), agriculture (including fertilizers and pesticides) (*75*), livestock raising, forestry, mining, and energy production (including fossil fuels and biofuels) (*76*). However, conservation policies (including incentives) could also play out unequally. For example, higher-income countries might contribute to the financing of environmental protection in lower-income countries but only to secure global benefits—such as the preservation of particular species and ecosystems, or carbon storage—whereas such policies can sometimes lower welfare locally (*77*, *78*).

### Progress toward internationally agreed goals

In view of the trends summarized above, it is not surprising that progress in meeting internationally agreed goals has been generally poor. Progress toward the 20 "Aichi Targets" in the Strategic Plan on Biodiversity 2011–2020 of the Convention on Biological Diversity has been mixed (Fig. 6A). Of the 54 elements comprising the 20 targets, good progress has been made toward five (9%), moderate progress toward 19 (35%), and poor progress or movement away from the target for 21 (39%). Progress is unknown for nine elements (17%). Overall, it is clear that the majority of Aichi Targets will not be met. More progress has been made in adopting and/or implementing policy responses and actions to conserve and use nature more sustainably than has been achieved in addressing the drivers of biodiversity loss. The strongest progress has been toward increasing protected area coverage

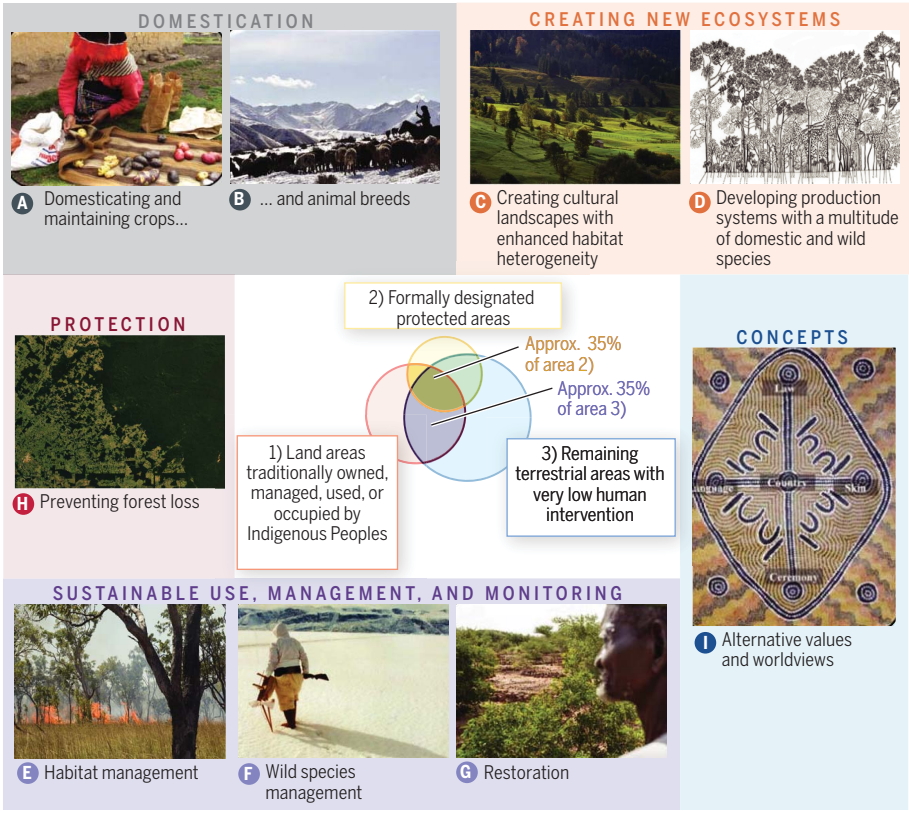



**Fig. 4. Contribution of Indigenous Peoples and local communities to biodiversity.** A wide range of practices of Indigenous Peoples and local communities maintain and enhance wild and domestic biodiversity. (**A** and **B**) Domestication and maintenance of locally adapted crop and fruit varieties and animal breeds (potatoes, Peru; rider and sheep, Kyrgyzstan) (*57*). (**C**) Creation of species-rich habitats and fine-grain habitat mosaics (hay meadows, Central Europe) (*115*). (**D**) Identification of useful plants and their cultivation in high-diversity agroecosystems (multispecies forest garden, Indonesia) (*116*). (**E** and **F**) Management and monitoring of wild species, habitats, and landscapes and (**G**) restoration of degraded lands (Australia, Alaska, and Niger) (*116*, *117*). (**H**) Prevention of deforestation in recognized Indigenous territories (Amazon basin, Brazil) (*68*). (**I**) Generation of alternative concepts of relations between humanity and nature (Northern Australia) (*118*). (Middle) Worldwide, the area traditionally owned, managed, used, or occupied by Indigenous Peoples (red circle), representing ~8 million km$^2$ in 87 countries, overlaps with at least 35 to 40% of the area that is formally protected (yellow circle), and a similar proportion of all remaining terrestrial areas with very low human intervention (blue circle) [<4 Human Footprint Index (*18*)] [based on (*60*)]. Circles and intersections are proportional in area. [Modified from (*1*).] [Photos credits: (A) FAO/Sandro Cespoli; (B) FAO/Vyacheslav Oseledko; (C) Daniel Babai; (D) (*118*); (E) Shutterstock_S. Todd; (F) Vadeve; (G) Rodrigo Ordonez/GLF; (H) Google Maps; (I) Daniel Rockman Jupurrurla.] The image in (I) represents Ngurra-kurlu, the Warlpiri people's understanding of how country contributes to people and vice versa. [Painting by Daniel Rockman Jupurrurla, from (*119*) and reproduced under the Creative Commons license.]

(Target 11) and developing national biodiversity strategy and action plans (Target 17). However, although protected areas now cover 14.9% of terrestrial and freshwater environments and 7.44% of the marine realm, they only partly cover areas of importance for biodiversity and are not yet fully ecologically representative, well-connected, and effectively and equitably managed. Although some species have been brought back from the brink of extinction (contributing toward Target 12 on preventing extinctions), the species in taxonomic groups with quantified trends are moving toward extinction at an increasing rate, meaning that this Target will not be met. Few data are available to quantify what the trends would have been in the absence of conservation action and policy responses to the Aichi Targets, although species' extinction risk trends would have been worse (*79*), and many island ecosystems that are recovering after eradications of invasive mammals would not have done so (*80*).

Nature and its contributions to people were found to underpin the achievement of all the United Nations Sustainable Development Goals (SDGs), either directly for goals on water, climate, oceans, and biodiversity (SDGs 6, 13, 14, and 15, respectively) or through more complex interactions for goals on poverty, hunger, health, cities, education, gender equality, reduced inequalities, and peace (SDGs 1, 2, 3, 4, 5, 10, 11, and 16, respectively). For SDGs on energy, economic growth, industry and infrastructure, and consumption and production (SDGs 7, 8, 9, and 12, respectively), important feedbacks between these goals and nature and its contributions to people were found with consequences for the achievement of all SDGs.

Declines in nature and its contributions to people therefore compromise our ability to meet the SDGs. At the target level, progress to 35 SDG targets that could be quantitatively assessed—on poverty, hunger, health, water, cities, climate, and biodiversity (SDGs 1, 2, 3, 6, 11, 13, 14, and 15 , respectively)—are being undermined by current trends in aspects of nature and its contributions to people relevant to these targets (Fig. 6B). However, current goal and target articulation omit or obscure the links to nature, preventing an assessment of other targets under these goals as well as targets in other goals linked to education, gender equality, reduced inequalities, and peace.

## Possible futures

We comprehensively reviewed both exploratory and target-seeking scenarios (*81*) of future change in direct and indirect drivers. These scenarios resulted in starkly different impacts on nature and its contributions to people and, in combination, enabled synthetic conclusions about the need for transformative change. We considered a wide range of exploratory scenarios on the basis of future plausible changes in direct and indirect drivers. A subset of the scenarios was based on Shared Socioeconomic Pathway (SSP) scenarios and Representative greenhouse gas Concentration Pathways (RCPs) developed in support of Intergovernmental Panel on Climate Change assessments (*82*). These combined scenarios ranged from "Global sustainability" [combining proactive environmental policy and sustainable production and consumption with low greenhouse gas emissions (SSP1 and RCP2.6)] to "Regional competition" [combining strong trade and other barriers and a growing gap between rich and poor with high emissions (SSP3 and RCP6.0)] to "Economic optimism" [combining rapid economic growth and low environmental regulation with very high greenhouse emissions (SSP5 and RCP8.5)]. Scenarios of a world with increased regional political and trade barriers tend to result in the greatest divergence across regions, scenarios that emphasize liberal financial markets result in intermediate levels of disparity, whereas scenarios that integrate actions toward sustainable development result in more modest differences between regions (*83*). Under business-as-usual future scenarios—meaning

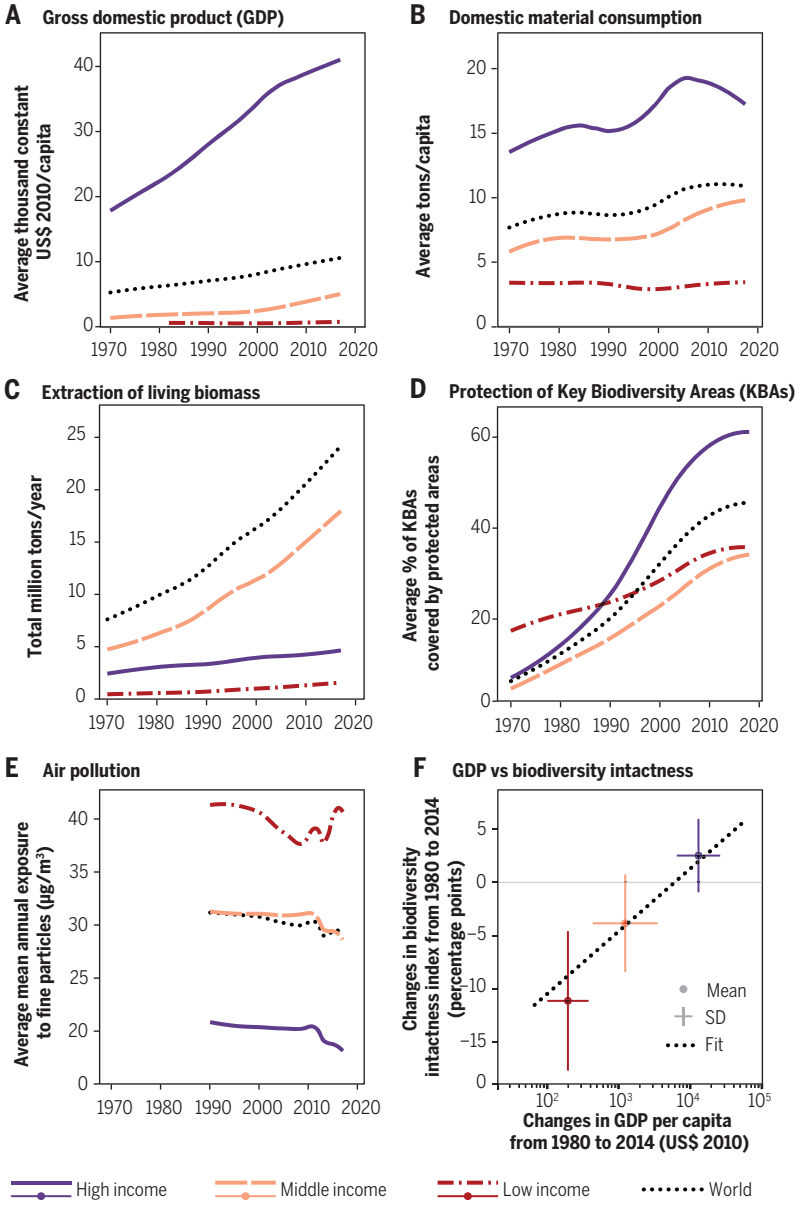



**Fig. 5. Development pathways since 1970 have featured unequal impacts on people and nature across countries.** (**A**, **B**, and **C**) Increased trade between countries has shifted the tradeoffs between environmental and other goals, with the footprints of increasing consumption in higher-income countries being exported to both middle-income and lower-income countries, who increased extraction of living materials. (**D**) Protection of key biodiversity areas has been highest in high-income countries, although international financing supported the protection of global public goods in low-income countries, which (**E**) have experienced much higher local air pollution given less support for local regulation and (**F**) not only the lowest increases in GDP but also the largest declines in some elements of nature. Countries are classified according to World Bank income categories. Data sources are (A) and (E), www.data.worldbank.org; (B) and (C), www.materialflows.net; (D), www.keybiodiversityareas.org and www.protectedplanet.net; and (F), www.data.worldbank.org and (*30*). [Modified from (*1*) and (*69*)]

that drivers of change do not deviate from the current socioeconomic and governance trajectory—nature in terrestrial, freshwater, and marine realms and most of its contributions to people will continue to decline sharply. Recent modeling of natural regulation of water quality, reduction of coastal risk, and crop pollination worldwide (*84*) found convergent conclusions.

Direct drivers of change that have predominated in the past 50 years (Fig. 3) will continue to play an important role (*19*, *83*, *85*), with climate change increasingly driving further biodiversity and ecosystem decline (*19*, *83*, *86*). These projections come with important uncertainties as to the degree of change or to the geographical differentiation of the impacts, depending on the underlying socioeconomic scenario. Given the interconnectedness of the world regions, future scenarios need to better address the impacts of telecouplings [socioeconomic and environmental interactions over distances (*70*)], such as trade, foreign direct investment, migrations, biological invasions, and pollutant flows (*87*). Projections also omit interconnections among species, which may cause domino effects that amplify the loss of diversity (*88*).

A different picture emerges from "target-seeking" scenarios (*81*), which start with a desirable target set in the future and then evaluate different pathways allowing to achieve it, including the innovations and policy interventions that are needed to reach such a target. Our analysis suggests that it is possible to achieve many of the global biodiversity targets and sustainability goals related to food, energy, climate, and water at local and global scales. The complexity of the challenges calls for an integrative (nexus) approach (*89*) that simultaneously examines interactions among multiple sectors along with synergies and tradeoffs among goals. An example of a key nexus are the simultaneous needs to mitigate climate change, arrest biodiversity loss, and ensure that all people have adequate nutrition on one hand, and the potentially negative consequences of large-scale land-based climate change mitigation on the other. Even moderate warming will likely be detrimental for biodiversity (*90*) and associated benefits to people (*91*). However, most scenarios projected to limit warming to 1.5°C or 2°C by the end of the 21st century rely on large-scale mitigation measures on land, in the form of bioenergy crops, reforestation, and/or afforestation, negatively affecting biodiversity and also food production and water demand (*19*, *92*). At the same time, expanding the amount of land devoted to agriculture to ensure that all people have adequate nutrition would negatively affect biodiversity as well (*93*) and would further exacerbate climate change (*19*, *92*). Both land-based climate change mitigation and agricultural expansion, when deployed at the large scale, can undermine local livelihoods, create access problems, and intensify social conflict (*94*). A suite of possible actions could be effective in navigating these tradeoffs (*19*, *95*)—for example, focusing on regeneration and restoration of high-carbon ecosystems (as well as reducing waste and overconsumption) rather than massive bioenergy monoculture plantations—to achieve climate change mitigation (*19*, *96*, *97*). Similarly, the increasing demands for food could be met without expanding agriculture's footprint by sustainably increasing yields, changing dietary choices, and

**Fig. 6. Summary of progress toward major internationally agreed goals.** (**A**) Progress toward the Aichi Biodiversity Targets contained in the Strategic Plan on Biodiversity 2011–2020 of the Convention on Biological Diversity. (**B**) Recent status of, and trends in, aspects of nature and nature's contributions to people that support progress toward achieving selected targets of the Sustainable Development Goals adopted through the United Nations in 2015. Scores in (A) for each of the 54 elements of the 20 targets are based on quantitative analysis of indicators, a systematic review of the literature, fifth National Reports to the Convention on Biological Diversity, and available information on countries' stated intentions to implement additional actions by 2020. Progress toward target elements is scored as "Good" (substantial positive trends at a global scale relating to most aspects of the element), "Moderate" (the overall global trend is positive but insubstantial or insufficient, or there may be substantial positive trends for some aspects of the element but little or no progress for others, or the trends are positive in some geographic regions but not in others), "Poor" (little or no progress toward the element or movement away from it; although there may be local, national, or case-specific successes and positive trends for some aspects, the overall global trend shows little or negative progress), or "Unknown" (insufficient information to score progress). In (B), selected targets are those for which current evidence and target wording enable assessment of the consequences for target achievement of trends in nature and nature's contribution to people. Scores for targets are based on systematic assessments of the literature and quantitative analysis of indicators where possible (*5*). "Poor/Negative" indicates poor status or substantial negative trends at a global scale; "Mixed" indicates the overall global status and trends are good or positive but insubstantial or insufficient, or there may be substantial positive trends for some relevant aspects but negative trends for others, or the trends are positive in some geographic regions but negative in others; "Unknown" indicates insufficient information to score the status and trends. An additional two targets under Goal 1 and two targets under Goal 3 were found to have evidence-based links to nature; however, because of the uncertain and complex relationships between nature and the target, they could not be assessed. [Data redrawn from (*1*) and (*5*).]

**A**

| Goal | Aichi Target (abbreviated) | Progress towards elements of each target | | | |
|---|---|---|---|---|---|
| | | Poor | Moderate | Good | Unknown |
| Drivers | Awareness | | ~~ | | |
| | Planning and accounting | ✗✗ | ~~ | | |
| | Incentives | ✗✗ | | | |
| | Production and consumption | ✗✗ | | | |
| Pressures | Habitat loss | ✗✗ | | | |
| | Fisheries | ✗✗ | | | ? |
| | Agriculture and forestry | ✗✗ | ~ | | |
| | Pollution | ✗✗ | | | |
| | Invasive alien species | ✗✗ | | ✓ | ? |
| | Coral reefs etc | ✗✗ | | | |
| Status | Protected and conserved areas | | ~~~~ | ✓✓ | |
| | Extinctions prevented | ✗✗ | | | |
| | Genetic diversity | | ~~~~ | | ? |
| Benefits | Ecosystem services | ✗ | | | ? |
| | Ecosystem restoration | | | | ?? |
| | Access and benefit sharing | | ~ | ✓ | |
| Implementation | Strategies and action plans | | ~~ | ✓ | |
| | Indigenous and local knowledge | | ~ | | ?? |
| | Biodiversity science | | ~ | | ? |
| | Financial resources | | ~ | | |

**B**

| Selected Sustainable Development Goals | Recent status and trends in aspects of nature and nature's contributions to people that support progress to selected targets | | |
|---|---|---|---|
| | Poor/Declining | Mixed | Unknown |
| No poverty | ↓↓ | | |
| Zero hunger | ↓ | →→→ | |
| Good health and well-being | | | ?? |
| Clean water and sanitation | ↓↓↓ | → | |
| Sustainable cities and communities | ↓↓↓↓ | → | |
| Climate action | ↓ | → | ??? |
| Life below water | ↓↓↓↓ | →→→ | |
| Life on land | ↓↓↓↓↓↓ | →→→→→ | |

reducing waste, among other measures (*19*, *98*, *99*).

More generally, solutions are needed that simultaneously address a nexus of relevant goals, such as feeding humanity, resourcing growing cities, mitigating climate change, protecting nature on land and at sea, maintaining freshwater, and ensuring animal welfare (Fig. 7). The futures that successfully address this suite of sustainability goals require rapid transition toward clean energy, a continued ramping up of biological conservation, large-scale restoration of degraded ecosystems, and transformation of supply chains to reduce resource extraction and environmental impacts. However, such comprehensive changes to direct drivers also require reform of indirect drivers, including innovations in economic and political structures and societal norms.

### Levers and leverage points for transformative change

Our assessment—the most comprehensive carried out to date, including the nexus analysis of scenarios and an expert input process with literature reviews—revealed clearly that reversing nature's ongoing decline (*100*) while also addressing inequality will require transformative change, namely a fundamental, system-wide reorganization across technological, economic, and social factors, making sustainability the norm rather than the altruistic exception. Achieving such a transformation for the broader current and future public good will have to overcome resistance from vested interests, including some powerful actors (*101*). One important avenue to transformation is the improved implementation and enforcement

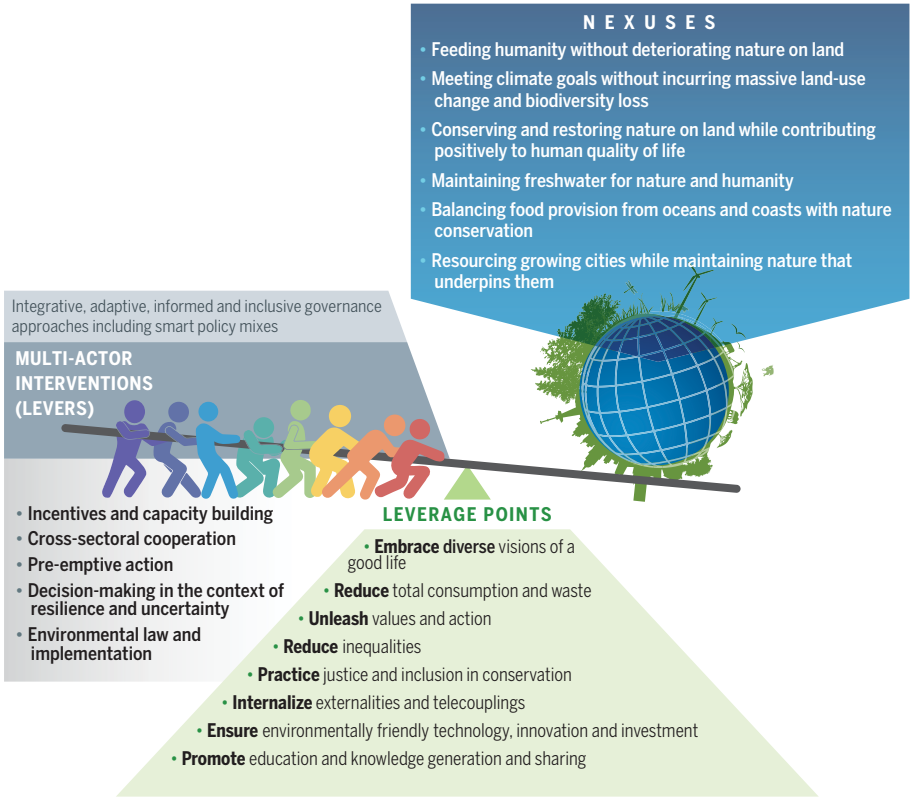



**Fig. 7. Enabling transformative change.** Collaborative implementation of priority interventions (levers) targeting key points of intervention (leverage points representing major indirect drivers) could enable transformative change from current trends toward more sustainable ones. Effectively addressing these levers and leverage points requires innovative governance approaches and organizing the process around nexuses, representing closely interdependent and complementary goals (*1*, *94*). [Modified from (*1*).]

of existing environmental policies and regulations and the removal and reform of harmful policies, such as subsidies for energy use or resource harvest (*102*). Another important step involves reforming global financial and economic systems, steering away from the current limited paradigm of economic growth to reward sustainability and penalize actions, resulting in the deterioration of the fabric of life (*103*, *104*). Such transformative change can be enabled, strengthened, and accelerated with the collaborative application of priority interventions (levers) to key points of intervention (leverage points) through innovative governance approaches (Fig. 7).

A comprehensive set of five levers (*95*) for this transformative change emerged from our unprecedentedly broad and rigorous analysis of the many possible levers that have been proposed previously: (i) developing incentives and widespread capacity for environmental responsibility and eliminating perverse incentives; (ii) reforming sectoral and segmented decision-making to promote integration across sectors and jurisdictions; (iii) taking preemptive and precautionary actions in regulatory and management institutions and businesses to avoid, mitigate, and remedy the deterioration of nature, and monitoring their outcomes; (iv) managing for resilient social and ecological systems in the face of uncertainty and complexity to deliver decisions that are robust in a wide range of scenarios; and (v) strengthening environmental laws and policies and their implementation, and the rule of law more generally.

The scenarios analysis and expert-input process further found that efforts focused on the following eight leverage points yield disproportionately large effects: (i) enabling visions of a good quality of life that do not entail ever-increasing material consumption; (ii) lowering total consumption and waste, including by addressing both population growth and per capita consumption differently in different contexts; (iii) unleashing existing, widely held values of responsibility to effect new social norms for sustainability, especially by extending notions of responsibility to include the impacts associated with consumption; (iv) addressing inequalities, especially regarding income and gender, that undermine the capacity for sustainability; (v) ensuring inclusive decision-making and the fair and equitable sharing of benefits arising from the use of and adherence to human rights in conservation decisions; (vi) accounting for nature's deterioration from both local economic activities and telecouplings (*70*), including, for example, international trade; (vii) ensuring environmentally friendly technological and social innovation, taking into account potential rebound effects and investment regimes; and (viii) promoting education, knowledge generation, and the maintenance of different knowledge systems, including in the sciences and indigenous and local knowledge, especially regarding nature, conservation, and nature's sustainable use. Although change at some of these levers and leverage points may encounter resistance individually, action at other levers and leverage points can enable such changes.

The review also revealed that innovative governance approaches that are integrative, inclusive, informed, and adaptive (*105*, *106*) are needed to effectively apply these levers to leverage points. Integrative approaches focus on the relationships between sectors and policies and ensure policy coherence and effectiveness, and inclusive approaches, including rights-based ones, reflect a plurality of values and thus promote equity. Informed governance entails new strategies for knowledge production and coproduction that are inclusive of diverse values and knowledge systems. Last, adaptive approaches—including learning, monitoring and feedback loops—help coping with inevitable uncertainties and complexities.

## Conclusions and outlook

The most comprehensive global review to date of the interrelationship between people and nature makes it evident that the challenges posed by biodiversity loss, climate change, and achieving a good quality of life for all are deeply interconnected and need to be addressed in an integrative manner—and urgently—from local to global levels. Maintaining a life-sustaining and life-fulfilling planet for humans and other species are thus one and the same challenge—a challenge that cannot be met by business as usual. However, a rich array of approaches and instruments are available that can, together, achieve sustainability. The transformations in economies, politics, and social systems that will be needed in order to deploy these changes in time and at scale can be triggered by a series of targeted interventions, especially at key points of leverage in indirect drivers. In this way, it is still possible to achieve a full suite of goals associated with feeding and resourcing humanity while maintaining and restoring the fabric of life that supports us all.

## REFERENCES AND NOTES

1. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), *Summary for Policymakers of the Global Assessment Report of the Intergovernmental*

*Science-Policy Platform on Biodiversity and Ecosystem Services*. S. Díaz *et al.*, Eds. (IPBES Secretariat, 2019).
2. G. M. Mace, Ecology. Whose conservation? *Science* **345**, 1558–1560 (2014). doi: 10.1126/science.1254704; pmid: 25258063
3. S. Díaz *et al.*, The IPBES Conceptual Framework - connecting nature and people. *Curr. Opin. Environ. Sustain.* **14**, 1–16 (2015). doi: 10.1016/j.cosust.2014.11.002
4. S. Díaz *et al.*, Assessing nature's contributions to people. *Science* **359**, 270–272 (2018). doi: 10.1126/science.aap8826; pmid: 29348221
5. S. H. M. Butchart *et al.*, in *Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, E. S. Brondízio *et al.*, Eds. (Secretariat of the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services, 2019).
6. With the exception of SDG 17 (strengthen the means of implementation and revitalize the global partnership for sustainable development).
7. A. D. Guerry *et al.*, Natural capital and ecosystem services informing decisions: From promise to practice. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7348–7355 (2015). doi: 10.1073/pnas.1503751112; pmid: 26082539
8. K. A. Brauman *et al.*, in *Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, E. S. Brondízio *et al.*, Eds. (Secretariat of the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services, 2019).
9. J. Hinkel *et al.*, Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3292–3297 (2014). doi: 10.1073/pnas.1222469111; pmid: 24596428
10. S. G. Potts *et al.*, Safeguarding pollinators and their values to human well-being. *Nature* **540**, 220–229 (2016). doi: 10.1038/nature20588; pmid: 27894123
11. Food and Agriculture Organization of the United Nations (FAO), *FAOSTAT Statistical Database* (FAO, 2019).
12. FAO, *The State of World's Fisheries and Aquaculture 2018. Meeting the Sustainable Development Goals* (Food and Agriculture Organization of the United Nations, Rome, 2018).
13. https://data.worldbank.org/indicator/SP.POP.TOTL
14. https://data.worldbank.org/indicator/NY.GDP.MKTP.KD
15. https://data.worldbank.org/indicator/NE.EXP.GNFS.KD
16. www.materialflows.net
17. Millennium Ecosystem Assessment, *Ecosystems and Human Well-being* (Island Press, 2005).
18. J. E. M. Watson *et al.*, Catastrophic declines in wilderness areas undermine global environment targets. *Curr. Biol.* **26**, 2929–2934 (2016). doi: 10.1016/j.cub.2016.08.049; pmid: 27618267
19. Intergovernmental Panel on Climate Change (IPCC), *IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*, A. Arneth *et al.*, Eds. (IPCC, 2019).
20. B. S. Halpern *et al.*, Patterns and emerging trends in global ocean health. *PLOS ONE* **10**, e0117863 (2015). doi: 10.1371/journal.pone.0117863; pmid: 25774678
21. N. C. Davidson, How much wetland has the world lost? Long-term and recent trends in global wetland area. *Mar. Freshw. Res.* **65**, 934 (2014). doi: 10.1071/MF14173
22. G. Grill *et al.*, Mapping the world's free-flowing rivers. *Nature* **569**, 215–221 (2019). doi: 10.1038/s41586-019-1111-9; pmid: 31068722
23. K. Frieler *et al.*, Limiting global warming to 2 C is unlikely to save most coral reefs. *Nat. Clim. Chang.* **3**, 165–170 (2013). doi: 10.1038/nclimate1674
24. K. A. Krumhansl *et al.*, Global patterns of kelp forest change over the past half-century. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13785–13790 (2016). doi: 10.1073/pnas.1606102113; pmid: 27849580
25. K.-H. Erb *et al.*, Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* **553**, 73–76 (2018). doi: 10.1038/nature25138; pmid: 29258288
26. Food and Agriculture Organization of the United Nations (FAO), *Global Forest Resources Assessment 2010* (FAO, 2010).
27. X.-P. Song *et al.*, Global land change from 1982 to 2016. *Nature* **560**, 639–643 (2018). doi: 10.1038/s41586-018-0411-9; pmid: 30089903
28. R. J. Keenan *et al.*, Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *For. Ecol. Manage.* **352**, 9–20 (2015). doi: 10.1016/j.foreco.2015.06.014
29. S. Gauthier, P. Bernier, T. Kuuluvainen, A. Z. Shvidenko, D. G. Schepaschenko, Boreal forest health and global change. *Science* **349**, 819–822 (2015). doi: 10.1126/science.aaa9092; pmid: 26293953
30. S. L. L. Hill *et al.*, Worldwide impacts of past and projected future land-use change on local species richness and the Biodiversity Intactness Index. *bioRxiv* 311787 [Preprint]. 1 May 2018.
31. J. Rice, in *The First Global Integrated Marine Assessment: World Ocean Assessment I*, United Nations (Cambridge Univ. Press, 2017).
32. E. Sala, N. Knowlton, Global marine biodiversity trends. *Annu. Rev. Environ. Resour.* **31**, 93–122 (2006). doi: 10.1146/annurev.energy.31.020105.100235
33. D. J. Agnew *et al.*, Estimating the worldwide extent of illegal fishing. *PLOS ONE* **4**, e4570 (2009). doi: 10.1371/journal.pone.0004570; pmid: 19240812
34. J. Spijkers *et al.*, Global patterns of fisheries conflict: Forty years of data. *Glob. Environ. Change* **57**, 101921 (2019). doi: 10.1016/j.gloenvcha.2019.05.005
35. W. J. Ripple *et al.*, Status and ecological effects of the world's largest carnivores. *Science* **343**, 1241484 (2014). doi: 10.1126/science.1241484; pmid: 24408439
36. N. K. Dulvy *et al.*, Extinction risk and conservation of the world's sharks and rays. *eLife* **3**, e00590 (2014). doi: 10.7554/eLife.00590; pmid: 24448405
37. A. Estrada *et al.*, Impending extinction crisis of the world's primates: Why primates matter. *Sci. Adv.* **3**, e1600946 (2017). doi: 10.1126/sciadv.1600946; pmid: 28116351
38. C. Wilkinson *et al.*, in *The First Global Integrated Marine Assessment: World Ocean Assessment I*, United Nations (Cambridge Univ. Press, 2017).
39. A. M. Humphreys, R. Govaerts, S. Z. Ficinski, E. Nic Lughadha, M. S. Vorontsova, Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nat. Ecol. Evol.* **3**, 1043–1047 (2019). doi: 10.1038/s41559-019-0906-2; pmid: 31182811
40. S. Faurby, J. C. Svenning, Historic and prehistoric human-driven extinctions have reshaped global mammal diversity patterns. *Divers. Distrib.* **21**, 1155–1166 (2015). doi: 10.1111/ddi.12369
41. J. Schipper *et al.*, The status of the world's land and marine mammals: Diversity, threat, and knowledge. *Science* **322**, 225–230 (2008). doi: 10.1126/science.1165115; pmid: 18845749
42. A. M. Magera, J. E. Mills Flemming, K. Kaschner, L. B. Christensen, H. K. Lotze, Recovery trends in marine mammal populations. *PLOS ONE* **8**, e77908 (2013). doi: 10.1371/journal.pone.0077908; pmid: 24205025
43. M. Tabarelli, C. A. Peres, F. P. L. Melo, The 'few winners and many losers' paradigm revisited: Emerging prospects for tropical forest biodiversity. *Biol. Conserv.* **155**, 136–140 (2012). doi: 10.1016/j.biocon.2012.06.020
44. S. Pagad *et al.*, IUCN SSC Invasive Species Specialist Group, Invasive alien species information management supporting practitioners, policy makers and decision takers. *Int. J. Appl. Res. Biologic. Invasions* **6**, 127 (2015).
45. T. Newbold *et al.*, Widespread winners and narrow-ranged losers: Land use homogenizes biodiversity in local assemblages worldwide. *PLOS Biol.* **16**, e2006841 (2018). doi: 10.1371/journal.pbio.2006841; pmid: 30513079
46. W. D. Pearse *et al.*, Homogenization of plant diversity, composition, and structure in North American urban yards. *Ecosphere* **9**, e02105 (2018). doi: 10.1002/ecs2.2105
47. International Union for the Conservation of Nature (IUCN), IUCN Red List Categories and Criteria. Version 3.1 (Gland, Switzerland, 2012).
48. www.iucn.org/content/european-red-list
49. V. Clausnitzer *et al.*, Odonata enter the biodiversity crisis debate: The first global assessment of an insect group. *Biol. Conserv.* **142**, 1864–1869 (2009). doi: 10.1016/j.biocon.2009.03.028
50. M. J. Caley, R. Fisher, K. Mengersen, Global species richness estimates have not converged. *Trends Ecol. Evol.* **29**, 187–188 (2014). doi: 10.1016/j.tree.2014.02.002; pmid: 24569101
51. A. Purvis *et al.*, in *Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, E. S. Brondízio *et al.*, Eds. (Secretariat of the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services, 2019).
52. Y. M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6506–6511 (2018). doi: 10.1073/pnas.1711842115; pmid: 29784790
53. V. Christensen *et al.*, A century of fish biomass decline in the ocean. *Mar. Ecol. Prog. Ser.* **512**, 155–166 (2014). doi: 10.3354/meps10946
54. IUCN, *The IUCN Red List of Threatened Species. Version 2019-2. Downloaded on 18 July 2019* (IUCN, 2019); www.iucnredlist.org.
55. S. L. Pimm *et al.*, The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344**, 1246752 (2014). doi: 10.1126/science.1246752; pmid: 24876501
56. G. Ceballos *et al.*, Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci. Adv.* **1**, e1400253 (2015). doi: 10.1126/sciadv.1400253; pmid: 26601195
57. FAO, *The State of the World's Animal Genetic Resources for Food and Agriculture*. B. Rischkowsky, D. Pilling, Eds. (FAO, 2016).
58. S. Pironon *et al.*, Potential adaptive strategies for 29 sub-Saharan crops under future climate change. *Nat. Clim. Chang.* **9**, 758–763 (2019). doi: 10.1038/s41558-019-0585-7
59. N. P. Castañeda-Álvarez *et al.*, Global conservation priorities for crop wild relatives. *Nat. Plants* **2**, 16022 (2016). doi: 10.1038/nplants.2016.22; pmid: 27249561
60. S. T. Garnett *et al.*, A spatial overview of the global importance of Indigenous lands for conservation. *Nat. Sustainabil.* **1**, 369–374 (2018). doi: 10.1038/s41893-018-0100-6
61. A. P. Hendry, K. M. Gotanda, E. I. Svensson, Human influences on evolution, and the ecological and societal consequences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20160028 (2017). doi: 10.1098/rstb.2016.0028; pmid: 27920373
62. F. Gould, Z. S. Brown, J. Kuzma, Wicked evolution: Can we address the sociobiological dilemma of pesticide resistance? *Science* **360**, 728–732 (2018). doi: 10.1126/science.aar3780; pmid: 29773742
63. A. Audzijonyte *et al.*, Trends and management implications of human-influenced life-history changes in marine ectotherms. *Fish Fish.* **17**, 1005–1028 (2016). doi: 10.1111/faf.12156
64. K. A. Thompson, M. Renaudin, M. T. J. Johnson, Urbanization drives the evolution of parallel clines in plant populations. *Proc. Biol. Sci.* **283**, 20162180 (2016). doi: 10.1098/rspb.2016.2180; pmid: 28003451
65. J. Merilä, A. P. Hendry, Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evol. Appl.* **7**, 1–14 (2014). doi: 10.1111/eva.12137; pmid: 24454544
66. J. C. Z. Woinarski, S. T. Garnett, S. M. Legge, D. B. Lindenmayer, The contribution of policy, law, management, research, and advocacy failings to the recent extinctions of three Australian vertebrate species. *Conserv. Biol.* **31**, 13–23 (2017). doi: 10.1111/cobi.12852; pmid: 27704619
67. UNEP, *Cultural and Spiritual Values of Biodiversity* (Intermediate Technology, 1999).
68. L. Porter-Bolland *et al.*, Community managed forests and forest protected areas: An assessment of their conservation effectiveness across the tropics. *For. Ecol. Manage.* **268**, 6–17 (2012). doi: 10.1016/j.foreco.2011.05.034
69. P. Balvanera *et al.*, in *Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, E. S. Brondizio *et al.*, Eds. (Secretariat of the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services, 2019).
70. J. Liu *et al.*, Framing sustainability in a telecoupled world. *Ecol. Soc.* **18**, art26 (2013). doi: 10.5751/ES-05873-180226
71. H. Österblom *et al.*, Transnational corporations as 'keystone actors' in marine ecosystems. *PLOS ONE* **10**, e0127533 (2015). doi: 10.1371/journal.pone.0127533; pmid: 26017777
72. C. Folke *et al.*, Transnational corporations and the challenge of biosphere stewardship. *Nat. Ecol. Evol.* **3**, 1396–1403 (2019). doi: 10.1038/s41559-019-0978-z; pmid: 31527729
73. V. Galaz *et al.*, Tax havens and global environmental degradation. *Nat. Ecol. Evol.* **2**, 1352–1357 (2018). doi: 10.1038/s41559-018-0497-3; pmid: 30104749
74. U. R. Sumaila, V. Lam, F. Le Manach, W. Swartz, D. Pauly, Global fisheries subsidies: An updated estimate. *Mar. Policy* **69**, 189–193 (2016). doi: 10.1016/j.marpol.2015.12.026
75. F. H. Oosterhuis, P. Ten Brink, Eds., *Paying the Polluter: Environmentally Harmful Subsidies and their Reform* (Edward Elgar, 2014).

76. P. Ten Brink, *The Economics of Ecosystems and Biodiversity in National and International Policy Making* (Routledge, 2012).
77. A. S. Pullin *et al*., Human well-being impacts of terrestrial protected areas. *Environ. Evid.* **2**, 19 (2013). doi: 10.1186/2047-2382-2-19
78. M. M. Bayrak, M. L. Marafa, Ten years of REDD+: A critical review of the impact of REDD+ on forest-dependent communities. *Sustainability* **8**, 620 (2016). doi: 10.3390/su8070620
79. M. Hoffmann *et al*., The impact of conservation on the status of the world's vertebrates. *Science* **330**, 1503–1509 (2010). doi: 10.1126/science.1194442; pmid: 20978281
80. H. P. Jones *et al*., Invasive mammal eradication on islands results in substantial conservation gains. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4033–4038 (2016). doi: 10.1073/pnas.1521179113; pmid: 27001852
81. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), *The Methodological Assessment on Scenarios and Models of Biodiversity and Ecosystem Services*, S. Ferrier *et al*., Eds. (Secretariat of the Intergovernmental Platform for Biodiversity and Ecosystem Services, 2016).
82. B. C. O'Neill *et al*., A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Clim. Change* **122**, 387–400 (2014). doi: 10.1007/s10584-013-0905-2
83. Y. J. Shin *et al*., in *Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, E. S. Brondízio *et al*., Eds. (Secretariat of the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services, 2019).
84. R. Chaplin-Kramer *et al*., Global modeling of nature's contributions to people. *Science* **366**, 255–258 (2019). doi: 10.1126/science.aaw3372; pmid: 31601772
85. N. Titeux *et al*., Biodiversity scenarios neglect future land-use changes. *Glob. Chang. Biol.* **22**, 2505–2515 (2016). doi: 10.1111/gcb.13272; pmid: 26950650
86. M. C. Urban, Climate change. Accelerating extinction risk from climate change. *Science* **348**, 571–573 (2015). doi: 10.1126/science.aaa4984; pmid: 25931559
87. E. K. Kapsar *et al*., Telecoupling research: The first five years. *Sustainability* **11**, 1033 (2019). doi: 10.3390/su11041033
88. G. Strona, C. J. A. Bradshaw, Co-extinctions annihilate planetary life during extreme environmental change. *Sci. Rep.* **8**, 16724 (2018). doi: 10.1038/s41598-018-35068-1; pmid: 30425270
89. J. Liu *et al*., Nexus approaches to global sustainable development. *Nat. Sustainabil.* **1**, 466–476 (2018). doi: 10.1038/s41893-018-0135-8
90. G. T. Pecl *et al*., Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* **355**, eaai9214 (2017). doi: 10.1126/science.aai9214; pmid: 28360268
91. O. Hoegh-Guldberg *et al*., in *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, V. Masson-Delmotte *et al*., Eds. (World Meteorological Organization, 2018).
92. Intergovernmental Panel on Climate Change (IPCC), *Summary for Policymakers* in *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, Masson-Delmotte *et al*., Eds. (World Meteorological Organization, 2018).
93. D. Tilman *et al*., Future threats to biodiversity and pathways to their prevention. *Nature* **546**, 73–81 (2017). doi: 10.1038/nature22900; pmid: 28569796
94. D. M. Cáceres, Accumulation by Dispossession and Socio-Environmental Conflicts Caused by the Expansion of Agribusiness in Argentina. *J. Agrar. Change* **15**, 116–147 (2015). doi: 10.1111/joac.12057
95. K. M. A. Chan *et al*., in *Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, E. S. Brondízio *et al*., Eds. (Secretariat of the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services, 2019).
96. J.-F. Bastin *et al*., The global tree restoration potential. *Science* **365**, 76–79 (2019). doi: 10.1126/science.aax0848; pmid: 31273120
97. S. L. Lewis, C. E. Wheeler, E. T. A. Mitchard, A. Koch, Restoring natural forests is the best way to remove atmospheric carbon. *Nature* **568**, 25–28 (2019). doi: 10.1038/d41586-019-01026-8; pmid: 30940972
98. J. Poore, T. Nemecek, Reducing food's environmental impacts through producers and consumers. *Science* **360**, 987–992 (2018). doi: 10.1126/science.aaq0216; pmid: 29853680
99. H. M. Tallis *et al*., An attainable global vision for conservation and human well-being. *Front. Ecol. Environ.* **16**, 563–570 (2018). doi: 10.1002/fee.1965
100. G. M. Mace *et al*., Aiming higher to bend the curve of biodiversity loss. *Nat. Sustainabil.* **1**, 448–451 (2018). doi: 10.1038/s41893-018-0130-0
101. L. Raymond, S. L. Weldon, D. Kelly, X. B. Arriaga, A. M. Clark, Making change: Norm-based strategies for institutional change to address intractable problems. *Polit. Res. Q.* **67**, 197–211 (2014). doi: 10.1177/1065912913510786
102. Organization for Economic Cooperation (OECD), *Environmentally Harmful Subsidies: Challenges for Reform* (OECD, 2005).
103. D. O'Rourke, N. Lollo, Transforming consumption: From decoupling, to behavior change, to system changes for sustainable consumption. *Annu. Rev. Environ. Resour.* **40**, 233–259 (2015). doi: 10.1146/annurev-environ-102014-021224
104. G. Kallis *et al*., Research on degrowth. *Annu. Rev. Environ. Resour.* **43**, 291–316 (2018). doi: 10.1146/annurev-environ-102017-025941
105. S. Sitch, P. M. Cox, W. J. Collins, C. Huntingford, Indirect radiative forcing of climate change through ozone effects on the land-carbon sink. *Nature* **448**, 791–794 (2007). doi: 10.1038/nature06059; pmid: 17653194
106. J. Razzaque *et al*., in *Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, E. S. Brondízio *et al*., Eds. (Secretariat of the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services, 2019).
107. Z. Q. Zhang, Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (Addenda 2013). *Zootaxa* **3703**, 1–82 (2013). doi: 10.11646/zootaxa.3148.1.1; pmid: 26146682
108. M. Böhm *et al*., The conservation status of the world's reptiles. *Biol. Conserv.* **157**, 372–385 (2013). doi: 10.1016/j.biocon.2012.07.015
109. N. A. Brummitt *et al*., Green plants in the red: A baseline global assessment for the IUCN sampled Red List Index for plants. *PLOS ONE* **10**, e0135152 (2015). doi: 10.1371/journal.pone.0135152; pmid: 26252495
110. E. N. Lughadha *et al*., Counting counts: Revised estimates of numbers of accepted species of flowering plants, seed plants, vascular plants and land plants with a review of other recent estimates. *Phytotaxa* **272**, 82 (2016). doi: 10.11646/phytotaxa.272.1.5
111. M. J. M. Christenhusz, J. W. Byng, The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201 (2016). doi: 10.11646/phytotaxa.261.3.1
112. *State of the World's Plants 2017*, K. J. Willis, Ed. (Royal Botanical Gardens, 2017).
113. *State of the World's Fungi 2018*, K. J. Willis, Ed. (Royal Botanical Gardens, 2018).
114. E. S. Brondízio *et al*., in *Global Assessment Report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*, E. S. Brondízio *et al*., Eds. (Secretariat of the Intergovernmental Science-Policy Platform for Biodiversity and Ecosystem Services, 2019).
115. Z. Molnár, F. Berkes, in *Reconnecting Natural and Cultural Capital: Contributions from Science and Policy*, M. L. Paracchini *et al*., Eds. (Joint Research Centre, 2018).
116. F. Berkes, *Sacred Ecology* (Routledge, 2017).
117. V. Reyes-García *et al*., The contributions of Indigenous Peoples and local communities to ecological restoration. *Restor. Ecol.* **27**, 3–8 (2019). doi: 10.1111/rec.12894
118. G. Michon, H. De Foresta, P. Levang, F. Verdeaux, Domestic forests: A new paradigm for integrating local communities' forestry into tropical forest science. *Ecol. Soc.* **12**, 1 (2007).
119. W. J. Pawu-Kurlpurlurnu *et al*., "Ngurra-kurlu: A way of working with Warlpiri people" (Desert Knowledge, 2008).


## ACKNOWLEDGMENTS

We thank the reviewers who provided thoughtful and constructive feedback on this manuscript, which resulted in considerable improvement to the manuscript. We are grateful to the IPBES (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services) Global Assessment Management Committee, review editors and numerous contributing authors, to the IPBES Secretariat and its Technical Support Units, and all the governments and organizations that supported author meetings. We are especially grateful to A. Larigauderie (IPBES executive secretary) and R. T. Watson (IPBES chair 2016–2019), for their strategic vision and continued advice, and to I. Baste (IPBES Bureau) and M. Stenseke, S. Demissew, and L. Dziba (MEP co-chairs) for their continued advice. We also acknowledge the contributions of D. Cooper, M. Guèze (GA Technical Support Unit); Y. O. Estrada (data visualization); and R. González, S. Hill, C. Hilton-Taylor, and M. Rivers (data and discussions for this paper). **Funding:** The authors received no specific funding for this work; all authors involved in IPBES assessments and deliverables are involved on a voluntary basis. The IPBES global assessment was made possible thanks to many generous contributions, including nonearmarked contributions to the IPBES trust fund from governments (Australia, Belgium, Bulgaria, Canada, Chile, China, Denmark, Estonia, European Union, Finland, France, Germany, India, Japan, Latvia, Luxembourg, Malaysia, Monaco, Netherlands, New Zealand, Norway, Republic of Korea, South Africa, Sweden, Switzerland, the United Kingdom, and the United States); earmarked contributions to the IPBES trust fund toward the global assessment [Germany, Canada, France (Agence Française pour la Biodiversité), Norway, the United Kingdom, and the United States]; and in-kind contributions targeted at the global assessment. All donors are listed on the IPBES web site: www.ipbes.net/donors. **Author contributions:** All authors volunteered and contributed their time in producing this publication and components of the underlying assessment report (*5*, *8*, *51*, *69*, *83*, *95*, *106*, *114*) and its summary for policymakers (*1*). We are grateful to the following lead authors, fellows, and chapter scientists of the IPBES Global Assessment: C. Adams, A. P. D. Aguiar, D. Armenteras, Y. Aumeeruddy-Thomas, X. Bai, L. Balint, Z. Basher, T. Bekele Gode, E. Bennett, Y. A. Boafo, A. K. Boedhihartono, P. Brancalion, E. Bukvareva, I. Chan, N. Chettri, W. L. Cheung, B. Czúcz, F. DeClerck, E. Dulloo, A. Fernandez-Llamazares, B. Gabrielyan, L. Galetto, K. Galvin, E. García Frapolli, A. P. Gautam, L. R. Gerber, A. Geschke, J. Gutt, S. Hashimoto, A. Heinimann, A. Hendry, G. C. Hernández Pedraza, T. Hickler, A. I. Horcea-Milcu, S. A. Hussain, M. Islar, U. Jacob, W. Jetz, J. Jetzkowitz, P. Jaureguiberry, M. S. Karim, E. Kelemen, E. Keskin, P. Kindlmann, M. Kok, M. Kolb, Z. Krenova, R. Krug, P. Leadley, M. Lim, J. Liu, G. Lui, A. J. Lynch, M. Mastrangelo, P. McElwee, L. Merino, P. A. Minang, A. Mohamed, A. Mohammed, I. B. Mphangwe Kosamu, E. Mungatana, R. Muradian, M. Murray-Hudson, T. H. Mwampamba, N. Nagabhatla, A. Niamir, N. Nkongolo, P. O'Farrell, T. Oberdorff, P. Osano, B. Öztürk, H. Palang, M. G. Palomo, I. Palomo, M. Panahi, U. Pascual, R. Pichs Madruga, P. Pliscoff, V. Reyes-García, C. Rondinini, G. M. Rusch, O. Saito, R. Salimov, J. A. Samakov, J. Sathyapalan, T. Satterfield, A. K. Saysel, E. R. Selig, O. Selomane, R. Seppelt, L. Shannon, A. U. B. Shrestha, A. Sidorovich, A. Simcock, G. S. Singh, B., J. Spangenberg, B. Strassburg, E. Strombom, D. Tarkhnishvili, N. Titeux, E. Turnhout, M. Verma, A. Viña, M. Wiemers, M. J. Williams H. Xu, D. Xue, T. Yue, and D. Zaleski. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/eaax3100/suppl/DC1
Table S1
References (*120*–*163*)

10.1126/science.aax3100

REVIEW SUMMARY

AGRICULTURE

# Realizing the potential of digital development: The case of agricultural advice

Raissa Fabregas, Michael Kremer, Frank Schilbach*

**BACKGROUND:** Sustainably raising agricultural productivity for the 2 billion people living in smallholder farming households in the developing world is critical for reducing world poverty and meeting rising food demand in the face of climate change. Nevertheless, most smallholder farmers have no access to science-based agricultural advice. The widespread adoption of basic mobile phone technology presents opportunities to improve upon existing in-person agricultural extension efforts that are expensive and fraught with accountability problems.

**ADVANCES:** Meta-analyses suggest that the transmission of agricultural information through mobile technologies in sub-Saharan Africa and India increased yields by 4% and the odds of adoption of recommended agrochemical inputs by 22%. The delivery of market information can have additional system-wide impacts, reducing price dispersion and lowering transaction costs. Given the low and rapidly declining cost of information transmission, benefits likely exceed costs by an order of magnitude. Even basic phones and inexpensive text and voice messages can influence farmer behavior. Smartphones with GPS systems create the potential for larger gains through the transmission of more sophisticated media, such as videos, and for locally customized information on soil characteristics, weather, and pest outbreaks, delivered at the appropriate time during the agricultural season.

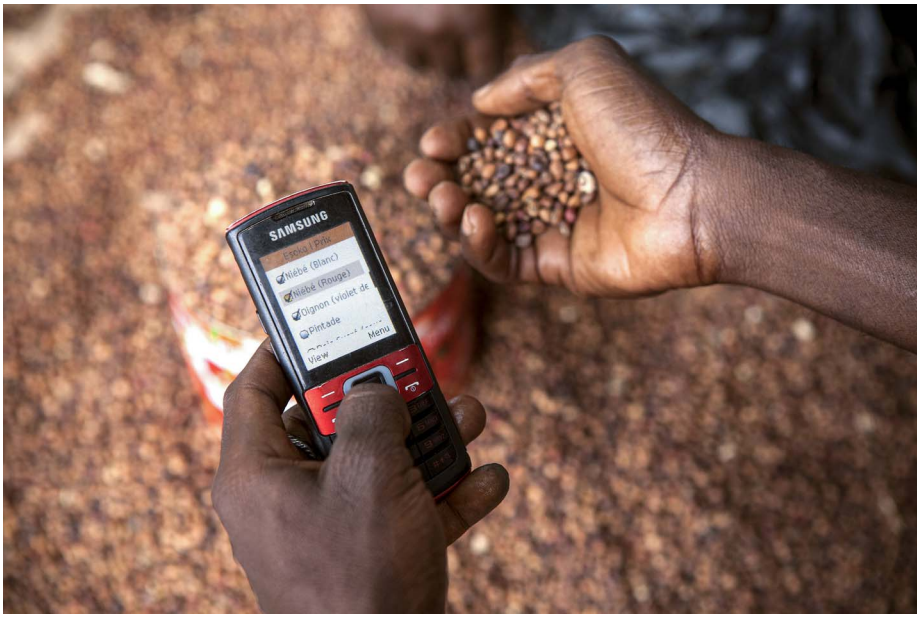
Messages could be customized on the basis of farmer characteristics, such as education or financial circumstances. Experimentation, machine learning, and two-way communication with and between farmers could facilitate improvements of information and other services over time. Advances from behavioral science can improve information transmission and address behavioral barriers to the adoption of improved agricultural techniques. Mobile phone–based systems could increase the productivity and accountability of in-person extension agents and enhance supply chain functionality. Realizing the potential of digital agriculture will require an interdisciplinary effort to develop and rigorously test a variety of approaches, incorporating insights from behavioral science, agriculture, economics, and data science.

**OUTLOOK:** Multiple market failures associated with information markets limit the ability of mobile phone–based extension systems to reach socially efficient scale through purely commercial financing. Because the marginal costs of disseminating information are close to zero, the optimal scale of such systems is very large. However, fixed system development costs still must be covered. Multiple organizations have introduced digital agricultural extension systems with financial models based on selling subscriptions to individual farmers, but such systems have been able to reach only a small fraction of farmers in the developing world. Farmer payments may be insufficient to cover the fixed costs, because information is difficult to exclude from nonpurchasers and because it is challenging for farmers to verify the quality of the information. Existing evidence suggests substantial gaps between farmers' willingness to pay for information and its social value. Advertising or agrochemical input sales could be used to finance information provision, but this approach could incentivize providers to distort information content in the absence of strong reputational costs of misinformation or appropriate regulation.

Public financing could cover fixed costs and enable scale-up. Although agriculture ministries often deliver messages in ways that farmers find difficult to understand and use, recent examples suggest that if feedback mechanisms are in place, governments can improve their services over time. Models that incentivize farmers to share their experiences create scope for customization and efficiency gains as systems grow, because this data may be used to improve recommendations for other farmers. If successful, digital agricultural advisory systems could supply a model for digital development more broadly. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aay3038




**Mobile phones can benefit farmers in low- and middle-income countries by improving access to agricultural advice and market price information.** Mobile technologies, particularly smartphones, have the potential to bring sophisticated science-based agricultural advice to smallholder farmers to improve productivity, especially under rapidly changing economic and environmental conditions. However, market failures likely preclude efficient scaling of valuable digital advice applications.

PHOTO: JAKE LYELL/ALAMY STOCK PHOTO

The list of author affiliations is available in the full article online.
*Corresponding author. Email: fschilb@mit.edu
Cite this article as R. Fabregas *et al.*, *Science* **366**, eaay3038 (2019). DOI: 10.1126/science.aay3038

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

REVIEW

AGRICULTURE

# Realizing the potential of digital development: The case of agricultural advice

Raissa Fabregas[1], Michael Kremer[2], Frank Schilbach[3]*

The rapid spread of mobile phones creates potential for sustainably raising agricultural productivity for the 2 billion people living in smallholder farming households. Meta-analyses suggest that providing agricultural information via digital technologies increased yields by 4% and the odds of adopting recommended inputs by 22%. Benefits likely exceed the cost of information transmission by an order of magnitude. The spread of GPS-enabled smartphones could increase these benefits by enabling customized information, thus incentivizing farmers to contribute information to the system. Well-known distortions in markets for information limit the ability of such systems to reach the socially efficient scale through purely commercial means. There is a clear role for public support for digital agricultural extension, but messages designed by agricultural ministries are often difficult for farmers to understand and use. Realizing the potential of mobile communication systems requires feedback mechanisms to enable rigorous testing and continuous improvement.

Mobile phones have penetrated the developing world to a greater extent than most other technologies (Fig. 1). More than three of four people in low- and middle-income countries (LMICs) own a phone. Approximately one in three people have internet access, and access is expected to increase markedly as smartphone costs decline (*1*).

The spread of phones presents opportunities for digital development by reducing information acquisition costs, allowing customization of information, and enabling monitoring and accountability in public services (*1*–*3*). Digital technologies have been deployed in a range of sectors—including finance, education, health, and civic participation—to improve development outcomes (*1*, *4*, *5*).

The proliferation of phones may also carry risks, such as the potential to exacerbate violent conflict (*6*), enable state surveillance and propaganda (*7*), accelerate the spread of fake news via social media, or further widen inequality because of uneven access to digital technologies (*8*, *9*). Finance and governance systems will affect the sustainability, scale, equitable reach, and effective design and implementation of these systems.

We review the evidence on "digital agriculture." With current technologies, impacts on farming practices and yields are modest in absolute terms but large relative to the cost of information delivery. The spread of GPS-enabled smartphones will create opportunities for customization and two-way communication, but an interdisciplinary effort will be required to experiment with different approaches and rigorously measure impact. Distortions in the markets for information limit the ability of systems to reach the socially efficient scale through purely commercial means, such that scaling programs beyond their current modest levels will likely involve an active public-sector role.

[1]Lyndon B. Johnson School of Public Affairs, University of Texas at Austin, Austin, TX 78712, USA. [2]Department of Economics, Harvard University, Cambridge, MA 02138, USA. [3]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: fschilb@mit.edu

## Traditional agricultural extension

Raising agricultural productivity is critical to reducing poverty and satisfying the growing global food demand (*10*) in the face of environmental stress and climate change. Improved access to agricultural information and targeting of agricultural inputs can raise agricultural productivity and reduce negative environmental footprints (*11*, *12*).

Nevertheless, most smallholder farmers lack access to science-based agricultural advice. Although ~400,000 agricultural extension agents (*13*) are employed by governments in LMICs, the ratio of farmers to extension workers exceeds 1000 to 1 in many countries (*14*). Transport budgets are often meager, and training, management, and accountability of extension workers are inadequate. In India, only 6% of farmers report having received any advice from an extension agent in the past year, and 70% of farmers distrust extension worker recommendations (*15*).

More generally, there is limited evidence of extension services' impact or cost effectiveness (*13*, *16*). Extension workers have been found to favor their own social networks (*17*) and neglect the most vulnerable farmers (*18*, *19*) and women (*20*, *21*).

## Digital agriculture: Potential and challenges

There is good reason to believe that emerging digital technologies can improve the functioning of agriculture markets at a very low cost per farmer. Establishing initial mobile phone coverage involves fixed costs, but the marginal cost of phone communication in rural areas is close to zero because cell phone towers typically operate below capacity. Cellular phone companies charge prices well above marginal cost, but they are often highly regulated, and governments could negotiate access at prices with lower markups.

Mobile phones, particularly GPS-enabled smartphones, facilitate the provision of tailored information. Recommendations for agrochemical inputs that address specific soil conditions on the basis of digital maps can improve yields while reducing environmentally harmful and wasteful use (*22*–*24*). Messages can target specific areas with reported pest outbreaks or be customized to other local conditions such as market prices. Farmers can tailor their investment decisions to expected weather patterns and benefit from improvements in weather forecasting (*25*, *26*). Customized information allows farmers to choose language, dialect, or literacy levels. Mobile technologies can also provide reminders and other nudges to address behavioral biases (*27*).

Running these systems at scale allows for testing variations to establish the most effective approaches (A/B testing) and feedback loops to improve accuracy and effectiveness of messages over time. Images taken from satellites can provide rich data about crop growth and, when linked with Geographic Information System (GIS) on plot boundaries, can improve measurements of productivity at scale and allow for ongoing experimentation (*28*, *29*). Mobile phones facilitate two-way communication, whereby farmers can ask questions and request information. Such platforms can also provide opportunities for networking and information exchange among farmers. Information from farmers using the system can further improve future recommendations for all users.

As smartphone use continues to expand, farmers will increasingly have the means to watch videos demonstrating new agricultural techniques or take pictures of pests affecting their crops and either request automatic identification and recommendations or raise questions with agronomists (*30*). Smartphones may also provide farmers access to interventions and apps that can enhance psychological well-being (*31*). Increased aspirations, grit, and improved mental health may boost farmer income by increasing investment and facilitating learning among farmers (*32*–*34*).

Mobile phones may create opportunities to complement and strengthen existing in-person

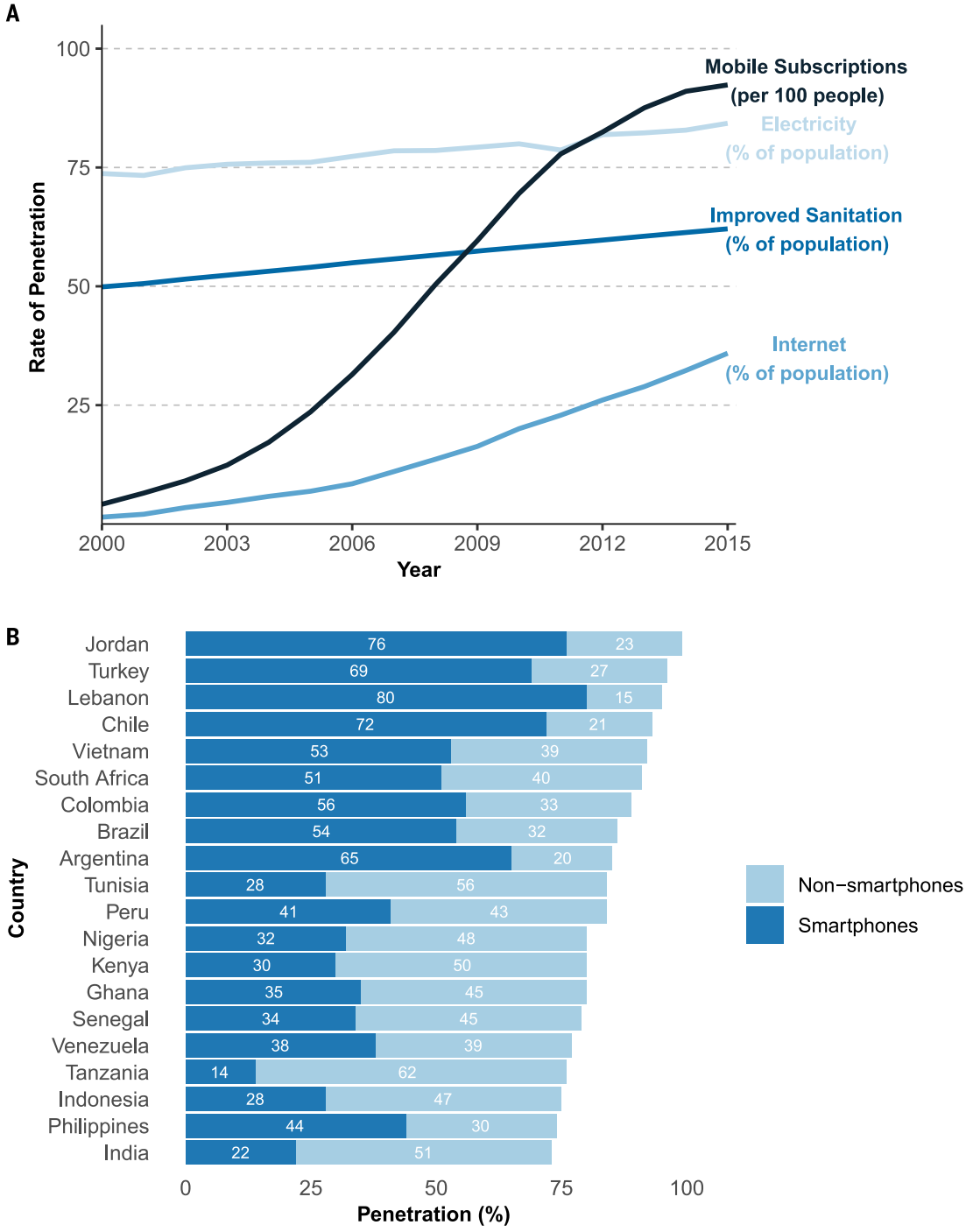



**Fig. 1. The spread of mobile phones in LMICs.** (**A**) Growth of mobile phone subscriptions relative to other services in LMICs. "Improved sanitation" denotes the percentage of people using basic sanitation services, at minimum [data from (*1*) and World Development Indicators: https://datacatalog.worldbank.org/dataset/world-development-indicators]. (**B**) Mobile phone penetration, as determined by the percentage of adults who report owning any mobile phone and those who own a smartphone [data from Pew Global Attitudes Survey, Spring 2017: www.pewresearch.org/global/dataset/spring-2017-survey-data/].

agricultural extension efforts. Many agricultural extension workers already have smartphones and thus could download information on pests, flooding, or other problems arising in their region, as well as information needed to respond to farmer queries. Automatic notifications can allow extension agents to alert farmers in their region when they are visiting demonstration plots or conducting training sessions. Mobile phones could also be used to improve accountability among extension workers—for example, by allowing extension workers and their supervisors to set goals and track performance, enabling automatic collection of feedback from farmers, or tracking whether extension agents actually visit farmers.

Finally, digital agricultural services can improve the functioning of agricultural supply chains. For example, these services could make it easier for farmers to check and compare input or output prices, potentially lowering markups; notify farmers whether inputs are in stock with particular dealers; and facilitate coordination among farmers in an area and

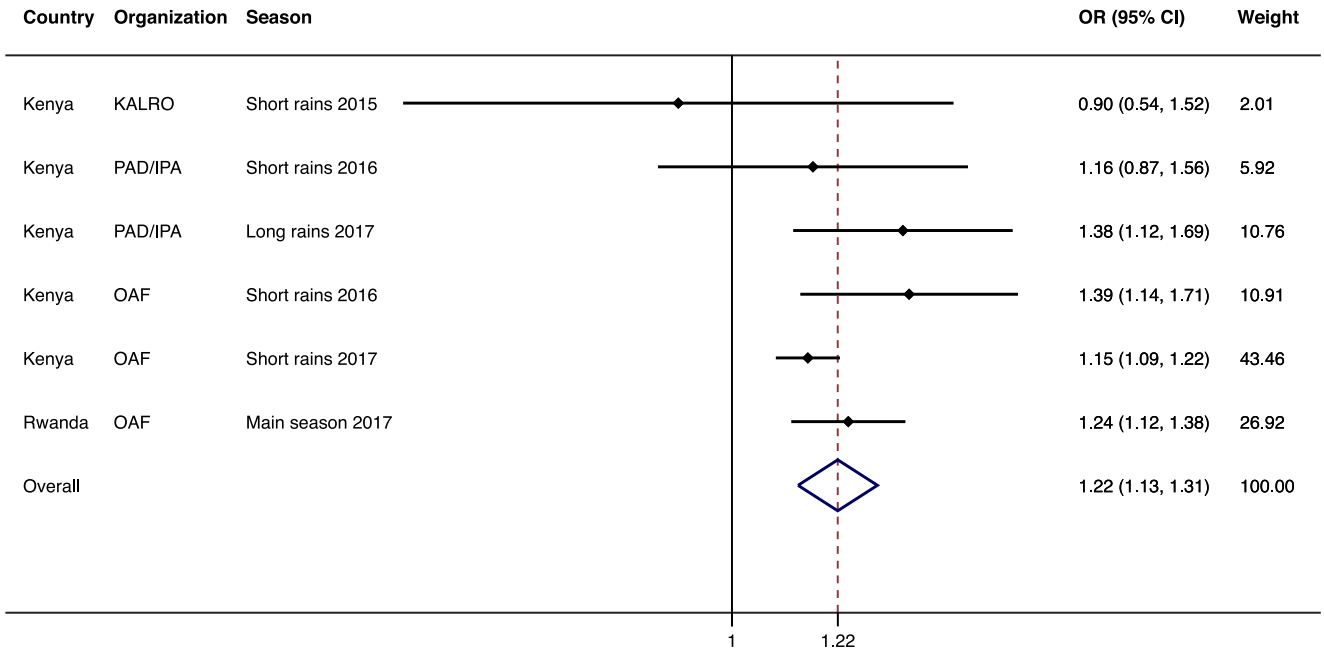



**Fig. 2. Effects of text messages on acquisition of recommended inputs.** Meta-analysis of the effects of text messages on following advice for purchasing agricultural lime, as measured by administrative data. Kenya Agriculture and Livestock Research Organization (KALRO), Precision Agriculture for Development and Innovations for Poverty Action (PAD/IPA), and One Acre Fund (OAF) implemented the programs. The effects are measured in odds ratios (OR). The OR is the ratio of the odds of following recommendations in the treatment group divided by the odds in the control group. The odds refer to the probability of following the advice over the probability of not following the advice. Weights are from a random effects analysis [data adapted from (*47*)].

with traders. Firms could use projections of harvest quality to inform lending decisions (*35*). Satellite-based yield assessment could be used to inform social insurance programs that provide support for farmers in response to weather or pest disasters.

However, despite the potential of digital agriculture advisory services, reasons for skepticism remain. Overcoming informational constraints may not result in substantially increased agricultural productivity, given the existence of other barriers such as credit constraints, input shortages at local markets, and missing insurance markets and infrastructure (*12*). Even to the extent that informational barriers are important, mobile phone messages may not overcome them: Some farmers ignore messages, especially from unknown sources, because phone spam is common in many LMICs. Some farmers are illiterate and have difficulty using voice menus. Senders may design obscure and confusing messages or may provide messages designed to target objectives at odds with farmer interests, such as messages aimed at increasing sales of inappropriate agricultural inputs. Certain kinds of information may be too complicated to convey by text or voice; effective communication may require pictures or video. Smartphones are thus required to receive these messages, but few smallholder farmers currently have access to this technology in the poorest countries. Finally, farmers may begin to ignore reminders or nudges if they are repeated too often, or they may be annoyed by unwanted spam messages or feel patronized by reminders and exhortations (*36*). Taken together, such issues could lead to reduced trust in the messaging system.

Realizing the potential of customization and two-way communication in LMICs carries particular challenges. Customization requires information about a farm's location, which is difficult to collect remotely unless farmers have GPS-enabled smartphones, because in many countries there is a lack of precisely-defined physical addresses (*37*), area names are often ambiguous, and user text entry is error prone (*38*). Gathering agricultural data from farmers is challenging because response rates to phone surveys are typically low; farmers may be hesitant to provide accurate information; and some information, such as exact yields, can be difficult to quantify.

## Impacts of digital agriculture: Empirical evidence

Earlier reviews of the impacts of digital agricultural extension report mixed results and considerable context dependence (*39*–*43*). However, sufficient evidence is now emerging to begin quantitatively assessing the farmer-level impact of digital agricultural extension by meta-analysis.

### *Impacts on individual farmers*

Several experimental studies have found that mobile phone–based programs increase farmer knowledge and self-reported adoption or planned adoption of recommended agricultural inputs and practices (*44*–*46*). Each of these outcome variables has limitations. Knowledge may or may not translate into behavior change. Relying on self-reported data on the use of inputs may lead to overestimation of impact. For example, farmers who receive messages advocating certain behavior may over-report this behavior because of social desirability bias. Indeed, a recent meta-analysis of four trials in Kenya found that the measured impact of mobile phone messages using self-reported data exceeded the impacts based on administrative data (*47*).

To alleviate such concerns, administrative data on input purchases from agricultural suppliers or redemption of discount coupons were used to measure farmer behavior in six experimental evaluations of text messaging programs that encouraged farmers in East Africa to use locally appropriate inputs (*47*). Figure 2 depicts the results from a meta-analysis of these studies, which found that the odds ratio for following the recommendation to purchase agricultural lime, an input to reduce soil acidity, is 1.22 [95% confidence interval (CI): 1.13 to 1.31]. For context, the proportion of people acquiring recommended inputs in each of the control groups ranged from 0.03 to 0.32.

Some of the individual experiments had statistically significant impacts and others did not. However, we cannot reject the hypothesis that the effects were the same across

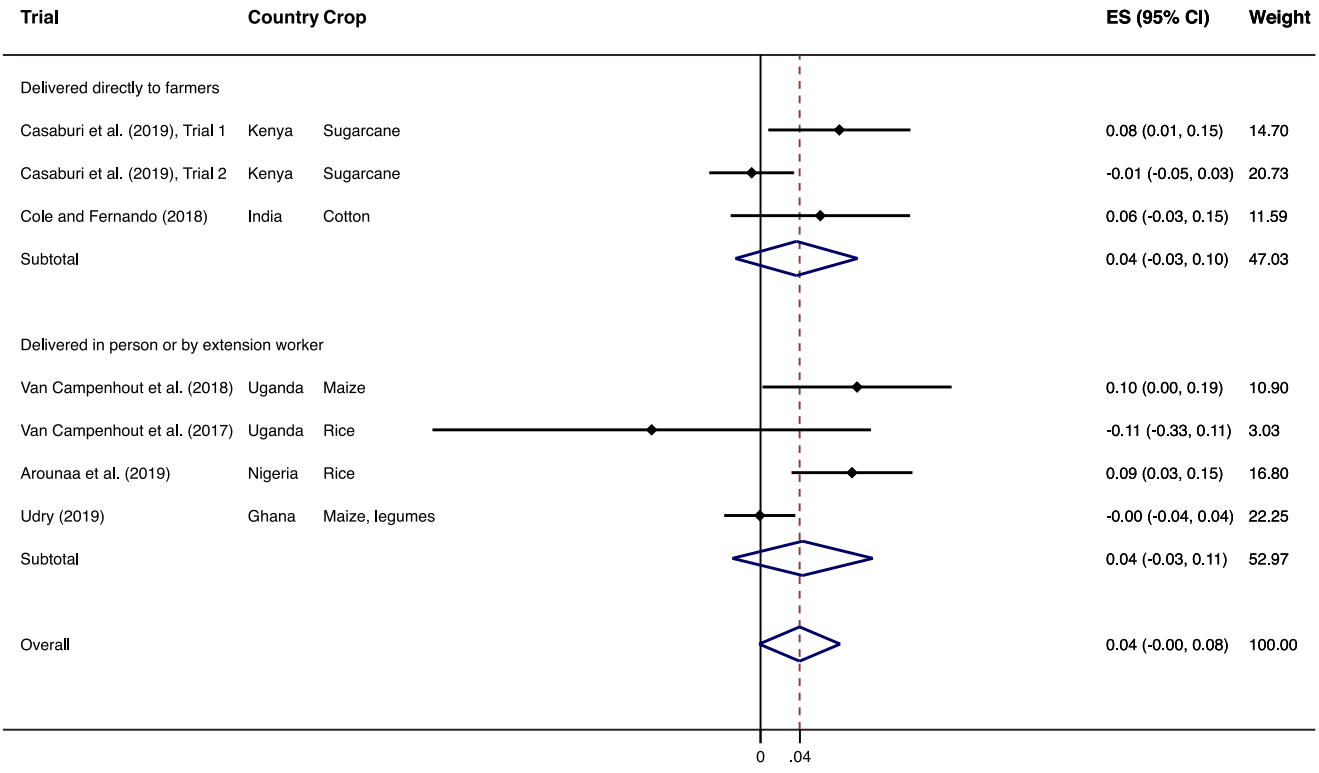



**Fig. 3. Effects of digital agriculture programs on yields.** Meta-analysis of effects of digital technologies on crop yields drawn from six studies that report yields [data from (*46*, *48*–*52*)]. For the study by Udry (*52*), the outcome variable is harvest value. The upper portion of the figure shows the impact of mobile-based programs delivered directly to farmers. The lower portion shows the impact of digital programs with an in-person component. Weights are from a random effects analysis. ES, effect size as a percentage increase.

contexts and that the estimated effects differed only because of sampling variation, which suggests that we need to be cautious in claims regarding heterogeneous treatment effects and, in particular, in interpreting the sources of differences in estimated effects across studies. Combining these estimates with agricultural trial data on the impact on yields of treating soil with lime suggests that farm profits increased by one to two orders of magnitude beyond the marginal cost of sending the messages. Similar estimates were found for fertilizer purchases (*47*).

Figure 3 reports on a complementary meta-analysis measuring the impact of experimentally evaluated digital agricultural extension interventions on farm yields or harvest value (unfortunately we do not have sufficient data on farm costs to estimate impacts on profits). This analysis encompasses four trials of messages delivered purely through mobile phones: two text message interventions with sugarcane farmers in Kenya (*48*) and two season measures for an interactive voice response (IVR) intervention with cotton farmers in India (*46*). It also includes four studies with an in-person component: two video interventions with maize and rice farmers in Uganda implemented via in-person visits (*49*, *50*), a program providing customized information on rice cultivation to Nigerian farmers offered through extension workers (*51*), and a program in Ghana delivered by community extension workers who relied on a mobile software application (*52*).

Several statistical approaches indicate that digitally delivered advice to farmers increases yields by ~4% (see supplementary materials). Notably, the impacts are not larger for services that include more costly in-person components. On average, the value of increased output greatly exceeds the marginal cost of delivery via mobile phones, such that policymakers would invest in mobile-based programs unless they are highly risk-averse.

Several factors suggest that the true returns to investment in digital agricultural extension may be higher than these numbers suggest. First, farmers who receive information via digital agricultural extension sometimes transmit it to other farmers, thus creating additional benefits (*46*, *47*). Second, to the extent that impacts vary across contexts and policy-makers have data to assess impact in their own context, there is value in testing such systems, assessing their effects, and adjusting policy accordingly. Unsuccessful programs can be abandoned and successful ones scaled up. Finally, impacts are likely to improve over time as farmers learn to use the systems, program operators improve message content and delivery through A/B trials, and smartphone use spreads, enabling digital extension services to incorporate more advanced technology, such as video, and better customization to local conditions. Video-based interventions and a gamified app have also been found to improve knowledge and farmers' practices (*30*, *50*, *53*, *54*).

As noted, traditional in-person agricultural extension has been found to favor certain groups. It seems likely that digital extension will also favor men, as well as richer, younger, and more educated farmers with better digital access. However, current data are inconclusive, and it is possible that biases will be less severe than with in-person extension. Cole and Fernando (*46*) report suggestive evidence that richer farmers were modestly more engaged in the service they studied and were more likely to adopt recommended practices. In contrast, Fabregas *et al.* (*47*) found little evidence of heterogeneous impact in their meta-analysis, although the underlying studies did not include farmers without phones.

### *Market- or system-level impacts*

Beyond individual-level effects, digital technologies can also affect farmers by altering entire markets or systems. In particular, improved access to price information can enable farmers to sell their products in markets with higher prices and reduce price dispersion across markets. By reducing waste of perishable goods and the need for middlemen,

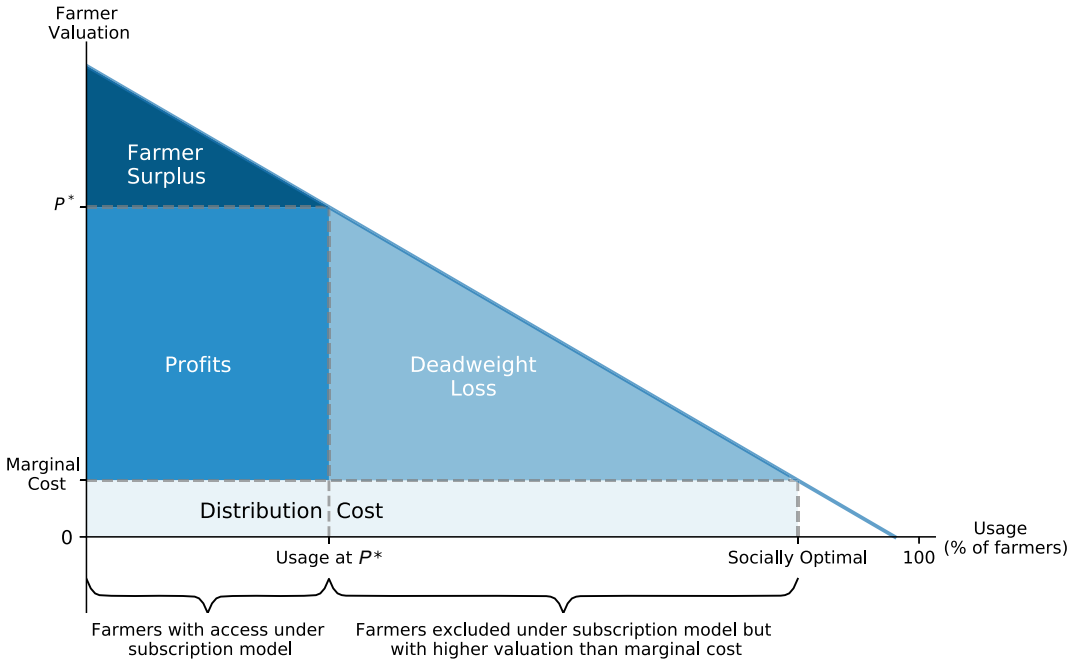



**Fig. 4. Farmer valuation curve and usage under a subscription model.** Given the downward-sloping valuation (or demand) curve, *P** is the profit-maximizing price. Usage at this price is lower than the socially efficient level, giving rise to a deadweight loss. Firms would not be able to cover fixed costs if they set prices at the distribution costs. Firms will invest in a system only if the development costs are less than the expected profits, whereas a hypothetical benevolent social planner would invest as long as costs are less than the sum of profits, farmer surplus, and deadweight loss.

improved information access can thus increase producer prices and lower average prices for consumers. Access to mobile phones allowed Indian fishermen to compare prices while still at sea and then transport their catch to the markets offering the highest prices, thus causing a reduction in price dispersion across markets (*55*). Fewer fish were wasted, profits increased by 8%, and consumer prices declined by 4%.

Studies in Uganda and Niger recorded similar results for other crops (*56*–*58*); delivering price information for staple grains was also found to cause positive effects in Ghana and Peru (*59*, *60*). In contrast, sending farmers price and weather information through text messages and phone calls did not affect average prices for crops from farmers in India or Colombia (*61*–*63*). These differences are hypothesized to result from a combination of factors, including differences in target populations, crop varieties, the importance of informational constraints, message design, and barriers to the effective use of information and communications technology (*41*).

Digital technologies may also improve supply chains by helping farmers shop for adequate inputs or report inefficiencies or fraud. Casaburi *et al.* (*64*) examined a contract-farm setting in which farmers sign contracts with a sugar company in Kenya. The company provides agrochemical inputs to farmers and then deducts the costs of the inputs from the amount it pays farmers for the sugarcane. Delays or failures in fertilizer deliveries to farmers are common, but by establishing a hotline for farmers to report problems, late delivery was reduced by 23% and nondelivery by 54%. The benefits spill over to neighbors because the company schedules deliveries to neighboring farms at the same time.

An important area for future work is exploring whether digital agriculture can address supply chain problems, such as limited competition and high markups (*65*) and adulteration and counterfeiting of inputs (*66*). Moreover, studying the distributional effects of different interventions, arising from a combination of the direct effects of receiving digital advice itself and through positive or negative spillovers of interventions, remains a fruitful area for investigation.

### *Learning from farmers*

A key open empirical question is the extent to which mobile systems can gather valuable information from farmers, which in turn can be used to inform other farmers. In the United States, Farmers Business Network applies machine learning to hundreds of thousands of acre-years of data to provide high-quality yield predictions for seed varieties (*67*). Mobile phone systems in LMICs could potentially be used to collect data to serve as inputs in machine learning applications, learn from farmers' experiences with particular agricultural technologies, and facilitate networking among farmers.

However, gathering high-quality data from farmers is challenging. At the most basic level, phone surveys allow collection of high-frequency survey data on agricultural production at a much lower cost than traditional methods (*68*). Yet despite some successes in eliciting information through phone-based feedback tools (*69*), phone surveys are plagued with low response rates and thus may be subject to selection bias. If users are required to provide information to access content, they may prioritize speed over accuracy, degrading the quality of information.

Systems that foster information exchange to facilitate participation and truthfulness by using a "Netflix model" are one solution. Netflix shares recommendations for video content with its users and tethers those recommendations to what users have liked in the past. This procedure incentivizes users to share information with the platform to improve the quality of its future recommendations. This information is then used to benefit other users of the service by improving the quality of Netflix's recommendations to them. A comparable model could potentially work for agriculture. Farmers could be convinced to supply information on what has recently worked for them, because doing so would improve the advice the mobile-advisory service provides them in the future. Such a system would incentivize farmers to share their experiences, because sharing would enable them to receive better-tailored recommendations. The resulting data could also be used to improve recommendations for other, similar farmers.

## Financing and governance of digital agriculture systems

Digital agricultural extension systems currently reach only a small proportion of farmers (*70*). Here, we discuss barriers to commercial scaling, as well as problems with public scaling and emerging evidence on ways to address them.

### *Barriers to scaling of subscription models*

Many of the efforts to establish digital extension systems, such as iCow Global in Kenya or RML Agtech in India, have sought to finance themselves by selling subscriptions to farmers, but these types of efforts have reached only a small fraction of the potential market (*70*). Economic theory suggests that three features of markets for agricultural information—nonrivalry, nonexcludability, and asymmetric information—make it difficult for pure subscription models to reach as many farmers as would be efficient from a social point of view.

#### *Nonrivalry*

Information differs from most other goods (*71*, *72*). Creation of information involves fixed costs—for example, collecting data from soil tests and weather stations or designing, testing, and refining comprehensible and actionable messages for farmers. However, unlike most other goods, such as agricultural products, information is a nonrival commodity: Once it has been created, it can be used by additional people at a minimal marginal distribution cost, with no cost to others.

From a social point of view, it is efficient for all who value the information at more than the distribution cost to have access to it. However, a firm using a pure subscription model would need to charge a higher fee than the price of distribution to cover the fixed cost of information creation. As a result, some farmers for whom it would be efficient to obtain information would be excluded (Fig. 4). For the particular curve drawn, the majority of farmers would value the information at more than the cost of distribution, but only about one-third would be willing to pay the profit-maximizing price of a commercial firm using a pure subscription model.

In addition, under nonrivalry, a potential provider operating via a subscription model may not have the commercial incentives to create the service, even if it is socially efficient to do so. It is socially efficient to invest if the total area between the farmer valuation curve and the cost of distributing information—profits plus farmer surplus and deadweight loss—exceeds the cost of information creation. However, a private firm will invest only if the profits from selling information are sufficient to cover the cost of information creation. Nonrivalry leads to a gap between the conditions under which it is socially efficient to invest in the creation of the service and the conditions under which it will be privately profitable to do so.

More sophisticated forms of subscription models may ameliorate these distortions. To the extent that firms can charge different prices for information on the basis of farmers' willingness to pay, these firms can serve more farmers and increase their profits. "Freemium" models are a step in this direction because they give consumers a chance to learn about the quality of advice before they pay for it.

Nonrivalry does not imply that no knowledge-creation investments will be commercially viable. Indeed, a considerable share of all research and development investment is made by the private sector. If the only market failure associated with agricultural information markets was nonrivalry, then a subscription model might become viable once technological progress and the spread of smartphones and data plans sufficiently drove down the costs of information production and distribution. However, markets for agricultural information are subject to two additional distortions that further reduce farmers' willingness to pay for information.

#### *Nonexcludability*

Agricultural information is also nonexcludable or only partially excludable—i.e., once an individual has access to the information, this person can easily share it with others. In their study of digital agricultural extension in India, Cole and Fernando (*46*) found significant knowledge spillovers to farmers who had not received the services in the trial. A rich literature documents the flow of agricultural information in rural communities (*73*–*75*). Sharing of information not only directly reduces the number of potential customers for digital agricultural extension services but may also reduce willingness to pay among those who do purchase, which could affect the financial viability of subscription-based services.

In a study of willingness to pay for local soil information in western Kenya, individual farmers were not willing to pay the full cost of local soil test results (*76*). However, the aggregate valuation of all farmers for a given test in an area exceeded the cost of testing, potentially making investment in this information worthwhile from a social standpoint. Farmers' willingness to pay for information was also lower in a setting where they could ask others for the information, which suggests that the option of resale or free riding depressed any willingness to pay.

#### *Asymmetric information*

Buyers do not necessarily know the value of the information sold to them, and they may not trust sellers' claims about its value (*77*). Because agricultural production is highly variable and the profitability of recommended agricultural practices may differ from year to year, it may be difficult for farmers to assess the quality of advice, even after they purchase it.

Weak regulation makes it difficult to trust those selling information. Fraudulent operators can set up firms and offer useless information. Even firms with legitimate information would have incentives to inflate the benefits of their information. Farmers may thus discount any claims and reduce their willingness to pay for information. Markets for information can unravel entirely, preventing any transactions. Sellers address this issue by investing in a reputation for trustworthiness, but this involves some costs, and farmers may still retain doubts.

Beyond these distortions specific to agricultural information markets, other factors make selling any investment product in these markets difficult. Many developing-country farmers have no readily available cash and may not be able to borrow money either. In addition, a variety of behavioral factors, ranging from present bias to loss aversion, may inhibit investment (*27*). Customer acquisition costs and transaction costs in payments are also likely obstacles to successful scaling with subscription models.

Together, these factors erode farmers' willingness to pay for information and thus limit the financial viability of digital agricultural extension efforts. Cole and Fernando (*46*) estimate that their IVR service in India increased farmers' incomes by more than the costs of the service, but despite a high rate of farmer engagement, the average price a farmer was willing to pay for a 9-month subscription was only $2, whereas the cost of provision was $7.

### *Other commercial financing models*

Beyond pure subscription models, other commercial models may partially address some of the market failures.

#### *Contract farming*

Some agricultural products, such as sugarcane or dairy, require local processing, often featuring a dominant local buyer. If the buyer profits sufficiently from having a greater input supply, this person may be willing to pay to provide digital agricultural extension services for all farmers in the area. Because the dominant buyer would want all farmers in the area to increase production, this would help address the problems of nonrivalry and nonexcludability. Additionally, a buyer with professional staff may be less subject to asymmetric information problems. The buyer may personally operate a digital agricultural extension service or purchase these services in bulk from another provider, thus reducing customer acquisition costs relative to the cost of selling to individual farmers.

This approach may be particularly effective when the buyer cares not only about the extra profits from greater input supply, but also about farmers. Dairy farmers, for example, often organize themselves into cooperatives that jointly buy milk-processing equipment. In some cases, lenders also provide digital agricultural advice.

### *Advertising and selling inputs*

Digital agricultural extension providers could also try to finance themselves by selling advertising, selling own-brand inputs, or entering strategic alliances with agricultural input providers. Incentives such as commissions can lead to biased advice (*78*); nevertheless, some firms succeed in developing a reputation for providing objective advice. For example, Farmers Business Network in the United States has grown rapidly with a financial model based on selling own-brand agricultural inputs, as well as providing information.

As smartphones and data plans become more common and it becomes possible to transmit more types of information at low cost, advertising and input sales may generate enough revenue to finance some provision of agricultural information, but markets are still unlikely to deliver socially optimal outcomes. Information will probably be undersupplied on agricultural techniques that do not involve input purchases or that involve only purchases of inputs that have become commodified and hence have low markups. False or misleading information may be supplied on the merits of different brands or the suitability of techniques that involve input purchase. Although regulation could potentially address such issues, designing and implementing appropriate regulation is likely to be difficult.

### *Public financing and government provision*

The above examples of market failures provide a rationale for the public sector to fully or partially finance provision of digital agricultural extension. A public-sector entity could either operate a digital agricultural extension service itself, as the government of Ethiopia currently does, or contract with or provide financial support for private providers, as donors in international development are currently doing. Governments could also use their position as regulators to encourage telecommunication firms to provide such services or to make capacity available to other digital agricultural extension providers.

Scaling through governments entails typically lower customer acquisition costs than faced by individual companies, because governments can leverage their relationships with telecommunications companies, the existing agricultural extension apparatus, and regulatory powers to draw farmers to their platforms. However, just as market solutions are subject to market failures, government solutions are subject to government failures. Governments are not known for nimble product development or user-friendly technology interfaces, and they lack the immediate customer feedback mechanism the market provides. Government agencies often provide agricultural information that might be too technical or detailed for communicating with the average farmer. For example, several Indian state governments distribute personalized soil health cards, based on local soil tests, to farmers. These cards are difficult for farmers to understand, and many farmers report never receiving them (*15*). Those who do receive the report cards often distrust the content (*79*). Simplifying the design, making the cards less technical, and complementing them with information delivered by mobile phones increased baseline comprehension from 8% to at least 40% (*15*). Similarly, a government-sponsored IVR helpline in Africa required farmers to answer a series of registration questions before they could access content, preventing many from reaching the agricultural content. Merely postponing user registration until after the farmers received some useful information increased the share of farmers getting to the content by approximately one-fifth, from 52 to 63% (*80*).

In these cases, governments responded to evidence by adjusting their programs, redesigning the soil health cards, and postponing registration requirements, raising the possibility that the systematic incorporation of ongoing surveys and A/B trials into government programs may serve as a partial substitute for the lack of market feedback.

However, governments, like private businesses, may distort information provision to farmers. For example, governments may want to increase production of certain commodities to achieve export goals, but they may not sufficiently value the time and effort required for farmers to adopt the corresponding agricultural practices. If government systems grow, input sellers may start lobbying or bribing government officials to recommend their brands as opposed to others.

### *Regulation*

Digital agricultural extension raises regulatory questions that require further research. Information providers with financial interests in selling certain products might send misleading messages. One approach to tackling this challenge would be to mandate certain disclosures of financial conflicts. However, evidence from the regulation of financial and medical products industries suggests that such mandates are not sufficient (*81*). Telecommunications authorities typically have rules limiting phone spam and will have to decide whether specific emergencies, such as severe pest outbreaks, warrant waiving rules against sending unsolicited information. As smartphones and data plans spread, it will become cheaper to distribute content to farmers, so nonexcludability of information will be less of a barrier to information provision. But it is likely to become more difficult to control the provision of misleading information, and hence asymmetric information may become more of a problem.

Customization could have great potential benefits, but it also raises questions about how to protect users' privacy. Governments must decide whether and under what circumstances to share contact information for agricultural extension agents or farmers.

Regulatory issues arise even for messages sent by government agencies. Messages from an agricultural ministry, for example, could crowd out equally important health messages. Messages that misleadingly imply that socially desirable behavior (such as environmentally favorable agricultural practices) has individual benefits could reduce trust in all messages from the government. Too many messages could annoy people and make them feel uncomfortable (*82*).

## Outlook

The available evidence suggests that the benefits of providing digital extension far exceed the costs but that subscription-based models will not reach optimal scale. This disparity creates a potential role for public financing, which LMIC governments and aid donors are increasingly supporting.

Delivering on the full promise of digital agriculture, including customization of information provision, will require sustained cycles of iteration and testing. The development of lessons that are viable and useful in multiple contexts will be essential to avoid reinventing the wheel for each application. Because these lessons may constitute a global public good, multilateral institutions and global donors may wish to financially support digital information provision efforts by governments or private actors in exchange for undertaking experimentation and making the results widely available. Equally, many of the emerging lessons on provision of information to farmers could also apply in other sectors, such as education or health.

## REFERENCES AND NOTES


1. World Bank, "World Development Report 2016: Digital Dividends" (World Bank, 2016). doi: 10.1596/978-1-4648-0671-1
2. J. C. Aker, J. E. Blumenstock, "The economic impacts of new technologies in Africa" in *The Oxford Handbook of Africa and Economics: Volume 2: Policies and Practices*, C. Monga, J. Y. Lin, Eds. (Oxford Univ. Press, 2015), pp. 354–371.
3. J. C. Aker, I. M. Mbiti, Mobile Phones and Economic Development in Africa. *J. Econ. Perspect.* **24**, 207–232 (2010). doi: 10.1257/jep.24.3.207
4. J. C. Aker, "Using digital technology for public service provision in developing countries" in *Digital Revolutions in Public Finance*, S. Gupta, M. Keen, A. Shah, G. Verdier, Eds. (International Monetary Fund, 2017), pp. 201–224.

5. T. Suri, W. Jack, The long-run poverty and gender impacts of mobile money. *Science* **354**, 1288–1292 (2016). doi: 10.1126/science.aah5309; pmid: 27940873
6. J. H. Pierskalla, F. M. Hollenbach, Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa. *Am. Polit. Sci. Rev.* **107**, 207–224 (2013). doi: 10.1017/S0003055413000075
7. B. Qin, D. Strömberg, Y. Wu, Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda. *J. Econ. Perspect.* **31**, 117–140 (2017). doi: 10.1257/jep.31.1.117
8. J. James, Digital Divide Complacency: Misconceptions and Dangers. *Inf. Soc.* **24**, 54–61 (2008). doi: 10.1080/01972240701774790
9. J. Blumenstock, N. Eagle, "Mobile divides: Gender, socioeconomic status, and mobile phone use in Rwanda" in *Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development - ICTD '10* (Association for Computing Machinery, 2010), article no. 6; .doi: 10.1145/2369220.2369225
10. Food and Agriculture Organization of the United Nations (FAO), "The economic lives of smallholder farmers: An analysis based on household data from nine countries" (FAO, 2015); www.fao.org/3/a-i5251e.pdf.
11. Z. Cui *et al*., Pursuing sustainable productivity with millions of smallholder farmers. *Nature* **555**, 363–366 (2018). doi: 10.1038/nature25785; pmid: 29513654
12. B. K. Jack, "Market inefficiencies and the adoption of agricultural technologies in developing countries," [White paper, Agricultural Technology Adoption Initiative, JPAL (MIT)/CEGA (Berkeley), 2011]; https://escholarship.org/content/qt6m25r19c/qt6m25r19c.pdf.
13. J. Anderson, G. Feder, in *Agricultural Development: Farmers, Farm Production and Farm Markets*, vol. 3 of *Handbook of Agricultural Economics*, R. Evenson, P. Pingali, Eds. (Elsevier, 2007), pp. 2343–2378.
14. K. Davis, B. Swanson, D. Amudavi, D. A. Mekonnen, A. Flohrs, J. Riese, C. Lamb, E. Zerfu, "In-depth assessment of the public agricultural extension system of Ethiopia and recommendations for improvement" [International Food Policy Research Institute (IFPRI) Discussion Paper, IFPRI, 2010]; www.ifpri.org/publication/depth-assessment-public-agricultural-extension-system-ethiopia-and-recommendations.
15. S. Cole, G. Sharma, "The Promise and Challenges in Implementing ICT for Agriculture," Working paper (2017); https://precisionag.org/uploads/cole-sharma-july1-2017.pdf.
16. M. Gautam, "Agricultural extension: The Kenya experience: An impact evaluation" (World Bank Publications, 2000).
17. O. Bandiera, R. Burgess, E. Deserranno, R. Morel, I. Rasul, M. Sulaiman, "Social Ties and the Delivery of Development Programs," Working paper (2018); https://drive.google.com/file/d/1uESM-AfGaiY-7eZ7Zq1WBzyP1IKE4fuS/view.
18. J. R. Alwang, P. B. Siegel, "Rural poverty in Zambia: An analysis of causes and policy recommendations" (World Bank, 1994).
19. B. Cunguara, K. Moder, Is Agricultural Extension Helping the Poor? Evidence from Rural Mozambique. *J. Afr. Econ.* **20**, 562–595 (2011). doi: 10.1093/jae/ejr015
20. F. Kondylis, V. Mueller, G. Sheriff, S. Zhu, Do Female Instructors Reduce Gender Bias in Diffusion of Sustainable Land Management Techniques? Experimental Evidence From Mozambique. *World Dev.* **78**, 436–449 (2016). doi: 10.1016/j.worlddev.2015.10.036
21. K. A. Saito, H. Mekonnen, D. Spurling, "Raising the Productivity of Women Farmers in Sub-Saharan Africa" (Discussion paper, World Bank, 1994); .doi: 10.1596/0-8213-2749-6
22. R. A. Viscarra Rossel *et al*., A global spectral library to characterize the world's soil. *Earth Sci. Rev.* **155**, 198–230 (2016). doi: 10.1016/j.earscirev.2016.01.012
23. D. Arrouays, P. Lagacherie, A. E. Hartemink, Digital soil mapping across the globe. *Geoderma Reg.* **9**, 1–4 (2017). doi: 10.1016/j.geodrs.2017.03.002
24. P. A. Sanchez *et al*., Digital soil map of the world. *Science* **325**, 680–681 (2009). doi: 10.1126/science.1175084; pmid: 19661405
25. M. Rosenzweig, C. R. Udry, "Forecasting Profitability," Working paper 19334 [National Bureau of Economic Research (NBER) Working Paper Series, 2013]; .doi: 10.3386/w19334
26. R. B. Alley, K. A. Emanuel, F. Zhang, Advances in weather prediction. *Science* **363**, 342–344 (2019). doi: 10.1126/science.aav7274; pmid: 30679358
27. M. Kremer, G. Rao, F. Schilbach, "Behavioral development economics" in *Handbook of Behavioral Economics - Foundations and Applications 2*, B. D. Bernheim, S. DellaVigna, D. Laibson, Eds. (North Holland, 2019), pp. 345–458.
28. M. Burke, D. B. Lobell, Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2189–2194 (2017). doi: 10.1073/pnas.1616919114; pmid: 28202728
29. D. Donaldson, A. Storeygard, The View from Above: Applications of Satellite Data in Economics. *J. Econ. Perspect.* **30**, 171–198 (2016). doi: 10.1257/jep.30.4.171
30. K. Vasilaky, K. Toyama, T. Baul, M. Mangal, U. Bhattacharya, "Learning Digitally: Evaluating the Impact of Farmer Training via Mediated Videos" presented at the Northeast Universities Development Consortium Conference, Providence, RI, 7 and 8 November 2015; https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=NEUDC2015&paper_id=519.
31. M. Ridley, G. Rao, P. Vikram, F. Schilbach, "Poverty and Mental Illness: Causal Evidence" (2019); https://economics.mit.edu/files/18694.
32. S. Alan, T. Boneva, S. Ertac, Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit. *Q. J. Econ.* **134**, 1121–1162 (2019). doi: 10.1093/qje/qjz006
33. E. La Ferrara, Aspirations, Social Norms, and Development. *J. Eur. Econ. Assoc.* 10.1093/jeea/jvz057 (2019). doi: 10.1093/jeea/jvz057
34. T. Bernard, S. Dercon, K. Orkin, A. S. Taffesse, "The future in mind: short and long-run impact of an aspirations intervention in rural Ethiopia," Working paper (Univ. of Oxford, 2018); http://ibread.org/bread/system/files/bread_wpapers/429.pdf.
35. G. Murthy, M. Fernandez-Vidal, X. Faz, R. Barreto, "Fintechs and Financial Inclusion: Looking past the hype and exploring their potential" (Consultative Group to Assist the Poor, 2019); www.cgap.org/sites/default/files/publications/2019_05_Focus_Note_Fintech_and_Financial_Inclusion_1_0.pdf.
36. M. T. Damgaard, C. Gravert, The hidden costs of nudging: Experimental evidence from reminders in fundraising. *J. Public Econ.* **157**, 15–26 (2018). doi: 10.1016/j.jpubeco.2017.11.005
37. U. P. Union, "Addressing the World–An Address for Everyone" (White paper, Universal Postal Union, 2012); www.upu.int/fileadmin/documentsFiles/activities/addressingAssistance/whitePaperAddressingTheWorldEn.pdf.
38. A. Erlich, D. F. Jung, J. D. Long, C. McIntosh, The double-edged sword of mobilizing citizens via mobile phone in developing countries. *Dev. Eng.* **3**, 34–46 (2018). doi: 10.1016/j.deveng.2017.11.001
39. J. C. Aker, Dial "A" for agriculture: A review of information and communication technologies for agricultural extension in developing countries. *Agric. Econ.* **42**, 631–647 (2011). doi: 10.1111/j.1574-0862.2011.00545.x
40. E. Nakasone, M. Torero, B. Minten, The power of information: The ICT revolution in agricultural development. *Annu. Rev. Resour. Econ.* **6**, 533–550 (2014). doi: 10.1146/annurev-resource-100913-012714
41. J. C. Aker, I. Ghosh, J. Burrell, The promise (and pitfalls) of ICT for agriculture initiatives. *Agric. Econ.* **47**, 35–48 (2016). doi: 10.1111/agec.12301
42. H. Baumüller, The Little We Know: An Exploratory Literature Review on the Utility of Mobile Phone-Enabled Services for Smallholder Farmers. *J. Int. Dev.* **30**, 134–154 (2018). doi: 10.1002/jid.3314
43. U. Deichmann, A. Goyal, D. Mishra, Will digital technologies transform agriculture in developing countries? *Agric. Econ.* **47**, 21–33 (2016). doi: 10.1111/agec.12300
44. C. Larochelle, J. Alwang, E. Travis, V. H. Barrera, J. M. D. Andrade, Did You Really Get the Message? Using Text Reminders to Stimulate Adoption of Agricultural Technologies. *J. Dev. Stud.* **55**, 548–564 (2019). doi: 10.1080/00220388.2017.1393522
45. X. Fu, S. Akter, The Impact of Mobile Phone Technology on Agricultural Extension Services Delivery: Evidence from India. *J. Dev. Stud.* **52**, 1561–1576 (2016). doi: 10.1080/00220388.2016.1146700
46. S. A. Cole, A. N. Fernando, "'Mobile'izing Agricultural Advice: Technology Adoption, Diffusion and Sustainability," Working paper (2018); www.dropbox.com/s/0rewpp8oeost0o5/ao_final_v24_tables.pdf?dl=0.
47. R. Fabregas, M. Kremer, M. Lowes, R. On, G. Zane, "SMS-extension and Farmer Behavior: Lessons from Six RCTs in East Africa," Working paper (2019); www.dropbox.com/s/guiog3ief0hnwl7/textfarmers.pdf?dl=0.
48. L. Casaburi, M. Kremer, S. Mullainathan, R. Ramrattan, "Harnessing ICT to increase agricultural production: Evidence from Kenya," Working paper (2019); www.econ.uzh.ch/dam/jcr:873845ce-de4d-4366-ba9a-d60accda577d/SMS_paper_with_tables_20190923_merged.pdf.
49. B. Van Campenhout, W. Walukano, F. Nattembo, L. Nazziwa-Nviiri, J. Blom, "The Role of Information in Agricultural Technology Adoption: Experimental Evidence from Rice Farmers in Uganda" (IFPRI Discussion Paper, IFPRI, 2017); https://core.ac.uk/download/pdf/153429888.pdf.
50. B. Van Campenhout, D. J. Spielman, E. Lecoutere, "Information and Communication Technologies (ICTs) to Provide Agricultural Advice to Smallholder Farmers: Experimental Evidence from Uganda" (IFPRI Discussion Paper, IFPRI, 2018); www.ifpri.org/publication/information-and-communication-technologies-icts-provide-agricultural-advice-smallholder.
51. A. Arounaa, J. D. Michlerb, W. G. Yergoa, K. Saitoa, "One Size Does Not Fit All: Experimental Evidence on the Digital Delivery of Personalized Extension Advice in Nigeria" (2019); www.researchgate.net/publication/336927609_One_Size_Does_Not_Fit_All_Experimental_Evidence_on_the_Digital_Delivery_of_Personalized_Extension_Advice_in_Nigeria.
52. C. Udry, "Information, Market Access and Risk: Addressing Constraints to Agricultural Transformation in Northern Ghana" (2019); https://sites.northwestern.edu/christopherudry/files/2019/03/DIRTS-Final-report-1362c0g.pdf.
53. R. Gandhi, R. Veeraraghavan, K. Toyama, V. Ramprasad, "Digital Green: Participatory video for agricultural extension," in *2007 International Conference on Information and Communication Technologies and Development* (IEEE, 2007). doi: 10.1109/ictd.2007.4937388
54. E. Tjernström, T. J. Lybbert, R. Frattarola Hernández, J. S. Correa, "Learning by (virtually) doing: experimentation and belief updating in smallholder agriculture" (2019); www.lafollette.wisc.edu/sites/tjernstrom/Learning_by_virtually_doing.pdf.
55. R. Jensen, The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. *Q. J. Econ.* **122**, 879–924 (2007). doi: 10.1162/qjec.122.3.879
56. J. C. Aker, Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger. *Am. Econ. J. Appl. Econ.* **2**, 46–59 (2010). doi: 10.1257/app.2.3.46
57. J. C. Aker, M. Fafchamps, Mobile Phone Coverage and Producer Markets: Evidence from West Africa. *World Bank Econ. Rev.* **29**, 262–292 (2015). doi: 10.1093/wber/lhu006
58. J. Svensson, D. Yanagizawa, Getting Prices Right: The Impact of the Market Information Service in Uganda. *J. Eur. Econ. Assoc.* **7**, 435–445 (2009). doi: 10.1162/JEEA.2009.7.2-3.435
59. P. Courtois, J. Subervie, Farmer Bargaining Power and Market Information Services. *Am. J. Agric. Econ.* **97**, 953–977 (2015). doi: 10.1093/ajae/aau051
60. E. Nakasone, "The Role of Price Information in Agricultural Markets: Experimental Evidence from Rural Peru," Working paper (2014); https://fc317c2d-a-62cb3a1a-s-sites.googlegroups.com/site/eduardonakasone/Nakasone%20-%20Price%20Info%20and%20Ag%20Markets%20-%20OCT2014.pdf?attachauth=ANoY7crbtOVyu8ZK0sal6LjZFX5YTjR1uGo1cHVxIjGqIlwCGBxMJ_IlNbLwB9TYQn6vfZSW7ol2aX2JLozIxPvAWqbjvGZDtpdAtqLTQTQyWHFL4OMZ2294iGS4VN2afdYzi-rbC5gUKq8ADwq0aKtPVCX0iAMynd5Q3QK2oljTngQPoyLOYuNdJrQtyFxJV3pnkb9kLrEE8X6_oO2JBZJiF2snZppL-Edb5LwSVErpxNJJa4DARTJXcQPIsm7suHAaddvPwWGSiX75x7mE-gG9osAIVdVGg%3D%3D&attredirects=0&d=1.
61. M. Fafchamps, B. Minten, Impact of SMS-Based Agricultural Information on Indian Farmers. *World Bank Econ. Rev.* **26**, 383–414 (2012). doi: 10.1093/wber/lhr056
62. A. Camacho, E. Conover, "The Impact of Receiving Price and Climate Information in the Agricultural Sector," Working paper no. IDB-WP-220 [Inter-American Development Bank (IDB) Working Paper Series, 2011]; doi: 10.2139/ssrn.1755057
63. S. Mitra, D. Mookherjee, M. Torero, S. Visaria, Asymmetric Information and Middleman Margins: An Experiment with Indian Potato Farmers. *Rev. Econ. Stat.* **100**, 1–13 (2018). doi: 10.1162/REST_a_00699
64. L. Casaburi, M. Kremer, R. Ramrattan, "Crony Capitalism, Collective Action, and ICT: Evidence from Kenyan Contract Farming," Working paper (2019); http://econ.uzh.ch/dam/jcr:e2ffc4e5-ab32-4405-bfa4-70b0e962aa81/hotline_paper_20191015_MERGED.pdf.
65. R. J. Sexton, Imperfect Competition in Agricultural Markets and the Role of Cooperatives: A Spatial Analysis. *Am. J. Agric. Econ.* **72**, 709–720 (1990). doi: 10.2307/1243041
66. T. Bold, K. C. Kaizzi, J. Svensson, D. Yanagizawa-Drott, Lemon Technologies and Adoption: Measurement, Theory and Evidence from Agricultural Markets in Uganda. *Q. J. Econ.* **132**, 1055–1100 (2017). doi: 10.1093/qje/qjx009

67. S. Cole, T. L. He, "Farmers Business Network: Putting Farmers First," Harvard Business School Case 217-025 (2018); www.hbs.edu/faculty/Pages/item.aspx?num=51649.
68. B. Dillon, Using mobile phones to collect panel data in developing countries. *J. Int. Dev.* **24**, 518–527 (2012). doi: 10.1002/jid.1771
69. M. Jones, F. Kondylis, Does feedback matter? Evidence from agricultural services. *J. Dev. Econ.* **131**, 28–41 (2018). doi: 10.1016/j.jdeveco.2017.10.013
70. M. Tsan, S. Totapally, M. Hailu, B. K. Addom, "The digitalization of African agriculture report 2018-2019: Executive summary" (CTA/Dalberg Advisers, 2019); https://cgspace.cgiar.org/bitstream/handle/10568/103198/Executive%20Summary%20V4.5%20ONLINE.pdf.
71. K. J. Arrow, "Economic welfare and the allocation of resources for invention" in *The Rate and Direction of Inventive Activity: Economic and Social Factors* (Princeton Univ. Press, 1962), pp. 609–626.
72. P. M. Romer, Increasing Returns and Long-Run Growth. *J. Polit. Econ.* **94**, 1002–1037 (1986). doi: 10.1086/261420
73. A. D. Foster, M. R. Rosenzweig, Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *J. Polit. Econ.* **103**, 1176–1209 (1995). doi: 10.1086/601447
74. T. G. Conley, C. R. Udry, Learning about a new technology: Pineapple in Ghana. *Am. Econ. Rev.* **100**, 35–69 (2010). doi: 10.1257/aer.100.1.35
75. L. Beaman, A. BenYishay, J. Magruder, A. M. Mobarak, "Can Network Theory-based Targeting Increase Technology Adoption?" Working paper 24912 (NBER Working Paper Series, 2018); .doi: 10.3386/w24912
76. R. Fabregas, M. Kremer, J. Robinson, F. Schilbach, "The Value of Local Agricultural Information: Evidence from Kenya," Working paper (2019); www.dropbox.com/s/bumw8uj5dyhnt9n/valueinformation.pdf?dl=0.
77. G. A. Akerlof, The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *Q. J. Econ.* **84**, 488 (1970). doi: 10.2307/1879431
78. S. Anagol, S. Cole, S. Sarkar, Understanding the Advice of Commissions-Motivated Agents: Evidence from the Indian Life Insurance Market. *Rev. Econ. Stat.* **99**, 1–15 (2017). doi: 10.1162/REST_a_00625
79. R. Fishman, A. Kishore, Y. Rothler, P. S. Ward, S. Jha, R. K. P. Singh, "Can information help reduce imbalanced application of fertilizers in India?: Experimental evidence from Bihar" (IFPRI Discussion Paper, IFPRI, 2016); http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/130249/filename/130460.pdf.
80. Precision Agriculture for Development, "Mid-term evaluation of IVR/SMS project: Findings and recommendations" (Ethiopian Agricultural Transformation Agency, 2018); https://precisionag.org/uploads/January-ATA-PAD-Report-FINAL-FOR-VALIDATION-WORKSHOP-compressed.pdf.
81. G. Loewenstein, S. Sah, D. M. Cain, The unintended consequences of conflict of interest disclosure. *JAMA* **307**, 669–670 (2012). doi: 10.1001/jama.2012.154; pmid: 22337676
82. H. Allcott, J. B. Kessler, The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons. *Am. Econ. J. Appl. Econ.* **11**, 236–276 (2019). doi: 10.1257/app.20170328



## ACKNOWLEDGMENTS

This paper reflects conversations and comments from many people, including D. Bjorkegren, S. Cole, T. Harigaya, J. Li, R. Maertens, R. Meager, G. Rao, W. Więcek, and G. Zane. We are grateful to F. Mashwama and J. Que for assistance in preparing this Review. **Competing interests:** R.F. and M.K. have received research grants from Precision Agriculture for Development. M.K. serves on the board of directors of Precision Agriculture for Development. The authors declare no other competing interests.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/eaay3038/suppl/DC1
Supplementary Text
References (*83*–*86*)

10.1126/science.aay3038

REVIEW SUMMARY

CLIMATE CHANGE

# Measuring the success of climate change adaptation and mitigation in terrestrial ecosystems

**Michael D. Morecroft*, Simon Duffield, Mike Harley, James W. Pearce-Higgins, Nicola Stevens, Olly Watts, Jeanette Whitaker**

**BACKGROUND:** Responding effectively to climate change requires urgent action to halt net greenhouse gas (GHG) emissions and to adapt to changes that cannot be prevented. The Paris Agreement of the United Nations Framework Convention on Climate Change has committed governments to the following: keeping global temperature rise below 2°C, pursuing efforts to limit it to 1.5°C, and adapting to reduce the vulnerability of people and ecosystems to the damaging consequences of a changing climate.

When protected, restored, or managed appropriately, natural and seminatural ecosystems make critical contributions to climate change mitigation and to helping people adapt to climate change. Ecosystems themselves are vulnerable to climate change, but by restoring natural ecosystem processes, resilience can be built, and a wide range of adaptation strategies can ameliorate the impacts.

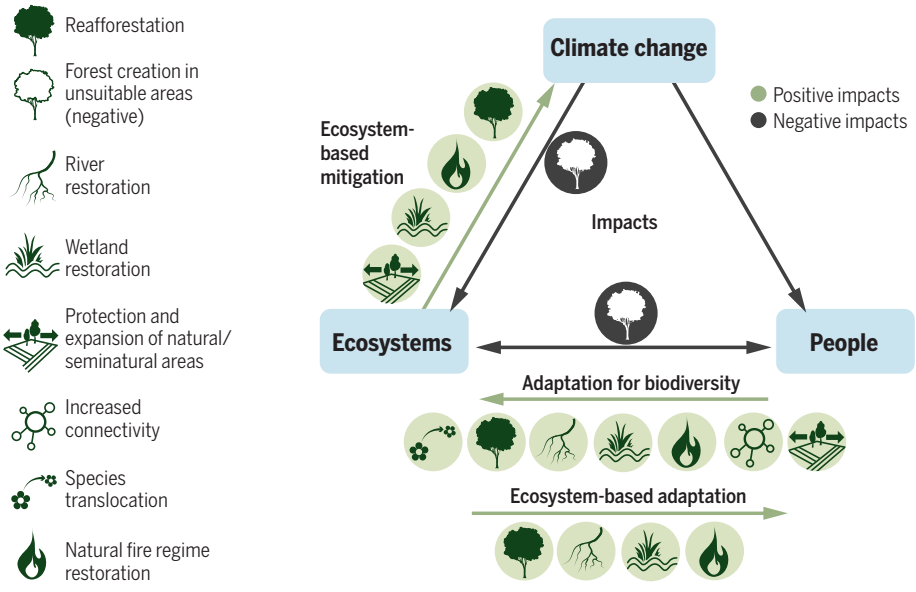
Both synergies and conflicts between different objectives can arise, and it is essential to have clarity about what constitutes success across the range of adaptation and mitigation outcomes and to track progress. The success of ecosystem-based mitigation can be measured in terms of falling net emissions and stabilization of atmospheric $CO_2$ concentration. Although this is conceptually straightforward, it can be difficult to measure ecosystem fluxes accurately. Adaptation is more complicated because it encompasses a wide range of objectives, with respect to people and biodiversity, including both reducing vulnerability and managing unavoidable change.

**ADVANCES:** Many studies have investigated how nature-based solutions can contribute to climate change mitigation and adaptation. The evidence is now clear that protecting and restoring ecosystems is essential to holding global temperature rise to between 1.5° and 2°C. The value of different interventions for reducing GHG emissions and promoting carbon sequestration can be quantified with varying degrees of confidence. The evidence for the effectiveness, opportunities, and limitations of ecosystem-based adaptation in enabling people to cope with climate change is also growing, and these approaches are starting to be implemented. Adaptation to reduce the vulnerability of biodiversity and ecosystems themselves to climate change has been discussed over many years but proposed measures remain largely untested. This is starting to change, with recent studies gathering empirical evidence of the factors that influence the vulnerability of ecosystems and biodiversity. Nevertheless, evaluation and reporting of adaptation is currently focused on planning and implementation of actions rather than on assessment of whether these programs have successfully reduced vulnerability.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaw9256

**OUTLOOK:** A picture is emerging of what successful adaptation and mitigation in terrestrial ecosystems looks like when it is built around protecting and restoring natural ecosystem processes. To realize the potential of ecosystems to ameliorate climate change requires integrated actions that are consistent with wider biodiversity and sustainable development goals. High-carbon ecosystems, particularly forests and peatlands, are essential, but other ecosystems, such as savannas, are also important elements of wider nature-based solutions and should be protected and restored. Pursuing mitigation objectives alone risks perverse outcomes that increase rather than reduce vulnerability. Further work is required to test the effectiveness of specific ecosystem-based mitigation and adaptation measures and what works best to support biodiversity in a changing climate. More-robust monitoring and evaluation are needed to drive progress. Measuring adaptation for biodiversity is particularly challenging, and monitoring and management will need to develop together as we learn from experience. ■




**The role of natural and seminatural ecosystems in adaptation to and mitigation of climate change.** The flow diagram shows the relationships between adaptation for biodiversity, ecosystem-based adaptation for people, and ecosystem-based mitigation. Negative impacts of climate change are shown in dark gray, and positive responses are shown in green. Successful ecosystem response to climate change depends on an integrated approach to ensure that synergistic effects are maximized and harms are avoided.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mike.morecroft@naturalengland.org.uk
Cite this article as M. D. Morecroft *et al.*, *Science* **366**, eaaw9256 (2019). DOI: 10.1126/science.aaw9256

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

REVIEW

CLIMATE CHANGE

# Measuring the success of climate change adaptation and mitigation in terrestrial ecosystems

Michael D. Morecroft[1,2]*, Simon Duffield[1], Mike Harley[3], James W. Pearce-Higgins[4,5], Nicola Stevens[6], Olly Watts[7], Jeanette Whitaker[8]

Natural and seminatural ecosystems must be at the forefront of efforts to mitigate and adapt to climate change. In the urgency of current circumstances, ecosystem restoration represents a range of available, efficient, and effective solutions to cut net greenhouse gas emissions and adapt to climate change. Although mitigation success can be measured by monitoring changing fluxes of greenhouse gases, adaptation is more complicated to measure, and reductions in a wide range of risks for biodiversity and people must be evaluated. Progress has been made in the monitoring and evaluation of adaptation and mitigation measures, but more emphasis on testing the effectiveness of proposed strategies is necessary. It is essential to take an integrated view of mitigation, adaptation, biodiversity, and the needs of people, to realize potential synergies and avoid conflict between different objectives.

Ecosystems are an essential element of climate change mitigation and adaptation, with the potential to reduce both net greenhouse gas (GHG) emissions and vulnerability to climate change. Approximately a quarter of postindustrial GHG emissions have come from the degradation of ecosystems (*1*), and the need for climate change mitigation based on restoring ecosystems or more sustainable land management is increasingly recognized (*2*). Over the past 10 years, carbon uptake by terrestrial ecosystems has been ~3.2 metric gigatons of carbon per year (GtC $y^{-1}$)—equivalent to one-third of emissions from fossil fuel burning (9.4 GtC $y^{-1}$) (*3*). However, land-use change simultaneously caused emissions of 1.5 GtC $y^{-1}$. Reversing this flux is an essential element of climate change mitigation. There is also increasing evidence that adaption responses that use ecosystems can reduce the risks of climate change–associated events to people (*4*), such as flooding or heat waves. Ecosystems are themselves vulnerable to climate change, but an increasing number of studies are showing that this vulnerability can be reduced when they are protected, restored, or managed for adaptation.

Ecosystem-based, or nature-based, solutions are distinctive in that they provide mitigation and adaptation benefits at the same time as benefitting biodiversity and human health and well-being (*5*). Furthermore, they have the potential to deliver adaptation and mitigation in a synergistic way. However, there is also the potential for conflicts between different objectives. A clear understanding of what success looks like (*6*) for both adaptation and mitigation alongside broader biodiversity and human factors is needed.

Monitoring and evaluation of actions aimed at adapting to and mitigating climate change is essential to driving progress and developing techniques (*6*). Progress with mitigation can be measured in terms of changes in GHG fluxes. However, adaptation is more conceptually challenging (*7*, *8*). First, because it may not be possible to fully assess the effectiveness of an adaptation strategy in preventing adverse impacts until decades later. Second, because no single metric or even a small range of metrics will adequately sum up progress across the many and varied aspects of adaptation. Third, because there are risks that reducing vulnerability in one sector may increase vulnerability in another. Finally, objectives may need to change over time, because what constitutes good adaptation at a global temperature rise of 1.5° to 2°C does not necessarily constitute good adaptation at 3° to 4°C. It is also possible that there may not be agreement among different actors about the goals of adaptation.

## What constitutes success in climate change adaptation and mitigation in ecosystems?

In simple terms, success in mitigation means preventing emissions and increasing carbon sequestration. Historically, the main source of GHG emissions from terrestrial ecosystems has been deforestation (*9*), but emissions from degraded peatlands, melting permafrost, more frequent or more intense wildfires, and other sources are compounding the problem (*10*). Natural and seminatural ecosystems are important elements of mitigation strategies because of their capacity to remove $CO_2$ from the atmosphere, which could partially offset emissions in sectors that are hard to decarbonize, such as aviation (*11*, *12*). Measures that have the greatest potential to deliver climate change mitigation in terrestrial ecosystems include protection of intact carbon stores, avoided deforestation, reforestation of formerly forested land, and restoration of degraded peatlands. (Other habitats, including coastal and marine systems, also store carbon but fall outside the scope of this Review.)

There is good evidence that reforestation of formerly forested land can create a large carbon sink in its early decades and, in the longer term, store considerable amounts of carbon. From 2001 to 2010, for the moist tropics and boreal Siberia, Pugh *et al.* (*13*) estimated a carbon sink of 1.30 GtC $y^{-1}$ in forest stands regrowing after past disturbance and 0.85 GtC $y^{-1}$ in intact old-growth forest. Likewise, allowing natural regeneration of forest on 350 Mha of formerly forested land in the tropics and subtropics has been estimated to store 42 GtC $y^{-1}$ by 2100 (*14*). Restoring degraded peatlands can substantially reduce (*15*) the large GHG emissions resulting from draining, burning, and cultivation (*16*). Climate change will have an increasingly negative impact on the capacity of many natural ecosystems to sequester or store carbon. There is also an increasing risk of exceeding tipping points that cannot easily be reversed, such as catastrophic permafrost melt (*17*). Rapid GHG emission reduction in all sectors is a priority to reduce such risks.

What constitutes success in adaptation has been widely discussed over the past two decades (*6*, *18*). For biodiversity conservation, adaptation includes a wide range of actions at different scales, from individual species to habitats and ecosystems (*19*, *20*). The evidence for these actions' effectiveness in supporting biodiversity conservation in a changing climate has grown rapidly in recent years. Nevertheless, many approaches have been based on theory, modeling, and observations comparing different locations or changes in time; relatively few studies have assessed the effectiveness of these measures experimentally (*21*). Three broad approaches to adaptation for biodiversity conservation can be identified: ecological restoration, direct intervention to reduce vulnerability of species and habitats, and adjusting conservation management and objectives.

Ecological restoration is important because the impacts of climate change are often exacerbated in degraded ecosystems. Restoration in this sense is focused on restoring natural

[1]Natural England, York YO1 7PX, UK. [2]Environmental Change Institute, University of Oxford, Oxford OX1 3QY, UK. [3]Climate Resilience Ltd., Stamford PE9 4AU, UK. [4]British Trust for Ornithology, Thetford, Norfolk IP24 2PU, UK. [5]Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge CB2 3QZ, UK. [6]Department of Zoology and Botany, Stellenbosch University, Stellenbosch 7600, South Africa. [7]Royal Society for the Protection of Birds, Sandy SG19 2DL, UK. [8]UK Centre for Ecology and Hydrology, Lancaster Environment Centre, Lancaster LA1 4AP, UK.
*Corresponding author. Email: mike.morecroft@naturalengland.org.uk

**Box 1. Assessing the potential for synergistic outcomes from ecosystem-based adaptation and mitigation actions.**

**Will the action:**

1. Reduce GHG emissions or promote sequestration of carbon?
2. Continue to be viable for a range of plausible future climate scenarios?
3. Increase the resilience of biodiversity to climate change?
4. Help people adapt to climate change?
5. Maintain or enhance the biodiversity of a region, now and under future climates?
6. Maintain or increase the provision of ecosystem services on which local people depend, including water, food, and materials, now and under future climates?
7. Lead to the displacement of emissions to another location?

processes, including catchment hydrology (*22*, *23*) and fire regimes (*24*, *25*). "Renovation" has been proposed as an alternate term (*20*), to distinguish this process from seeking to restore a former state, which may no longer be possible with climate change. Increasing habitat patch size can support more-resilient populations (*26*), and increasing habitat connectivity (*27*) enables some species to track changing climatic conditions in fragmented landscapes. Direct interventions range from targeted management of vulnerable species (*28*) and species translocation (*29*, *30*) to manipulating habitats, e.g., by shading watercourses with trees to reduce water temperatures (*31*, *32*). Adjusting conservation objectives and ways of working is becoming increasingly necessary as climate change impacts increase; for example, changing species distributions may mean that species will need protection in places they did not formerly inhabit (*33*, *34*). The longer that warming continues, the more that adaptation for biodiversity will need to shift toward managing change (*20*, *35*, *36*) rather than building resilience. However, recent research has demonstrated the existence of refugia (*37*, *38*) in which species survival is more likely because of microclimatic conditions, for example, on north-facing slopes and other locally cool spots in a landscape. Identifying and protecting these areas is therefore a good strategy for promoting species survival and is consistent with protecting and restoring large heterogeneous areas.

Ecosystem-based adaptation (EbA) is the use of biodiversity and ecosystem services to help people adapt to climate change. EbA includes natural flood management, in which peak flows are reduced and flood storage capacity is increased by restoring wetlands and natural features of rivers, such as meanders and woody debris; creation of green space and planting of trees to provide local cooling for people or livestock; and establishment of vegetation on slopes prone to landslip during extreme rainfall events. The evidence base for EbA has developed rapidly, with many studies demonstrating that the approach can be more cost-effective in the delivery of adaptation outcomes, and that it provides multiple cobenefits and represents a more sustainable approach than engineered adaptation measures (*39*). However, the adoption of EbA approaches is patchy (*40*) and the involvement and empowerment of local communities and stakeholders is essential for successful EbA (*41*, *42*).

The particular value of nature-based solutions in addressing climate change is their capacity to simultaneously provide mitigation and adaptation along with a wide range of other benefits for biodiversity and people. Box 1 summarizes the key issues that determine success or failure across all these areas.

### *Synergies*

Protecting existing natural areas is a cornerstone of conservation, which is even more important in a changing climate, especially areas where there are important carbon stores or potential refugia for species. Intact forests are important for a range of climate-related, ecological, and societal reasons, including carbon storage, carbon sequestration, local climate regulation, water supply, and biodiversity (*43*). Similarly, peatlands are important carbon stores and, like other natural wetlands, support biodiversity and contribute to water resources and flood management.

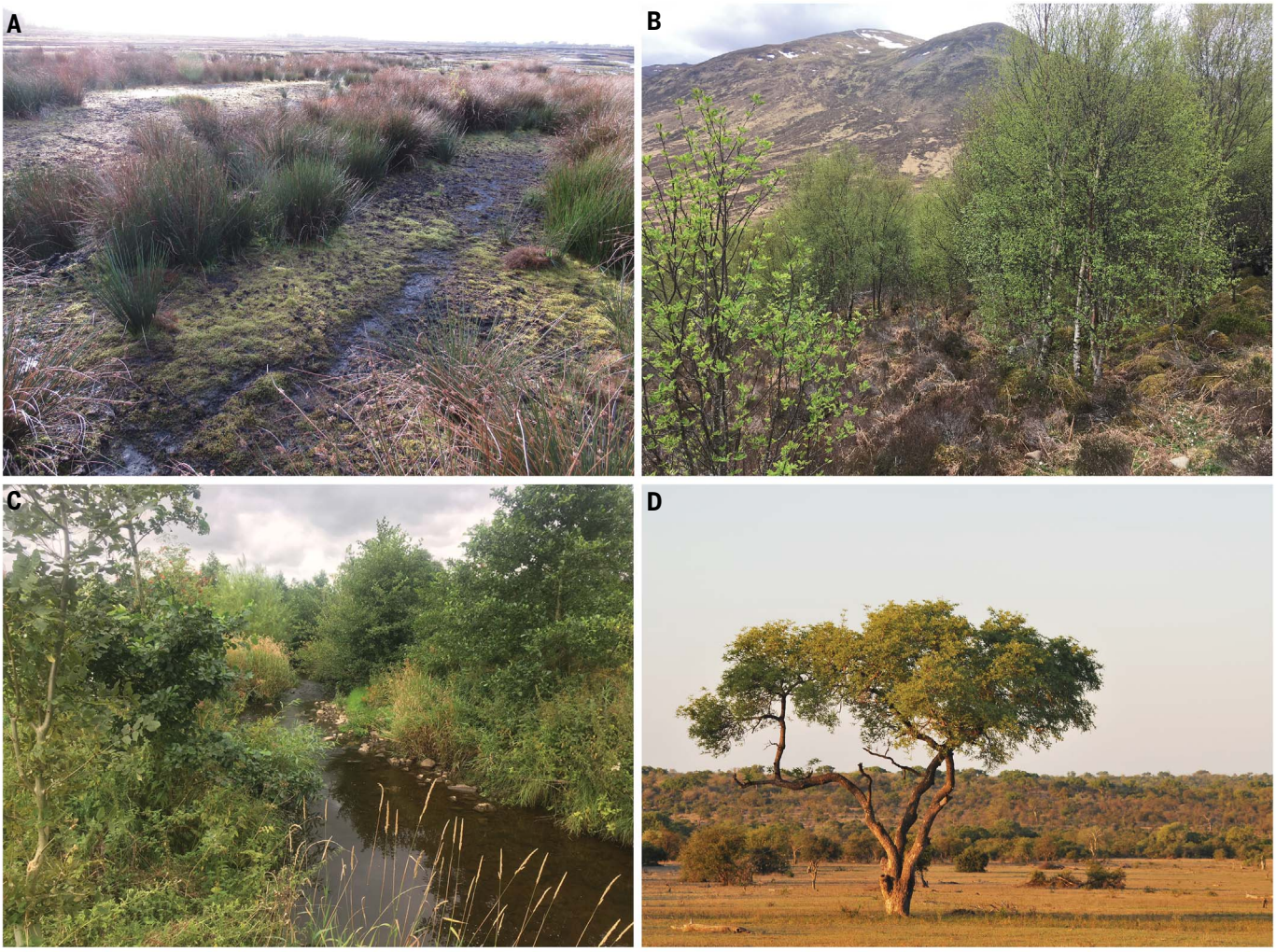
Given the degraded state of most of Earth's ecosystems, restoration is essential to realize their full potential for adaptation and mitigation. Restoring natural ecosystem functions, particularly hydrology and carbon dynamics, is central to this goal (Fig. 1). Catchment restoration (*44*), including regeneration of wetlands and reversing the canalization of rivers, can reduce flood risk by retaining water higher in the catchment. The additional advantages of such measures include maintaining water supplies during periods of drought, enhancing biodiversity, and contributing to carbon storage and sequestration. Reforestation can both sequester carbon and have benefits for adaptation, including increased rainfall infiltration into soil (*45*, *46*), floodwater impedance (*44*), and provision of shade (*31*). Restoring savannas, by removing trees, reseeding grasslands, and reinstating natural fire regimes (*47*), increases resilience, supports carbon storage in soils (*48*), protects water resources (*49*), and reduces the risk of catastrophic fires (*50*).

Restoration of natural processes is likely to have multiple benefits in tackling climate change. However, it will increasingly need to be complemented with active intervention or adjustment of conservation objectives, as described above, to reduce vulnerability to climate change and continue delivering adaptation and mitigation services.

### *Conflicts*

Although nature-based solutions offer the potential for win-win-win outcomes for mitigation, adaptation, and biodiversity, these are not guaranteed, and there is potential for conflict, especially when inappropriate interventions are employed. There are real risks in pursuing one objective without taking proper account of others. An important current concern is tree planting in inappropriate places. Although reforestation of formerly forested land can bring great benefits for adaptation, mitigation, and biodiversity, tree planting in other places can cause serious problems and can even exacerbate climate change impacts (*48*). Naturally open ecosystems are uniquely adapted to local conditions. Grass-dominated savannas have diverse communities and many endemic species (*51*) that have evolved in a high-light environment with vegetation-environment feedbacks, including high levels of herbivory and fire (*52*). Tree planting in historically open ecosystems would be harmful. Unfortunately, global scale analyses aimed at identifying degraded forest areas suitable for afforestation (*53*) cannot reliably separate naturally open ecosystems, such as savanna, from degraded forests. Here, other sources of knowledge and data, including the involvement of local communities, are essential. Similarly, the afforestation of formerly treeless peatlands that have been drained may not supply any mitigation benefits from growing timber because of release of GHGs from the drained peat (*54*) and may, in addition, lead to biodiversity losses (*55*).

Ecosystem restoration will not be possible in all places. Many formerly forested and peatland areas have been cleared and drained for agriculture. Ecosystem restoration in these regions may reduce crop production or displace it to other locations where it may have an equally or even more negative impact. A wider and coordinated strategy to tackle climate change and promote sustainable development therefore needs to include actions such as reducing food waste and dietary change and sustainable increases in agricultural productivity in some places to allow




CREDITS: (A) OLLY WATTS/RSPB; (B) MIKE MORECROFT/NATURAL ENGLAND; (C) ENVIRONMENT AGENCY; (D) NICOLA STEVENS

**Fig. 1. Examples of nature-based solutions to climate change.** (**A**) Restoration of a lowland raised bog at Bolton Fell Moss, Cumbria, England—a former commercial peat extraction site. (**B**) Natural regeneration of trees following reduction in grazing at Creag Meagaidh National Nature Reserve, Scotland. This is a landscape that would once have been forested below the natural treeline, but the forest was subsequently cleared, and grazing prevented reestablishment. (**C**) Establishment of woodland next to a stream provides shading to maintain lower water temperatures and contributes to carbon sequestration. (**D**) Savanna in a natural state, with large areas of open grassland and scattered trees. Protection and restoration of this state using native herbivores and natural fire regimes is the best approach to climate change adaptation as well as biodiversity. This ecosystem is not suitable for tree planting.

restoration in others. Sustainable approaches, including organic and regenerative farming (*56*), which can support some aspects of biodiversity and contribute to climate change mitigation, can also play a role in wider land-use strategies. Negative emissions technologies, especially biofuels with carbon capture and storage (BECCS), are often included in scenarios for meeting the Paris Agreement commitments, but these technologies create a demand for land (*11*, *57*). The benefits of BECCS for mitigation need to be considered alongside the implications for adaptation, biodiversity, and food security and will require careful management and monitoring (*58*). When a broader frame of reference is considered, the protection and restoration of natural ecosystems may, in many circumstances, be a better option than resorting to BECCS.

## Measuring and reporting progress

It is important that success in deploying nature-based solutions can be monitored and evaluated, to ensure that all intended benefits are delivered, that progress is being made, and that actions can be prioritized on the basis of evidence. To enable this to happen, metrics are required for biodiversity, mitigation, and adaptation parameters, which reflect the needs of local communities as well as international reporting.

Mitigation is measured by quantifying emissions and removal of GHGs in terms of $CO_2$ equivalents. The Paris Agreement of the United Nations Framework Convention on Climate Change (UNFCCC) requires countries to report emissions on the basis of observation-constrained models (*3*). However, there are considerable gaps and challenges in reporting ecosystem emissions and removals. Many confounding factors can influence assessment of mitigation outcomes, including the slow rate of carbon sequestration in many ecosystems and the risk of stored carbon being released by wildfire or land-use change. Hence, long-term measurement is required, as patterns of carbon uptake change over time.

Under the UNFCCC, adaptation reporting is based on the development and delivery of broad-ranging National Adaptation Plans (*59*, *60*). The Paris Agreement also stipulates a "global stocktake," to review adaptation effectiveness and progress made toward the

global adaptation goals of enhancing adaptive capacity, strengthening resilience, and reducing vulnerability to climate change. Many variables that affect human vulnerability can be measured in a straightforward way. For example, fluvial flood risk relates closely to water flow in rivers, which can be measured directly (*61*). However, for biodiversity, especially at the species level, vulnerability is harder to measure and is associated with a high degree of uncertainty (*62*–*64*).

Lessons can be learned from the long history of species monitoring. Although well-chosen individual species can act as indicators for a wider range of species (*65*), in the context of climate change, more-robust indicators can be developed from the contrasting responses of different species (*66*, *67*), including sophisticated assemblage structure indicators (*68*).

Several indicators of climate change impact have been developed that track observed species' population and community responses to climate change (*69*–*71*). It may be possible to develop indicators of adaptation on the basis of further development of these impact indicators, but when doing so, it will be important that attribution to climate change is not confounded with other drivers of change, which may vary between species. In addition, the efficacy of species as indicators may change over time, as a result of altering interspecific interactions, density dependence, and even evolutionary adaptation. Finally, we need to consider that the past, upon which the indicators are devised, may not be a good predictor of the future.

Given these concerns, the development of true outcome-based adaptation indicators is challenging, particularly where adaptation actions are targeted at managing change. An alternative approach is to monitor the deployment of adaptation measures for which there is good evidence of effectiveness, such as increasing patch size (*26*) and protecting refugia (*36*), in combination with existing biodiversity surveillance. At present, the main limitation to this approach is our understanding of the effectiveness of adaptation measures (*21*) and the narrow geographical representation of existing studies, but this evidence base should improve over time. The concept of adaptive monitoring linked with adaptive management (*72*, *73*), in which both monitoring and management of the natural environment develop and evolve as situations and knowledge change, is a powerful one that will need to be embraced given the need for prompt action in the face of an uncertain future.

## Outlook

Tackling climate change is an urgent priority and a drive to restore degraded ecosystems needs to be at the heart of it. Natural ecosystems take up approximately a third of current fossil fuel emissions and will be vital for future carbon sequestration. The distinctive value of natural ecosystems in climate change mitigation is that they can protect biodiversity and help societies to adapt to climate change. To realize the potential of nature-based solutions to climate change requires more investment, technical expertise, and the active participation of local communities. There is a pressing need for clear measures of success and effective monitoring and evaluation that cover mitigation, adaptation, and biodiversity outcomes to drive progress.

## REFERENCES AND NOTES


1. T. F. Stocker *et al*., Eds., *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2014).
2. B. W. Griscom *et al*., Natural climate solutions. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11645–11650 (2017). doi: 10.1073/pnas.1710465114; pmid: 29078344
3. C. Le Quéré *et al*., Global carbon budget 2018. *Earth Syst. Sci. Data* **10**, 2141–2194 (2018). doi: 10.5194/essd-10-2141-2018
4. R. Munang *et al*., Climate change and Ecosystem-based Adaptation: A new pragmatic approach to buffering climate change impacts. *Curr. Opin. Environ. Sustain.* **5**, 67–71 (2013). doi: 10.1016/j.cosust.2012.12.001
5. S. Naumann *et al*., "Assessment of the potential of ecosystem-based approaches to climate change adaptation and mitigation in Europe. Final report to the European Commission, DG Environment," (Contract no. 070307/2010/580412/SER/B2, Ecologic Institute and Environmental Change Institute, Oxford University Centre for the Environment, 2011).
6. S. C. Moser, M. T. Boykoff, in *Successful Adaptation to Climate Change: Linking Science and Policy in a Rapidly Changing World*, S. C. Moser, M. T. Boykoff, Eds. (Routledge, 2013), pp. 25–58.
7. L. Dilling *et al*., Is adaptation success a flawed concept? *Nat. Clim. Chang.* **9**, 572–574 (2019). doi: 10.1038/s41558-019-0539-0
8. L. Berrang-Ford *et al*., Tracking global climate change adaptation among governments. *Nat. Clim. Chang.* **9**, 440–449 (2019). doi: 10.1038/s41558-019-0490-0
9. S. Federici, F. N. Tubiello, M. Salvatore, H. Jacobs, J. Schmidhuber, New estimates of $CO_2$ forest emissions and removals: 1990–2015. *For. Ecol. Manage.* **352**, 89–98 (2015). doi: 10.1016/j.foreco.2015.04.022
10. S. Rossi *et al*., FAOSTAT estimates of greenhouse gas emissions from biomass and peat fires. *Clim. Change* **135**, 699–711 (2016). doi: 10.1007/s10584-015-1584-y
11. V. Masson-Delmotte *et al*., Eds., "Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty" (Intergovernmental Panel on Climate Change, 2018).
12. S. Roe *et al*., Contribution of the land sector to a 1.5° C world. *Nat. Clim. Chang.* **9**, 817–828 (2019). doi: 10.1038/s41558-019-0591-9
13. T. A. M. Pugh *et al*., Role of forest regrowth in global carbon sink dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4382–4387 (2019). doi: 10.1073/pnas.1810512116; pmid: 30782807
14. S. L. Lewis, C. E. Wheeler, E. T. A. Mitchard, A. Koch, Restoring natural forests is the best way to remove atmospheric carbon. *Nature* **568**, 25–28 (2019). doi: 10.1038/d41586-019-01026-8; pmid: 30940972
15. D. Wilson *et al*., Greenhouse gas emission factors associated with rewetting of organic soils. *Mires Peat* **17**, 1–28 (2016).
16. J. Leifeld, L. Menichetti, The underappreciated potential of peatlands in global climate change mitigation strategies. *Nat. Commun.* **9**, 1071 (2018). doi: 10.1038/s41467-018-03406-6; pmid: 29540695
17. K. Schaefer, H. Lantuit, V. E. Romanovsky, E. A. Schuur, R. Witt, The impact of the permafrost carbon feedback on global climate. *Environ. Res. Lett.* **9**, 085003 (2014). doi: 10.1088/1748-9326/9/8/085003
18. W. N. Adger, N. W. Arnell, E. L. Tompkins, Successful adaptation to climate change across scales. *Glob. Environ. Change* **15**, 77–86 (2005). doi: 10.1016/j.gloenvcha.2004.12.005
19. N. E. Heller, E. S. Zavaleta, Biodiversity management in the face of climate change: A review of 22 years of recommendations. *Biol. Conserv.* **142**, 14–32 (2009). doi: 10.1016/j.biocon.2008.10.006
20. S. M. Prober, V. A. Doerr, L. M. Broadhurst, K. J. Williams, F. Dickson, Shifting the conservation paradigm: A synthesis of options for renovating nature under climate change. *Ecol. Monogr.* **89**, e01333 (2019). doi: 10.1002/ecm.1333
21. O. Greenwood, H. L. Mossman, A. J. Suggitt, R. J. Curtis, I. M. D. Maclean, Using *in situ* management to conserve biodiversity under climate change. *J. Appl. Ecol.* **53**, 885–894 (2016). doi: 10.1111/1365-2664.12602; pmid: 27609987
22. W. R. Moomaw *et al*., Wetlands in a changing climate: Science, policy and management. *Wetlands* **38**, 183–205 (2018). doi: 10.1007/s13157-018-1023-8
23. B. L. Timpane-Padgham, T. Beechie, T. Klinger, A systematic review of ecological attributes that confer resilience to climate change in environmental restoration. *PLOS ONE* **12**, e0173812 (2017). doi: 10.1371/journal.pone.0173812; pmid: 28301560
24. J. S. Halofsky *et al*., The nature of the beast: Examining climate adaptation options in forests with stand-replacing fire regimes. *Ecosphere* **9**, e02140 (2018). doi: 10.1002/ecs2.2140
25. P. F. Hessburg *et al*., Restoring fire-prone Inland Pacific landscapes: Seven core principles. *Landsc. Ecol.* **30**, 1805–1835 (2015). doi: 10.1007/s10980-015-0218-0
26. T. H. Oliver *et al*., Interacting effects of climate change and habitat fragmentation on drought-sensitive butterflies. *Nat. Clim. Chang.* **5**, 941–945 (2015). doi: 10.1038/nclimate2746
27. A. T. Keeley *et al*., New concepts, models, and assessments of climate-wise connectivity. *Environ. Res. Lett.* **13**, 073002 (2018). doi: 10.1088/1748-9326/aacb85
28. J. W. Pearce-Higgins *et al*., Site-based adaptation reduces the negative effects of weather upon a southern range margin Welsh black grouse *Tetrao tetrix* population that is vulnerable to climate change. *Clim. Change* **153**, 253–265 (2019). doi: 10.1007/s10584-019-02372-2
29. R. W. Brooker *et al*., Tiny niches and translocations: The challenge of identifying suitable recipient sites for small and immobile species. *J. Appl. Ecol.* **55**, 621–630 (2018). doi: 10.1111/1365-2664.13008
30. S. G. Willis *et al*., Assisted colonization in a changing climate: A test-study using two UK butterflies. *Conserv. Lett.* **2**, 46–52 (2009). doi: 10.1111/j.1755-263X.2008.00043.x
31. S. M. Thomas, S. W. Griffiths, S. J. Ormerod, Beyond cool: Adapting upland streams for climate change using riparian woodlands. *Glob. Change Biol.* **22**, 310–324 (2016). doi: 10.1111/gcb.13103; pmid: 26395251
32. S. Broadmeadow, J. Jones, T. Langford, P. Shaw, T. Nisbet, The influence of riparian shade on lowland stream water temperatures in southern England and their viability for brown trout. *River Res. Appl.* **27**, 226–237 (2011). doi: 10.1002/rra.1354
33. P. Glick, H. Chmura, B. A. Stein, "Moving the conservation goalposts: A review of climate change adaptation literature" (National Wildlife Federation, 2011).
34. R. L. Pressey, M. Cabeza, M. E. Watts, R. M. Cowling, K. A. Wilson, Conservation planning in a changing world. *Trends Ecol. Evol.* **22**, 583–592 (2007). doi: 10.1016/j.tree.2007.10.001; pmid: 17981360
35. B. A. Stein, M. R. Shaw, in *Successful Adaptation to Climate Change: Linking Science and Policy in a Rapidly Changing World*, S. C. Moser, M. T. Boykoff, Eds. (Routledge, 2013), pp. 74–90.
36. M. D. Morecroft, H. Q. Crick, S. J. Duffield, N. A. Macgregor, Resilience to climate change: Translating principles into practice. *J. Appl. Ecol.* **49**, 547–551 (2012). doi: 10.1111/j.1365-2664.2012.02136.x
37. A. J. Suggitt *et al*., Extinction risk from climate change is reduced by microclimatic buffering. *Nat. Clim. Chang.* **8**, 713–717 (2018). doi: 10.1038/s41558-018-0231-9
38. T. L. Morelli *et al*., Managing climate change refugia for climate adaptation. *PLOS ONE* **11**, e0159909 (2016). doi: 10.1371/journal.pone.0159909; pmid: 27509088

39. H. P. Jones, D. G. Hole, E. S. Zavaleta, Harnessing nature to help people adapt to climate change. *Nat. Clim. Chang.* **2**, 504–509 (2012). doi: 10.1038/nclimate1463
40. J. Nalau, S. Becken, B. Mackey, Ecosystem-based Adaptation: A review of the constraints. *Environ. Sci. Policy* **89**, 357–364 (2018). doi: 10.1016/j.envsci.2018.08.014
41. S. Woroniecki, C. Wamsler, E. Boyd, The promises and pitfalls of ecosystem-based adaptation to climate change as a vehicle for social empowerment. *Ecol. Soc.* **24**, 4 (2019). doi: 10.5751/ES-10854-240204
42. A. McVittie, L. Cole, A. Wreford, A. Sgobbi, B. Yordi, Ecosystem-based solutions for disaster risk reduction: Lessons from European applications of ecosystem-based adaptation measures. *Int. J. Disaster Risk Reduct.* **32**, 42–54 (2018). doi: 10.1016/j.ijdrr.2017.12.014
43. J. E. M. Watson *et al*., The exceptional value of intact forest ecosystems. *Nat. Ecol. Evol.* **2**, 599–610 (2018). doi: 10.1038/s41559-018-0490-x; pmid: 29483681
44. S. J. Dixon, D. A. Sear, N. A. Odoni, T. Sykes, S. N. Lane, The effects of river restoration on catchment scale flood risk and flood hydrology. *Earth Surf. Process. Landf.* **41**, 997–1008 (2016). doi: 10.1002/esp.3919
45. Z. L. Carroll, S. B. Bird, B. A. Emmett, B. Reynolds, F. L. Sinclair, Can tree shelterbelts on agricultural land reduce flood risk? *Soil Use Manage.* **20**, 357–359 (2004). doi: 10.1079/SUM2004266
46. M. G. Anderson, C. E. Ferree, A. P. Olivero, F. Zhao, Assessing floodplain forests: Using flow modeling and remote sensing to determine the best places for conservation. *Nat. Areas J.* **30**, 39–52 (2010). doi: 10.3375/043.030.0105
47. E. Buisson *et al*., Resilience and restoration of tropical and subtropical grasslands, savannas, and grassy woodlands. *Biol. Rev. Camb. Philos. Soc.* **94**, 590–609 (2019). doi: 10.1111/brv.12470; pmid: 30251329
48. J. W. Veldman *et al*., Comment on "The global tree restoration potential". *Science* **366**, eaay7976 (2019). doi: 10.1126/science.aay7976; pmid: 31624182
49. E. A. Honda, G. Durigan, Woody encroachment and its consequences on hydrological processes in the savannah. *Philos. Trans. R. Soc. London Ser. B* **371**, 20150313 (2016). doi: 10.1098/rstb.2015.0313; pmid: 27502378
50. S. Archibald, C. E. Lehmann, J. L. Gómez-Dans, R. A. Bradstock, Defining pyromes and global syndromes of fire regimes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6442–6447 (2013). doi: 10.1073/pnas.1211466110; pmid: 23559374
51. B. P. Murphy, A. N. Andersen, C. L. Parr, The underestimated biodiversity of tropical grassy biomes. *Philos. Trans. R. Soc. London Ser. B* **371**, 20150319 (2016). doi: 10.1098/rstb.2015.0319; pmid: 27502382
52. C. E. Lehmann *et al*., Savanna vegetation-fire-climate relationships differ among continents. *Science* **343**, 548–552 (2014). doi: 10.1126/science.1247355; pmid: 24482480
53. J.-F. Bastin *et al*., The global tree restoration potential. *Science* **365**, 76–79 (2019). doi: 10.1126/science.aax0848; pmid: 31273120
54. H. Simola, A. Pitkänen, J. Turunen, Carbon loss in drained forestry peatlands in Finland, estimated by re-sampling peatlands surveyed in the 1980s. *Eur. J. Soil Sci.* **63**, 798–807 (2012). doi: 10.1111/j.1365-2389.2012.01499.x
55. J. D. Wilson *et al*., Modelling edge effects of mature forest plantations on peatland waders informs landscape-scale conservation. *J. Appl. Ecol.* **51**, 204–213 (2014). doi: 10.1111/1365-2664.12173
56. C. E. LaCanne, J. G. Lundgren, Regenerative agriculture: Merging farming and natural resource conservation profitably. *PeerJ* **6**, e4428 (2018). doi: 10.7717/peerj.4428; pmid: 29503771
57. H. Muri, The role of large—scale BECCS in the pursuit of the 1.5° C target: An Earth system model perspective. *Environ. Res. Lett.* **13**, 044010 (2018). doi: 10.1088/1748-9326/aab324
58. P. A. Turner *et al*., The global overlap of bioenergy and carbon sequestration potential. *Clim. Change* **148**, 1–10 (2018). doi: 10.1007/s10584-018-2189-z
59. Least Developed Countries Expert Group, "National Adaptation Plans: Technical guidelines for the national adaptation plan process" (United Nations Framework Convention on Climate Change, 2012).
60. H. Price-Kelly, A. Hammill, J. Dekens, T. Leiter, J. Olivier, "Developing national adaptation monitoring and evaluation systems: A guidebook" (Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, 2015).
61. O. Iacob, J. S. Rowan, I. Brown, C. Ellis, Evaluating wider benefits of natural flood management strategies: An ecosystem-based adaptation perspective. *Hydrol. Res.* **45**, 774–787 (2014). doi: 10.2166/nh.2014.184
62. J. W. Pearce-Higgins *et al*., A national-scale assessment of climate change impacts on species: Assessing the balance of risks and opportunities for multiple taxa. *Biol. Conserv.* **213**, 124–134 (2017). doi: 10.1016/j.biocon.2017.06.035
63. C. J. Wheatley *et al*., Climate change vulnerability for species—Assessing the assessments. *Glob. Change Biol.* **23**, 3704–3715 (2017). doi: 10.1111/gcb.13759; pmid: 28660715
64. W. B. Foden *et al*., Climate change vulnerability assessment of species. *Wiley Interdiscip. Rev. Clim. Change* **10**, e551 (2019). doi: 10.1002/wcc.551
65. D. B. Lindenmayer, G. E. Likens, Direct measurement versus surrogate indicator species for evaluating environmental change and biodiversity loss. *Ecosystems* **14**, 47–59 (2011). doi: 10.1007/s10021-010-9394-6
66. R. D. Gregory *et al*., Developing indicators for European birds. *Philos. Trans. R. Soc. London Ser. B* **360**, 269–288 (2005). doi: 10.1098/rstb.2004.1602; pmid: 15814345
67. M. Diekmann, Species indicator values as an important tool in applied plant ecology–A review. *Basic Appl. Ecol.* **4**, 493–506 (2003). doi: 10.1078/1439-1791-00185
68. G. M. Siriwardena, I. G. Henderson, D. G. Noble, R. J. Fuller, How can assemblage structure indices improve monitoring of change in bird communities using ongoing survey data? *Ecol. Indic.* **104**, 669–685 (2019). doi: 10.1016/j.ecolind.2019.05.046
69. V. Devictor *et al*., Differences in the climatic debts of birds and butterflies at a continental scale. *Nat. Clim. Chang.* **2**, 121–124 (2012). doi: 10.1038/nclimate1347
70. B. Martay *et al*., An indicator highlights seasonal variation in the response of Lepidoptera communities to warming. *Ecol. Indic.* **68**, 126–133 (2016). doi: 10.1016/j.ecolind.2016.01.057
71. P. A. Stephens *et al*., Consistent response of bird populations to climate change on two continents. *Science* **352**, 84–87 (2016). doi: 10.1126/science.aac4858; pmid: 27034371
72. D. B. Lindenmayer, G. E. Likens, A. Haywood, L. Miezis, Adaptive monitoring in the real world: Proof of concept. *Trends Ecol. Evol.* **26**, 641–646 (2011). doi: 10.1016/j.tree.2011.08.002; pmid: 21880394
73. R. A. Ims, N. G. Yoccoz, Ecosystem-based monitoring in the age of rapid climate change and new technologies. *Curr. Opin. Environ. Sustain.* **29**, 170–176 (2017). doi: 10.1016/j.cosust.2018.01.003


## ACKNOWLEDGMENTS

**Funding:** M.D.M. and S.D. were funded by Natural England; M.H. was funded by Climate Resilience Ltd.; J.W.P.-H. was funded by the British Trust for Ornithology; N.S. was funded by National Research Foundation, Alliance for Collaboration on Climate and Earth Systems Science (ACCESS) grant 114695; O.W. was funded by the Royal Society for the Protection of Birds; and J.W. was funded by the Natural Environment Research Council grant Locked Up (NE/S005137/1). **Competing interests:** The authors are not aware of any conflicts of interest.

10.1126/science.aaw9256

RESEARCH ARTICLE SUMMARY

STRUCTURAL BIOLOGY

# p27 allosterically activates cyclin-dependent kinase 4 and antagonizes palbociclib inhibition

**Keelan Z. Guiley, Jack W. Stevenson, Kevin Lou, Krister J. Barkovich, Vishnu Kumarasamy, Tilini U. Wijeratne, Katharine L. Bunch, Sarvind Tripathi, Erik S. Knudsen, Agnieszka K. Witkiewicz, Kevan M. Shokat, Seth M. Rubin***

**INTRODUCTION:** The cyclin D (CycD)–dependent kinases CDK4 and CDK6 (CDK4/6) phosphorylate the retinoblastoma (Rb) tumor suppressor protein to cause entry into the cell cycle in normal cells and in many cancers. Three small-molecule CDK4/6 inhibitors—palbociclib, ribociclib, and abemaciclib—are clinically approved in HER2-negative, ER-positive breast cancer in combination with antiestrogens. Although these drugs are already being used in clinical trials for diverse cancers, ongoing research is needed to better understand the mechanisms of inherent or acquired resistance to CDK4/6 inhibition.

**RATIONALE:** The intrinsically disordered proteins p21 and p27 are commonly known as CDK inhibitors, yet CDK4-CycD requires p21 or p27 (p21/p27) for assembly and activity in vivo. Moreover, all approved CDK4/6 inhibitors were developed against purified CDK4/6-CycD complexes that lacked p21/p27. To date, there are no structures or kinetic data explaining whether and how p21/p27 activates CDK4-CycD and influences the response of chemical inhibitors. We therefore set out to structurally and functionally characterize the trimeric p21/p27-CDK4-CycD complexes, as well as to investigate the mechanism of clinically approved CDK4/6 inhibitors.

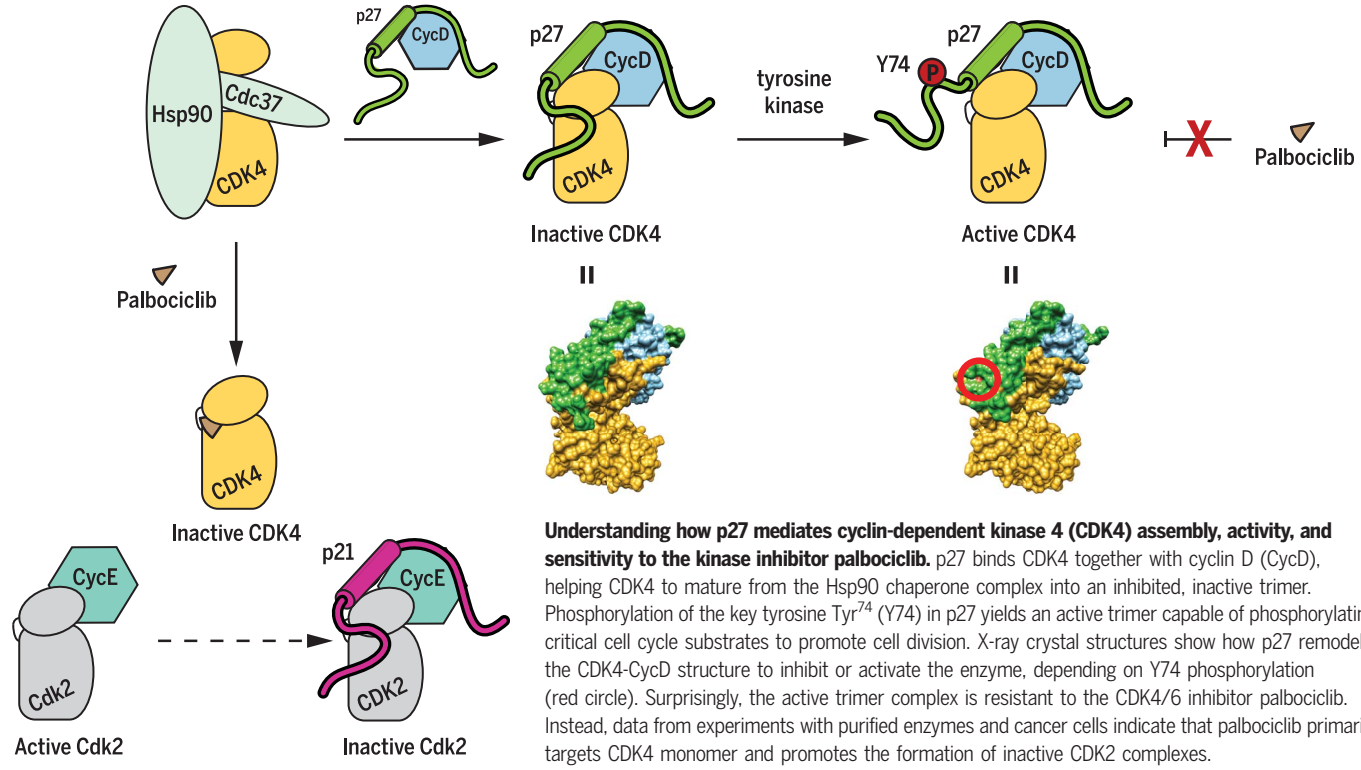
**RESULTS:** We present a crystal structure of the CDK4 holoenzyme, which reveals that p27 allosterically activates CDK4 to phosphorylate Rb by remodeling the adenosine triphosphate–binding site and by promoting release of the kinase activation segment. We find that tyrosine phosphorylation of p27 is required to activate CDK4 and that the lack of a key tyrosine in p21 makes it a poor activator. Surprisingly, we also find that the purified p27-CDK4-CycD1 complex is refractory to inhibition by approved CDK4/6 inhibitors and that endogenous p27-, CDK4-, and CycD-associated activity is also insensitive. Instead, palbociclib primarily targets CDK4/6 monomer in breast cancer cells, and this association indirectly inhibits the downstream Rb-inactivating kinase CDK2.

**CONCLUSION:** The success of CDK4/6 inhibitors demonstrates the clinical importance of Rb inactivation in cancer. Although treatment with palbociclib, abemaciclib, or ribociclib leads to low concentrations of phosphorylated Rb, we conclude that these small molecules do not directly inhibit CDK4 kinase activity in the breast cancer cells we examined. The mechanism that we propose, inhibition of complex assembly, parallels that of the endogenous CDK4/6 inhibitor protein p16. We conclude that mechanisms leading to CDK2 inhibition are critical for inducing cell cycle arrest and are likely critical determinants of whether cells are sensitive to the CDK4/6 inhibitors. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaw2106

The list of author affiliations is available in the full article online.
*Corresponding author. Email: srubin@ucsc.edu
Cite this article as K. Z. Guiley *et al.*, *Science* **366**, eaaw2106 (2019). DOI: 10.1126/science.aaw2106




**Understanding how p27 mediates cyclin-dependent kinase 4 (CDK4) assembly, activity, and sensitivity to the kinase inhibitor palbociclib.** p27 binds CDK4 together with cyclin D (CycD), helping CDK4 to mature from the Hsp90 chaperone complex into an inhibited, inactive trimer. Phosphorylation of the key tyrosine $Tyr^{74}$ (Y74) in p27 yields an active trimer capable of phosphorylating critical cell cycle substrates to promote cell division. X-ray crystal structures show how p27 remodels the CDK4-CycD structure to inhibit or activate the enzyme, depending on Y74 phosphorylation (red circle). Surprisingly, the active trimer complex is resistant to the CDK4/6 inhibitor palbociclib. Instead, data from experiments with purified enzymes and cancer cells indicate that palbociclib primarily targets CDK4 monomer and promotes the formation of inactive CDK2 complexes.

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

# p27 allosterically activates cyclin-dependent kinase 4 and antagonizes palbociclib inhibition

Keelan Z. Guiley[1,2], Jack W. Stevenson[2], Kevin Lou[2], Krister J. Barkovich[2], Vishnu Kumarasamy[3], Tilini U. Wijeratne[1], Katharine L. Bunch[1], Sarvind Tripathi[1], Erik S. Knudsen[3], Agnieszka K. Witkiewicz[3], Kevan M. Shokat[2], Seth M. Rubin[1]*

The p27 protein is a canonical negative regulator of cell proliferation and acts primarily by inhibiting cyclin-dependent kinases (CDKs). Under some circumstances, p27 is associated with active CDK4, but no mechanism for activation has been described. We found that p27, when phosphorylated by tyrosine kinases, allosterically activated CDK4 in complex with cyclin D1 (CDK4-CycD1). Structural and biochemical data revealed that binding of phosphorylated p27 (phosp27) to CDK4 altered the kinase adenosine triphosphate site to promote phosphorylation of the retinoblastoma tumor suppressor protein (Rb) and other substrates. Surprisingly, purified and endogenous phosp27-CDK4-CycD1 complexes were insensitive to the CDK4-targeting drug palbociclib. Palbociclib instead primarily targeted monomeric CDK4 and CDK6 (CDK4/6) in breast tumor cells. Our data characterize phosp27-CDK4-CycD1 as an active Rb kinase that is refractory to clinically relevant CDK4/6 inhibitors.

Cyclin-dependent kinases 4 and 6 (CDK4/6) drive cell proliferation by partnering with D-type cyclins (CycD) to phosphorylate the retinoblastoma protein (Rb). Rb is subsequently hyperphosphorylated and inactivated by CDK2 to trigger passage through the $G_1$ phase of the cell cycle (*1*–*3*). Disruption of this CDK4/6-Rb signaling pathway is ubiquitous in tumors and typically occurs through overexpression of CycD1 or loss of the CDK4/6-specific inhibitor $p16^{INK4a}$, both of which increase CDK4/6 activity, leading to uncontrolled proliferation (*4*–*6*). Specific CDK4/6-targeting adenosine triphosphate (ATP)–competitive drugs such as palbociclib are approved for estrogen receptor–positive breast cancer and are being tested in clinical trials for application in diverse cancer types (*4*, *5*, *7*, *8*). As the use of CDK4/6 inhibitors as therapies increases, it becomes critical to understand their mechanism and factors that promote sensitivity or resistance.

CDK4/6 regulation is multilayered, reflecting the need to integrate diverse growth signals to control the cell cycle (*3*). Canonical CDKs require cyclin binding to properly structure their catalytic site (*9*). CDK4-CycD is unique in that CycD binding alone does not induce an active kinase structure (*10*, *11*). CDK4-CycD also has relatively fewer characterized substrates and poorer catalytic activity compared to other CDKs (*12*–*16*). There are several cofactors that interact with CDK4-CycD to modulate complex activity, assembly, and localization. The Rb family members (Rb, p107, and p130), which are the best-characterized substrates of CDK4, contain a specific activating interaction sequence (*15*, *17*, *18*). The Hsp90-Cdc37 chaperone complex stabilizes monomeric CDK4 by binding the unfolded N-lobe of the kinase (*19*, *20*). The INK4 family (p19, p18, p16, and p15) inhibits CDK4/6 by obstructing cyclin binding and by pulling the activation segment (also called the T-loop) into an inactive conformation (*21*, *22*). Activation segment phosphorylation by the CDK-activating kinase (CAK) stimulates CDK4 activity (*23*, *24*).

CIP (p21) and KIP (p27 and p57) proteins are CDK2 inhibitors in vitro and in cells under conditions of growth arrest (*25*). They are intrinsically disordered proteins that fold onto a cyclin and then a CDK sequentially to form ternary complexes (*26*). Mice lacking p21 or p27 are susceptible to tumorigenesis (*27*, *28*), which is consistent with the important roles of CIP and KIP proteins in negatively regulating the cell cycle through CDK2 inhibition. p27 degradation is critical for licensing entry into S phase, and p21 is a key effector of p53-activated senescence (*25*, *29*). p27 directly inhibits CDK2-CycA by occluding a substrate-docking site and by inserting a small helix within the p27 CDK-inhibitory domain into the CDK2 ATP site (*30*).

p21 and p27 have a more complex role in regulating CDK4. Although they can inhibit CDK4 under some conditions, they are also necessary for CDK4 activity. Embryonic fibroblasts that lack both p21 and p27 fail to assemble active CDK4-CycD complexes (*31*). Much of p27 is found in a complex with CDK4-CycD in proliferating cells, and active CDK4 complexes in cells contain both CycD and p27 (*25*, *32*–*36*). High levels of p21 are inhibitory, whereas low levels induce assembly and nuclear localization of enzymatically active CDK4 complexes (*37*). The activity of CDK4 complexes requires phosphorylation of p27 by nonreceptor tyrosine kinases (NRTKs) (*34*, *35*, *38*), including the breast tumor kinase Brk (also called PTK6). However, it is unclear whether and how p21 and p27 directly stimulate CDK4 catalytic activity, how this activation is mediated by p27 phosphorylation, and how p27 influences CDK4's sensitivity to chemical inhibitors such as palbociclib.

## Crystal structures of p21-CDK4-CycD1 and p27-CDK4-CycD1 complexes

To better understand p21 and p27 regulation of CDK4, we determined the crystal structures of p21-CDK4-CycD1 and p27-CDK4-CycD1 complexes at 3.2 Å and 2.3 Å resolution, respectively (Fig. 1 and tables S1 and S2). p21 and p27 similarly fold into a single helix that spans CDK4-CycD1. The structures demonstrate why both proteins function as assembly factors. p21 and p27 contain a subdomain 1 (D1), which docks into a hydrophobic cleft in CycD1, and a subdomain 2 (D2), which binds the N-lobe of CDK4 (Figs. 1 and 2). CDK4 and CycD1 are joined through the bridging helix (α1), which provides a rigid constraint to define the relative orientation of the cyclin and kinase N-lobe domains (Fig. 1, A and B).

The trimer structures demonstrate two key mechanisms of inhibition. First, p21 and p27 use their RxLF motif in D1 to block access to a critical substrate-docking site on CycD (Fig. 2, A and B) (*39*, *40*). Similar D1 domain–mediated interactions are also observed between p27 and CDK2-CycA (fig. S1, A and B). Second, the p21 and p27 D2 domains displace the first strand (β1) of the CDK4 N-lobe β sheet (Fig. 2C). As a result, the glycine-rich loop (Gly-loop; residues 13 to 19 in CDK4) is dislodged from the CDK4 active site and the ATP-binding pocket is disrupted. A striking difference in how p27 binds CDK2 and CDK4 was observed in the interaction between p27 D2 and the kinase domain. Whereas p27 inserts a small $3_{10}$ helix into the CDK2 ATP site (*30*), a $3_{10}$ helix was not observed at the corresponding location in CDK4 (Fig. 2D and fig. S1C). Differences in the CDK4 hinge region from CDK2 suggest that $3_{10}$-helix insertion of p27 into the CDK4 active site would be sterically hindered (fig. S1D).

## p21 and p27 binding rotates the CDK4 N-lobe and releases the kinase activation segment

We compared the trimer structures to the known CDK4-CycD dimer structures (*10*, *11*). Unexpectedly, p21 and p27 induced structural changes that better shape the active site for

[1]Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064, USA. [2]Department of Cellular and Molecular Pharmacology and Howard Hughes Medical Institute, University of California, San Francisco, CA 94158, USA. [3]Center for Personalized Medicine, Roswell Park Cancer Center, Buffalo, NY 14263, USA.
*Corresponding author. Email: srubin@ucsc.edu

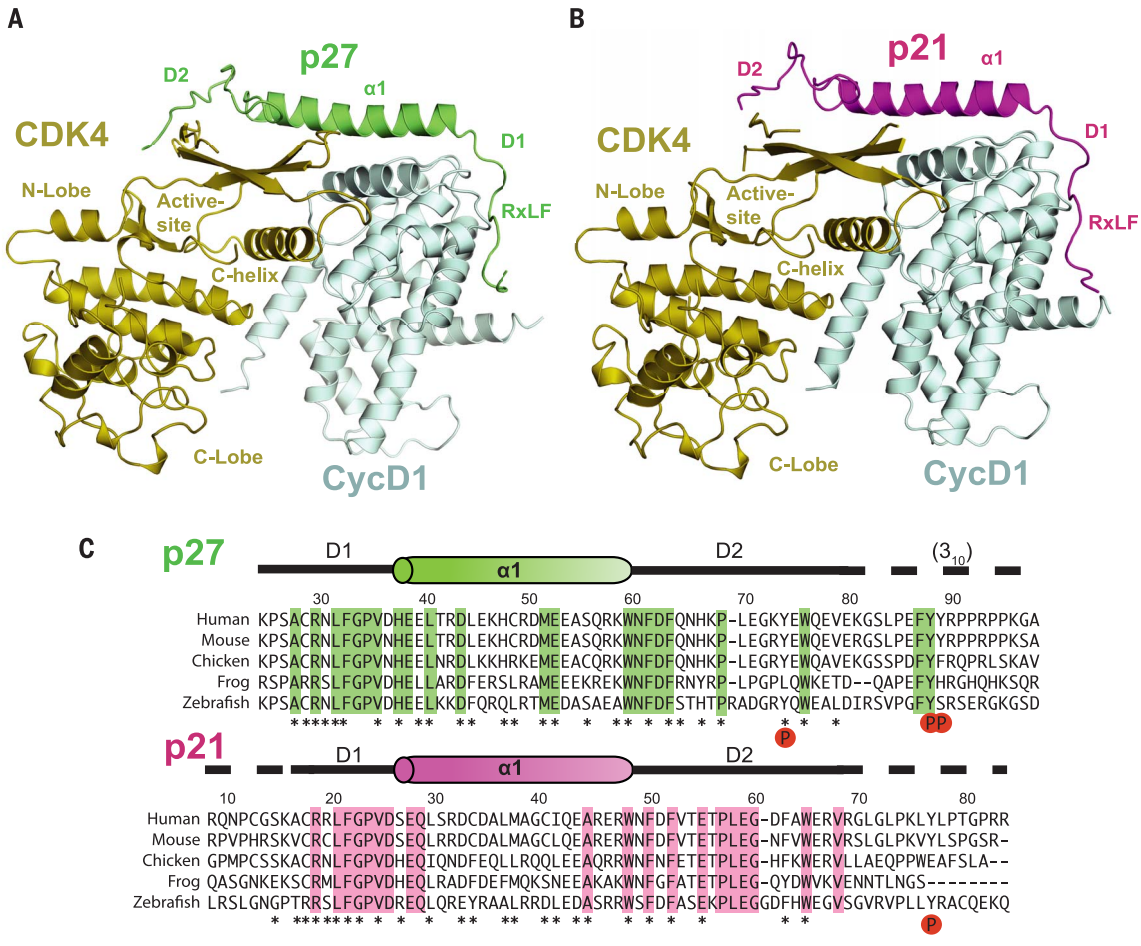



**Fig. 1. Structures of the p27-CDK4-CycD1 and p21-CDK4-CycD1 complexes.** (**A**) Overall structure of p27-CDK4-CycD1. p27 (green) binds CycD1 (cyan) with its D1 domain and CDK4 (gold) with its D2 domain. (**B**) Structure of p21-CDK4-CycD1. p21 (magenta) adopts a fold similar to that of p27, bridging CDK4 (gold) and CycD1 (cyan). (**C**) Sequence alignment of p27 and p21. Asterisks represent residues directly interacting with CDK4 or CycD1. The known tyrosine phosphorylation sites are noted. Secondary structure observed in the crystal is indicated above the sequences. Dashed lines indicate sequences in the crystallized protein that are not visible in the electron density, including the C-terminal sequence in p27 that forms a $3_{10}$ helix when bound to CDK2 (in parentheses). Amino acid abbreviations: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

catalysis. p27 binding rotated the N-lobe of CDK4 relative to the C-lobe (Fig. 3, A to D). The strands in the N-lobe β sheet were shifted by ~4 to 6 Å such that the β-strand 2 (β2) of the trimer complex replaced the β-strand 3 (β3) position of the dimer, β3 replaced β5, and β5 replaced β4. An important consequence of this altered N-lobe conformation is that the activation segment was released from the CDK4 active site, because specific interactions between β3 and β5 and the activation segment helix were broken (Fig. 3B). The release represents an important step that allows substrate binding in the kinase active site (*41*). For CDK1 and CDK2, activation segment release occurs upon binding to a cognate cyclin, and phosphorylation of the activation segment by CAK further improves substrate binding (*42*). However, in the crystal structure of the CDK4-CycD1 dimer, the activation segment remains in a substrate-blocking conformation despite the introduction of a phosphomimetic (T172D) at the CAK phosphorylation site (Fig. 3B) (*10*). The release of the activation segment observed upon p27 binding (Fig. 3B) is a unique mechanism to CDK4 and likely explains the requirement of p27 for CDK4 activation segment phosphorylation by CAK (*24*).

A second result of the β-sheet rearrangement was that the catalytic lysine ($Lys^{35}$) on β3 was pulled into a position to accept the β- and γ-phosphates of ATP (Fig. 3, C and D). This position of $Lys^{35}$ is similar to that of the corresponding catalytic lysine ($Lys^{33}$) in active CDK2-CycA (fig. S1E) (*9*, *30*). In contrast, the position of $Lys^{35}$ in the CDK4-CycD1 dimer is similar to its position in the inactive CDK2 monomer structure (*10*, *11*). This conformational change in the N-lobe does not occur when p27 binds CDK2 and is therefore an allosteric activating mechanism of p27 specific to CDK4 (fig. S1, A and F) (*30*).

Although the p21 and p27 trimer complexes are primed for catalysis, in that $Lys^{35}$ is positioned to coordinate ATP and the activation segment is released, there are still aspects of the structure that indicate a requirement for additional activating mechanisms. The D2 domain must be displaced to permit formation of the ATP-binding G-loop. In addition, comparison of the CDK4 trimer structure with an active CDK2 structure indicates that the C helix (also known as the PSTAIRE helix) in CDK4 remains in an inactive position, with its critical catalytic glutamate pointing away from the active site (Figs. 2C and 3D). We propose that the structure determined here is an intermediate along a pathway to activation, and that substrate binding, ATP binding, or activation segment phosphorylation in conjunction with p27 binding may induce the required rotation of the C helix (fig. S2A).

### Tyrosine phosphorylation of p27 but not p21 activates CDK4 trimer complexes

To quantify the kinase activity of p21- and p27-CDK4-CycD1, we assayed the activity of recombinant CDK4 complexes (Fig. 3, E to G). We used a phosphorylation site mimetic in the CDK4 activation segment (CDK4 T172E) because there was heterogeneity in phosphorylation of the $Thr^{172}$ site in our purified protein (fig. S2, B and C). We tested the CDK4-CycD1 dimer (called K4D1/– because it lacks p21 or p27), the trimer with p21 or p27 (K4D1/p21 or

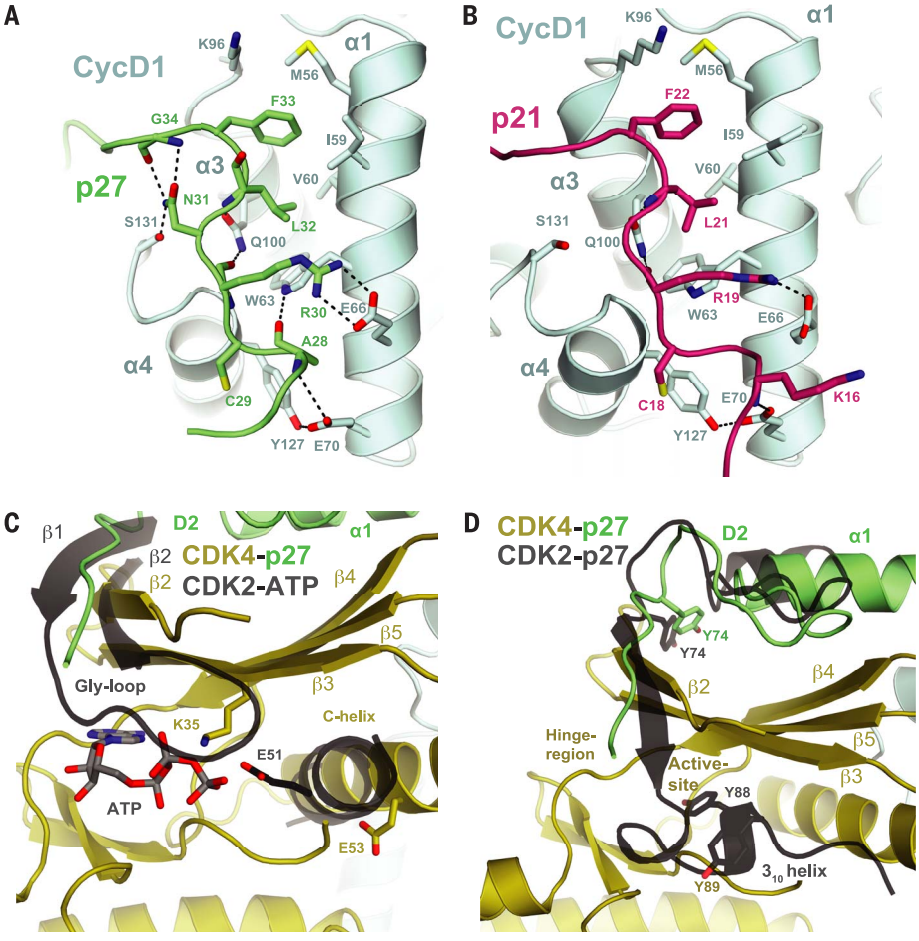



**Fig. 2. p27 and p21 inhibit substrate binding and catalytic activity.** (**A**) Association between the p27 RxLF motif (green) and the MVRIL cleft in CycD1 (cyan) competes for substrate docking. (**B**) The p21 RxLF (magenta) bound to CycD1 (cyan). (**C**) Binding of the D2 region in p27 (green) displaces the β1 strand in the CDK4 N-lobe (gold), disrupting the ATP-binding site. The structure shown in gray for comparison, including the ATP, is from CDK2 in the active CDK2-CycA dimer. The CDK4 β1 strand and following Gly loop are not visible in the p27-CDK4-CycD1 trimer structure, indicating they are disordered. The C helix remains in an inactive conformation in the CDK4-p27 structure. (**D**) Comparison of binding of p27 to CDK2 (p27 in gray) and to CDK4 (p27 in green, CDK4 in gold). The $3_{10}$ helix in p27 that binds the CDK2 ATP site is not visible and likely remains disordered upon CDK4 binding. Tyrosine phosphorylation sites are shown.

K4D1/p27), and the trimer with tyrosine kinase (Brk/PTK6)–phosphorylated p21 and p27 (K4D1/phosp21 or K4D1/phosp27) (Fig. 3E and fig. S3). The dimer demonstrated strong activity toward the purified Rb C-terminal domain (residues 771 to 928, $Rb^{771\text{-}928}$), whereas the trimer with unphosphorylated p21 or unphosphorylated p27 demonstrated poor activity. Consistent with observations in cell culture (*34*, *35*), tyrosine phosphorylation of p27 restored its recombinant trimer enzyme activity. The tyrosine kinase–phosphorylated p21 had only slight additional activity, indicating that p21 trimer complexes are mostly inhibited despite tyrosine phosphorylation.

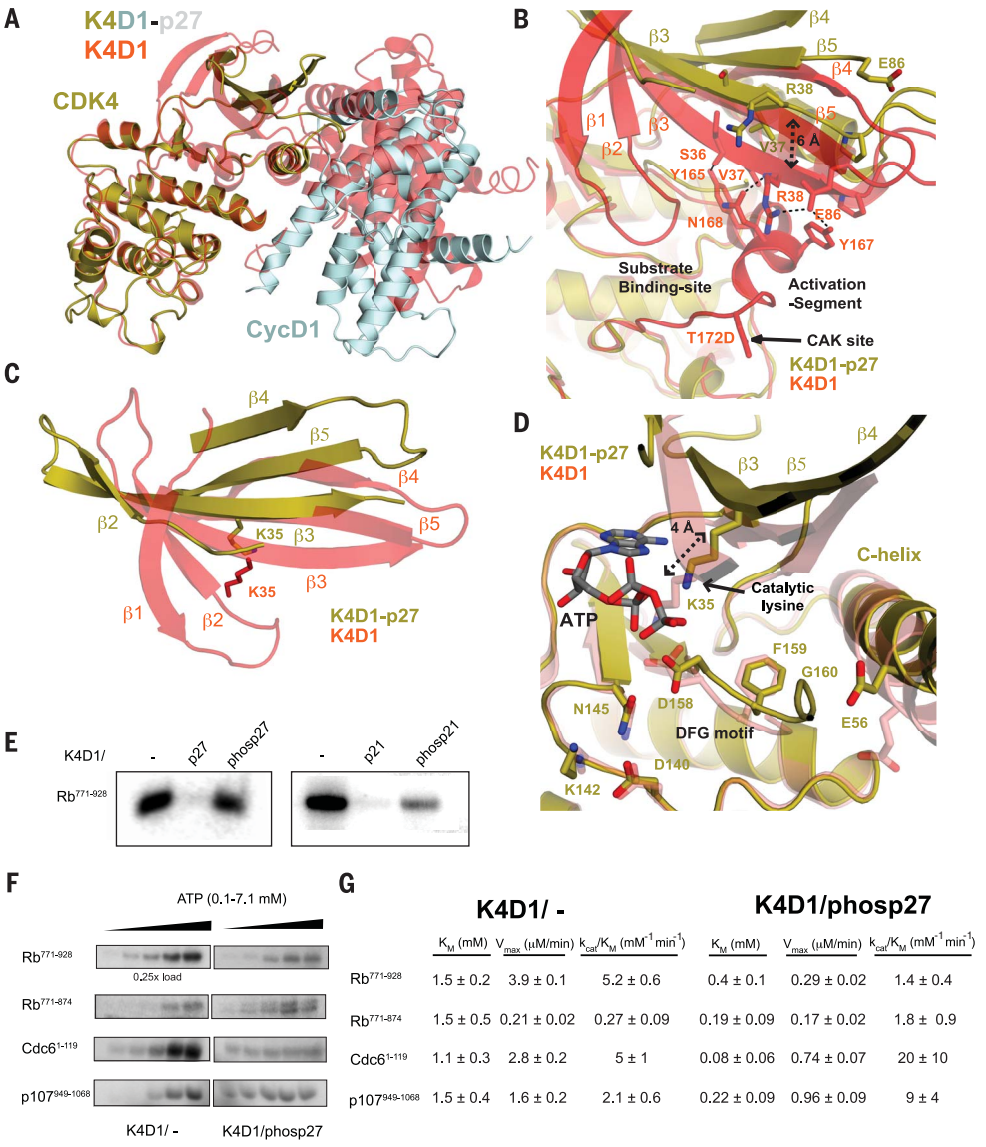
We focused on the strong activity of the phosphorylated p27 trimer and used steady-state kinetic experiments to quantify enzyme processing of ATP (Fig. 3, F and G, and fig. S4). We tested CDK4 activity of the dimer and the phosp27-trimer complex toward $Rb^{771\text{-}928}$, $Rb^{771\text{-}874}$ (residues 771 to 784; this shorter Rb fragment lacks a CDK4 docking sequence) (*15*, *17*, *18*), $Cdc6^{1\text{-}119}$ (residues 1 to 119, a CDK4 substrate that contains an RxLF docking sequence) (*13*), and $p107^{949\text{-}1068}$ (residues 949 to 1068, which lacks any known docking sequence). In the case of all four substrates, the $K_M$ of ATP for the K4D1/phosp27 trimer was reduced relative to that of K4D1/– and was more similar to the $K_M$ of CDK2 and that of most serine/threonine kinases (*16*). This decrease in $K_M$, which we also observed for CDK6-CycD1 but not for CDK2-CycA (fig. S5), is consistent with our structural results that p27 binding induces an N-lobe conformation in CDK4 that supports ATP coordination. We note that although we observed activity from reconstituted phosp27-CDK6-CycD1 trimers in vitro, it is unclear whether these complexes exist in cycling cells (*43*).

In the activity assays, phosp27 also had an inhibitory effect that is specific for substrates containing a CDK docking sequence. When assaying the K4D1/– dimer, the ATP -dependent maximum reaction rate $V_{max}$ and the Rb substrate $k_{cat}/K_M$ both decreased upon deletion of the docking sequence in Rb (compare $Rb^{771\text{-}928}$ to $Rb^{771\text{-}874}$), supporting the importance of docking in dimer phosphorylation of Rb (Fig. 3G and fig. S4B) (*17*, *18*). The ATP $V_{max}$ decreased upon addition of phosp27 when assaying $Rb^{771\text{-}928}$ (factor of 13 decrease) and $Cdc6^{1\text{-}119}$ (factor of 4 decrease), which both contain docking sequences (Fig. 3G). In contrast, the ATP $V_{max}$ only modestly decreased for $p107^{949\text{-}1068}$ and for $Rb^{771\text{-}874}$, which lack docking sequences. The $Rb^{771\text{-}928}$ substrate $k_{cat}/K_M$, but not the $Rb^{771\text{-}874}$ substrate $k_{cat}/K_M$, decreased upon addition of p27 (fig. S4B), and the kinetics of the trimer activity toward both Rb substrates was similar (Fig. 3G). These observations indicate that even in the active complex, p27 inhibits CDK4 activity toward some substrates by occluding the docking site in CycD (Fig. 2A).

### Structural mechanism of p27-CDK4-CycD activation by tyrosine phosphorylation

We solved the structure of a phosp27-CDK4-CycD1 complex to determine how tyrosine phosphorylation relieves p27 inhibition (table S1). $Tyr^{74}$ was in a position similar to that in the unphosphorylated trimer, but there was clear electron density for the phosphate (Fig. 4, A and B, and fig. S6). $Tyr^{74}$ is part of an aromatic cluster in p27 D2 that binds the CDK4 N-lobe. The structural data predict that $Tyr^{74}$ phosphorylation would weaken this association by destabilizing the hydrophobic interface between $Tyr^{74}$ and the N-lobe of CDK4 (Fig. 4, A and B). In p21, a phenylalanine ($Phe^{62}$) is in the position of $Tyr^{74}$ in p27 such that the p21 D2 domain would remain bound without the addition of a phosphate (Figs. 1C and 4C).

In a structure we solved of the trimer containing p27 with phosphomimetics (table S1), temperature B-factors were ~10 to 20 $Å^2$ higher for side chains in the aromatic cluster in D2 and the electron density was weaker around $Trp^{60}$ and $Glu^{74}$ (fig. S6). The higher B-factors and weaker density are consistent with lower D2 occupancy and higher disorder. Deletion of the D2 domain from p27 or mutation of the tyrosines to glutamate phosphomimetics generated a trimer complex that phosphorylated $Rb^{771\text{-}928}$ (Fig. 4D and fig. S7). Solution studies with p27 binding to CDK2-CycA demonstrate that phosphorylation of $Tyr^{74}$ results in loss of interactions between the D2 domain of p27




G

| | K4D1/ - | | | K4D1/phosp27 | | |
|---|---|---|---|---|---|---|
| | $K_M$ (mM) | $V_{max}$ (μM/min) | $k_{cat}/K_M$ ($mM^{-1}min^{-1}$) | $K_M$ (mM) | $V_{max}$ (μM/min) | $k_{cat}/K_M$ ($mM^{-1}min^{-1}$) |
| $Rb^{771-928}$ | 1.5 ± 0.2 | 3.9 ± 0.1 | 5.2 ± 0.6 | 0.4 ± 0.1 | 0.29 ± 0.02 | 1.4 ± 0.4 |
| $Rb^{771-874}$ | 1.5 ± 0.5 | 0.21 ± 0.02 | 0.27 ± 0.09 | 0.19 ± 0.09 | 0.17 ± 0.02 | 1.8 ± 0.9 |
| $Cdc6^{1-119}$ | 1.1 ± 0.3 | 2.8 ± 0.2 | 5 ± 1 | 0.08 ± 0.06 | 0.74 ± 0.07 | 20 ± 10 |
| $p107^{949-1068}$ | 1.5 ± 0.4 | 1.6 ± 0.2 | 2.1 ± 0.6 | 0.22 ± 0.09 | 0.96 ± 0.09 | 9 ± 4 |

**Fig. 3. p27 induces structural changes that promote ATP coordination and processing.** (**A** to **D**) Structural alignment of CDK4-CycD1 with p27 (gold-cyan) and without p27 (red, PDB code 2W96) reveals movement of both the CDK4 N-lobe and CycD1 domains relative to the CDK4 C-lobe. (**E**) Phosphorylation of purified $Rb^{771-928}$ with [$^{32}$P]ATP and the indicated dimer (K4D1/−) or trimer complex. (**F** and **G**) Steady-state kinase assays measuring effects of ATP concentration on initial reaction rate, as measured by incorporation of [$^{32}$P]ATP. Reactions include CDK4 dimer (K4D1/−) or active trimer (K4D1/phosp27) and the indicated substrate.

and the N-lobe of CycA (*44*). Together these results support a model in which phosphorylation or phosphomimetics weaken the affinity of the D2 domain for the CDK4 N-lobe, allowing for the formation of the Gly loop and activation of the kinase. The activation segment was not homogeneously phosphorylated in our structure (fig. S2B), so we cannot rule out the possibility that activation segment phosphorylation and the ultimate repositioning of the C helix also play a role in removing the D2 domain from the N-lobe.

$Tyr^{88}$ and $Tyr^{89}$ were disordered in all the p27-CDK4-CycD1 crystal structures, and the CDK4 trimer assembled with p27 containing a phosphomimetic at $Tyr^{74}$ had similar (although slightly less) activity to that of the enzyme assembled with phosphomimetics at all the tyrosines (Fig. 4D). Thus, $Tyr^{74}$ phosphorylation appeared to be nearly sufficient for p27 activation of CDK4, although the data indicate that there may have been some contribution from $Tyr^{88}$/$Tyr^{89}$ phosphomimetics in the p27 triple glutamate mutant (Fig. 4D). This mechanism for turning off p27 inhibition is different from the mechanism in CDK2, in which phosphorylation of p27 $Tyr^{88}$ and $Tyr^{89}$ is necessary for ejection of a p27 $3_{10}$ helix from the catalytic site (fig. S1C) (*45*). These phosphorylation sites in p27 and the corresponding site in p21 ($Tyr^{77}$) are located in a sequence that remains disordered upon CDK4 binding (Fig. 1C), which is consistent with their weaker role in the mechanism of CDK4 activation. The more important role of the $Tyr^{74}$ phosphorylation site also explains the observed poor activity of Brk-phosphorylated p21 complexes (Fig. 3E), as p21 lacks a tyrosine at the equivalent position of $Tyr^{74}$ in p27 (Figs. 1C and 4C). We conclude that although p21, like p27, primes CDK4 for catalysis by releasing the activation segment, p21 tyrosine phosphorylation does not relieve D2 inhibition, and the complex remains mostly inhibited.

### p27-CDK4-CycD1 complexes are insensitive to CDK4/6-specific inhibitors

The structural changes in the kinase N-lobe that are induced by p27 binding indicate that palbociclib-type inhibitors should have poor potency toward active CDK4 trimer complexes (Fig. 5A). The "gatekeeper" residue $Phe^{93}$ in β5 and also $Ala^{33}$ in β3 are rotated away from making contacts with the C5-methyl and C6-acetyl groups in the pyridopyrimidine scaffold of palbociclib (*46*, *47*). We tested the effects of palbociclib, ribociclib, and abemaciclib on the activity of the reconstituted CDK4 enzyme complexes (Fig. 5B and fig. S8A). All three compounds inhibited $Rb^{771-928}$ phosphorylation by CDK4-CycD1 dimer (K4D1/−), as expected; however, the active phosp27-CDK4-CycD1 trimer (K4D1/phosp27) was relatively insensitive

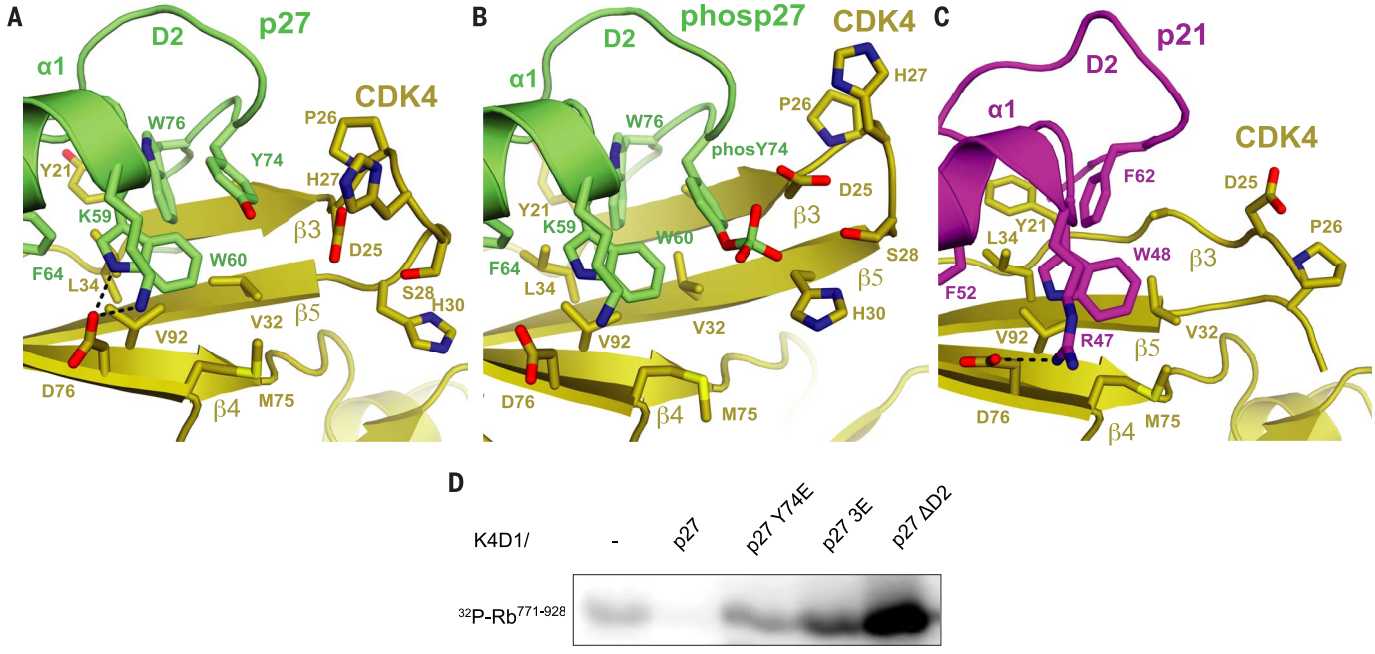



**Fig. 4. Tyr$^{74}$ phosphorylation disrupts the p27 D2-CDK4 interface.** (**A** to **C**) Comparison of CDK4-CycD1 structures with unphosphorylated p27 (A), phosphorylated p27 (B), and p21 (C). (**D**) [$^{32}$P]ATP phosphorylation of Rb$^{771-928}$ using K4D1 dimer or trimer enzymes assembled with the indicated p27 kinase inhibitory domain construct (residues 25 to 93). 3E contains three glutamate phosphomimetics at Tyr$^{74}$, Tyr$^{88}$, and Tyr$^{89}$. ΔD2 contains p27 residues 25 to 60, and therefore lacks the D2 CDK4-binding domain.

to the drugs, with apparent inhibition constants of ~10 μM or higher. We measured the binding affinity of palbociclib for CDK4 complexes by isothermal titration calorimetry (ITC). The drug bound only CDK4 monomer and K4D1/– dimer tightly. Consistent with our structural and kinetic data, we detected no affinity for K4D1/phosp27 trimer (Fig. 5C and fig. S8B). We also measured binding of p27 to CDK4 monomer and found no affinity if the enzyme was first saturated with palbociclib (Fig. 5C and fig. S8C). These data demonstrate that binding of p27 and of palbociclib to CDK4 are mutually exclusive.

### Palbociclib does not inhibit endogenous CDK4 activity

We tested whether the activity of endogenous cellular CDK4 complexes is inhibited by palbociclib. We immunoprecipitated p27 complexes from MCF7, MDA-MB-231, and T98G cells. MCF7 and MDA-MB-231 cells are Rb-positive and palbociclib-sensitive breast cancer cells that differ in estrogen receptor (ER) status (*46*). T98G cells are Rb (+) and ER (–) glioma cells that are relatively less sensitive to palbociclib (see below). The immunoprecipitates phosphorylated purified Rb$^{771-928}$ in the [$^{32}$P]ATP assay; however, the activity was insensitive to palbociclib that was added to the kinase reactions (Fig. 5D). We considered that this palbociclib-insensitive activity might be from CDK2. Although cellular CDK2 complexes with p27 are thought to be inactive (*32*), we found that reconstituted K2A/phosp27 complexes phosphorylated Rb (fig. S5C). However, the p27 immunoprecipitate activity was also not inhibited by the CDK2 inhibitor dinaciclib, and therefore the immunoprecipitate likely contained a CDK4 activity and not a CDK2 activity (fig. S9).

To test the entire cellular pool of CDK4 activity, we immunoprecipitated CDK4 from the same cell lines using an antiserum raised against the CDK4 C terminus, which contains a sequence distinct from CDK2 (*48*). Palbociclib also poorly inhibited the immunoprecipitated CDK4 kinase activity (Fig. 5E), which is consistent with previous observations that CDK4 activity arises from trimer complexes (*25*, *31*–*37*) and our observation that the trimer is insensitive to the drug (Fig. 5B).

### An ATP-site occupancy probe targets CDK4 monomers in cells

To examine the association of an inhibitor with endogenous CDK4 complexes, we treated MCF7 and MDA-MB-231 cells with the covalent ATP-site occupancy probe XO44 (*49*). XO44 labeled both CDK4 and CDK2 in these cells and could be detected by a gel shift upon coupling to an analysis probe (Fig. 6A and fig. S10, A and B). Although XO44 labeled most CDK2 and the mitotic kinase Aurora B (fig. S10A), it labeled only ~40 to 60% of CDK4 even at high concentrations, indicating that a resistant population of the protein exists. CDK4 labeling in cells was inhibited by palbociclib (Fig. 6B and fig. S10C); this finding shows that XO44 and palbociclib bind the ATP site competitively and that the population of CDK4 that is targeted by XO44 is also targeted by palbociclib-type inhibitors. XO44 labeled recombinant CDK4 efficiently in dimer complexes but not in unphosphorylated or phosphorylated trimer complexes (Fig. 6C).

We treated cells with XO44 and palbociclib and fractionated the lysates with size-exclusion chromatography (Fig. 6D and fig. S10D). We observed CDK4 in high–molecular size fractions comigrating with Hsp90 and Cdc37 (e.g., fractions 5 and 6), in mid-size fractions with CycD1 and p27 (fractions 9 and 10), and in low-size fractions that likely contain monomers (fractions 15 and 16) (*34*, *36*). Consistent with ATP-competitive probes sequestering obligate Hsp90-client kinases from the chaperone system (*20*, *50*), we observed a reduction in the abundance of large-sized, putatively Hsp90/Cdc37-bound CDK4 in extracts from XO44- and palbociclib-treated cells. We also detected greater quantities of CDK4 in small-sized monomer fractions upon treatment with either compound, and only the CDK4 population observed in these fractions was labeled with XO44 (Fig. 6D, fractions 15 and 16, XO-44 + click). Considering that CDK4 inhibitors do not associate with CDK4 bound by INK proteins (*51*), we propose that the tested inhibitors primarily

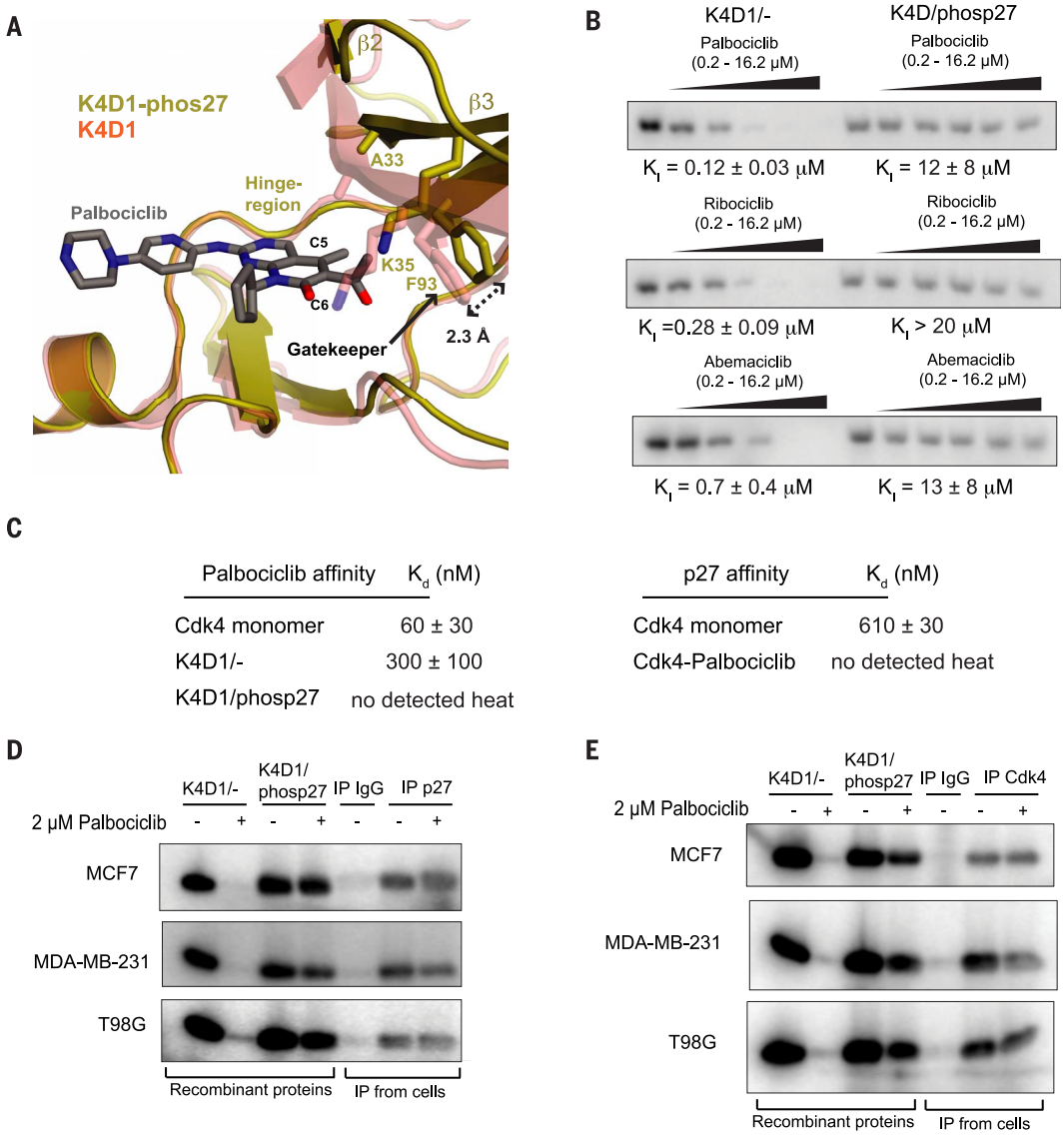



**Fig. 5. Palbociclib does not bind and poorly inhibits purified and endogenous CDK4 trimer complexes.** (**A**) K4D1 dimer (PDB ID 2W96) and the phosp27-K4D1 trimer structures were aligned with palbociclib-bound CDK6 (PDB ID 2EUF, not shown) to model the position and interactions of the drug when bound to CDK4. (**B**) [$^{32}$P]ATP phosphorylation of Rb$^{771\text{-}928}$ using CDK4-CycD1 dimer (K4D1/−) and phosp27-CDK4-CycD1 trimer (K4D1/phosp27) enzymes in the absence (leftmost lane in each titration) and presence of increasing inhibitor concentrations (see fig. S8A for quantification). Each drug is dosed from 0.2 μM to 16.2 μM in factor of 3 increments. (**C**) ITC affinities for palbociclib (left) or p27 (right) titrated into the indicated enzyme. (**D**) The indicated cell lysates were immunoprecipitated with control or with p27 antibody, and the activity of the immunoprecipitate was used to phosphorylate Rb$^{771\text{-}928}$ with [$^{32}$P]ATP in the absence or presence of palbociclib. Reactions with the indicated recombinant dimer (K4D1/−) or trimer (K4D1/phosp27) enzymes are shown for comparison in the first four lanes. (**E**) As in (D), except lysates were precipitated with antiserum raised against a CDK4 C-terminal peptide.

bind CDK4 monomers in cells and increase the abundance of the inactive monomer population.

We synthesized a palbociclib analog linked to biotin, and we used it to precipitate CDK4/6 from cell extracts to determine which proteins exist in the complexes targeted by palbociclib-type inhibitors (Fig. 7A). Consistent with our activity and ITC binding assays, the palbociclib-biotin precipitated recombinant CDK4-CycD1 dimer but not recombinant p27-CDK4-CycD1 trimer (Fig. 7B). Strikingly, in breast cancer cell lysates, the palbociclib-biotin strongly precipitated CDK4 and CDK6 but did not precipitate any detectable p21 or p27 (Fig. 7, C and D). This observation further demonstrates that the drug does not bind trimer complexes with p21 and p27 in these cells. A faint band reproducibly appeared in the CycD1 blot, indicating that there may be some small population of dimer complexes targeted by the drug. However, we conclude that monomer CDK4 and monomer CDK6 are the predominant targets of palbociclib in these cells.

### Palbociclib-induced cell cycle arrest coincides with loss of CDK2 but not CDK4 activity

We treated cells with palbociclib for 48 hours, which, similar to previous descriptions, led to strong arrest of the breast cancer cells and to weaker arrest of T98G cells, as determined by 5-ethynyl-2-deoxyuridine (EdU) incorporation and propidium iodide staining (*52*, *53*) (Fig. 8A and fig. S11). Endogenous CDK4 and CycD1 complexes retained kinase activity despite the cells arresting with drug treatment (Fig. 8B and fig. S12A). This similar activity in palbociclib-treated cells was observed despite increased levels of CycD1 (Fig. 8C and fig. S12B) (*52*). In contrast, CDK4 complexes immunoprecipitated from serum-starved NIH/3T3 cells lack Rb-directed kinase activity (fig. S12C), and the lack of activity coincides with reduced CycD1 levels. In the breast cancer cells tested, we did not observe palbociclib inhibiting CDK4- or CycD1-associated kinase activity, which is consistent with the lack of susceptible dimer complexes that could be detected in their lysates (Fig. 6D and Fig. 7, C and D).

Unlike CDK4 complexes, CDK2 complexes that were immunoprecipitated from the breast cancer cells had diminished activity after 48 hours of palbociclib treatment (Fig. 8B). CDK2 inhibition at this drug concentration is likely indirect (*5*, *46*), resulting at least in part from lower levels of CycA (Fig. 8C) (*52*). We also observed evidence for an increase in p21 but not p27 in CDK2-CycE complexes (Fig. 8, C and D, and fig. S12, D and E). In reciprocal coimmunoprecipitation experiments, additional p21-CycE association was detected after

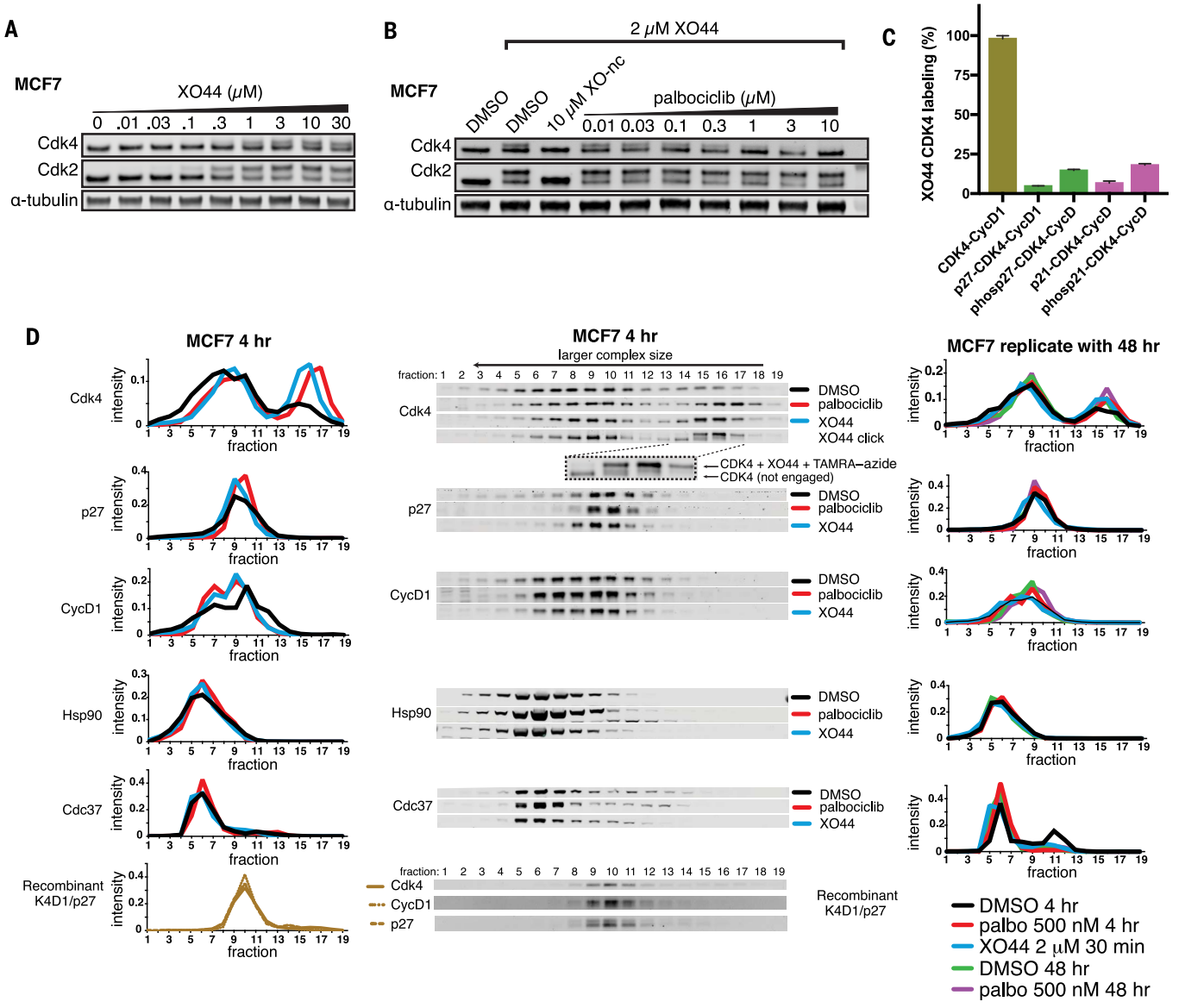



**Fig. 6. The ATP-site occupancy probe XO44 labels monomer CDK4 in MCF7 cells.** (**A**) Labeling of endogenous CDK4 and CDK2 by the promiscuous covalent ATP-site probe XO44. MCF7 cells were treated with DMSO vehicle or XO44 at the indicated concentrations for 30 min. Lysates were subjected to click reaction with TAMRA-azide to visualize XO44 labeling of proteins by gel mobility shift. (**B**) Palbociclib competes with XO44 for CDK4 binding but not CDK2 binding. Experiment performed as in (A), but cells were pretreated for 60 min with DMSO, a nonclickable analog of XO44 (XO-nc), or increasing concentrations of palbociclib. (**C**) XO44 efficiently labels purified recombinant CDK4-CycD1 dimer but not trimer complexes with p21 or p27, as determined by electrospray ionization mass spectrometry. Protein complexes were treated with DMSO or XO44. Average percent labeling was determined for three replicates; error bars denote SD. (**D**) Asynchronous MCF7 cells were treated with DMSO, XO44 (2 μM, 30 min) or palbociclib (500 nM, 4 hours), and lysates were fractionated using Superdex200 size-exclusion chromatography. XO44 labeling was monitored by gel mobility shift after click reaction with TAMRA-azide ("XO44 click"). Normalized quantifications of band signals from the Western blots using the indicated antibody are shown. A benchmark chromatography experiment using the recombinant protein complex is displayed at the bottom of the panel. See fig. S10 for data using MDA-MB-231 cells.

48 hours. We could not detect p27 in the CycE1 IP with or without palbociclib treatment, so we cannot determine whether p27 also plays a role in down-regulating the CDK2-CycE activity. We conclude that palbociclib treatment indirectly leads to CDK2 inhibition and that this CDK2 inhibition, and not CDK4 inhibition, correlates with cell cycle arrest.

## Discussion

CIP/KIP proteins were first characterized as CDK inhibitors, in particular as potent inhibitors of CDK2-CycA/E (*25*, *54*–*58*). In contrast, other evidence implicated noninhibitory roles for p21 and p27 in mediating CDK4-CycD function, including complex assembly and nuclear localization (*31*, *37*). Moreover, the observations that active CDK4-CycD tolerated the presence of p27 and that Rb-directed kinase activity in cells contained p27 led to models that CDK4-CycD may titrate inhibitory p27 away from CDK2 (*25*, *33*, *37*, *59*). Our data demonstrate that p27 is not merely a noninhibitor, but in fact allosterically activates CDK4-CycD, remodeling the kinase to increase the catalytic

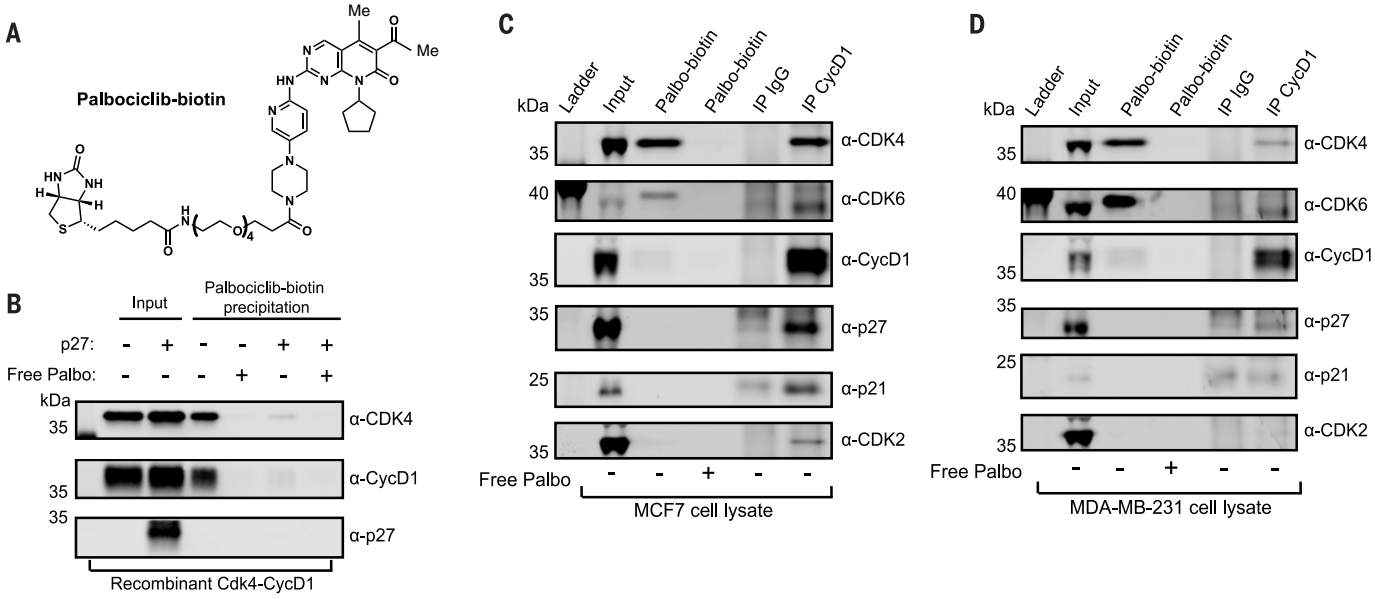



**Fig. 7. Palbociclib directly targets CDK4/6 monomer in cell lysate.** (**A**) Chemical structure of palbociclib-biotin synthesized here. (**B**) Palbociclib-biotin precipitated purified recombinant CDK4-CycD1 dimer, but not p27-CDK4-CycD1 trimer. (**C** and **D**) MCF7 or MDA-MB-231 cell lysates were used in a precipitation reaction with palbociclib-biotin. A CycD1 antibody was also used to precipitate CDK4/6 from cell lysates to compare the total pool of CycD1 complexes to palbociclib-bound complexes.

efficiency of ATP processing. Our structural data demonstrate that this effect is specific to CDK4 (versus CDK2) and is primarily induced by $Tyr^{74}$ phosphorylation in p27. Although p21 also acts as an assembly factor for active CDK4-CycD1 (*31*), p21 contains a phenylalanine at a position equivalent to that of $Tyr^{74}$ (Figs. 1C and 4C), and p21 trimer complexes cannot be activated by D2 displacement as in p27. As a result, phosphorylated p21 trimer complexes have relatively less activity than phosphorylated p27 complexes.

The catalytic efficiency of the phosp27-CDK4-CycD1 trimer was greater than the catalytic efficiency of the dimer for most of the substrates we investigated (Fig. 3G). The requirement of p27 for efficient phosphorylation may explain why fewer cell cycle substrates have been identified for CDK4/6 than for (e.g.) CDK2, and why it has been thought that Rb is a preferred CDK4/6 substrate (*12*, *14*, *15*, *17*, *18*). Additional CDK4/6 substrates have been identified more recently (*13*, *60*, *61*), and it will be important to explore the role of p27 in their regulation by CDK4/6 phosphorylation. Our observation that the CDK4-CycD1 dimer catalytic efficiency was dependent on the last ~50 amino acids of Rb is consistent with reports of a CDK4 docking site in the Rb C-terminal domain (*17*, *18*). Docking allows Rb to more tightly associate with the dimer and be phosphorylated even when ATP affinity is so low. In contrast, we observed that the trimer activity did not depend on a substrate docking sequence, which suggests that a role of p27 activation may be to broaden CDK4 substrate specificity. Our data also implicate a competitive association between Rb and p27 for CDK4-CycD and indicate that p27 still functions to temper Rb-directed CDK4 activity even in the context of active trimer complexes.

CDK4/6-specific inhibitors were developed through optimizing potency and selectivity toward CDK4-CycD dimer, despite evidence that the physiological complex contains CIP or KIP family inhibitors such as p27 (*25*, *31*–*37*, *46*). Our data demonstrate that active recombinant CDK4 complexes containing p27 are not sensitive to these inhibitors and that palbociclib does not associate with active endogenous p27 trimer complexes. Moreover, we did not detect inhibition of CDK4 activity in drug-arrested breast cancer cells, and we did not robustly detect CDK4-CycD1 dimer complexes in cell lysates. Still, we recognize that dimer inhibition may occur in some contexts and that some small population of dimer may be targeted even in the breast cancer cells tested here. In light of the differences between dimer and trimer response to CDK4/6 palbociclib-type inhibitors in vitro, it will be important to explore how the balance of dimer and trimer complexes may influence sensitivity to drugs in cells or in vivo.

Our data implicate an additional mechanism toward cell cycle arrest in which palbociclib associates with CDK4 monomer instead of binding and inhibiting CDK4 dimer or trimer activity. One outcome is that the population of CDK4 monomer increases, and it is possible that this CDK4-palbociclib complex is recognized and induces a response, such as the decrease in CycA abundance (*52*) or an increase of p21 in CDK2-CycE complexes. Both of these outcomes correlate with our observed lower CDK2 activity. A reported CDK4/6 degrader drug is not as effective as palbociclib in inhibiting Rb phosphorylation (*62*), and, a recent study finds that cells expressing a catalytically inactive CDK4 are arrested when treated with palbociclib (*63*). These results perhaps reflect the fact that palbociclib-induced cell cycle arrest does not require inhibition of CDK4 activity but alternatively results from indirect inhibition of CDK2. This indirect role of CDK4 in modulating CDK2 activity is consistent with observations that CDK4-CycD expression can activate CDK2 through sequestration of p27 and without the requirement of CDK4 catalytic activity (*25*, *58*).

We note the similarity of this mechanism of palbociclib inhibition to that of the p16-INK family of CDK4 protein inhibitors, which sequester inactive monomer CDK4 (Fig. 8E). The ability of palbociclib to mimic p16 may in part explain why cells lacking p16 are often most sensitive to palbociclib-type inhibitors (*52*, *53*, *64*, *65*). We suggest that, similar to the downstream effects of p16, accumulation of CDK4 monomer upon palbociclib treatment indirectly lowers CDK2 activity, potentially through the shuttling of p21 to CDK2 complexes (*25*, *36*, *66*, *67*). In cycling cells, this drug-induced reorganization of CDK complexes may occur when p21 is resynthesized after its degradation before S phase and/or when CycD1 is low in S phase and would lead to cell cycle arrest (*68*). The targeting of an inactive kinase, rather than inhibition of assembled kinase activity, may apply to

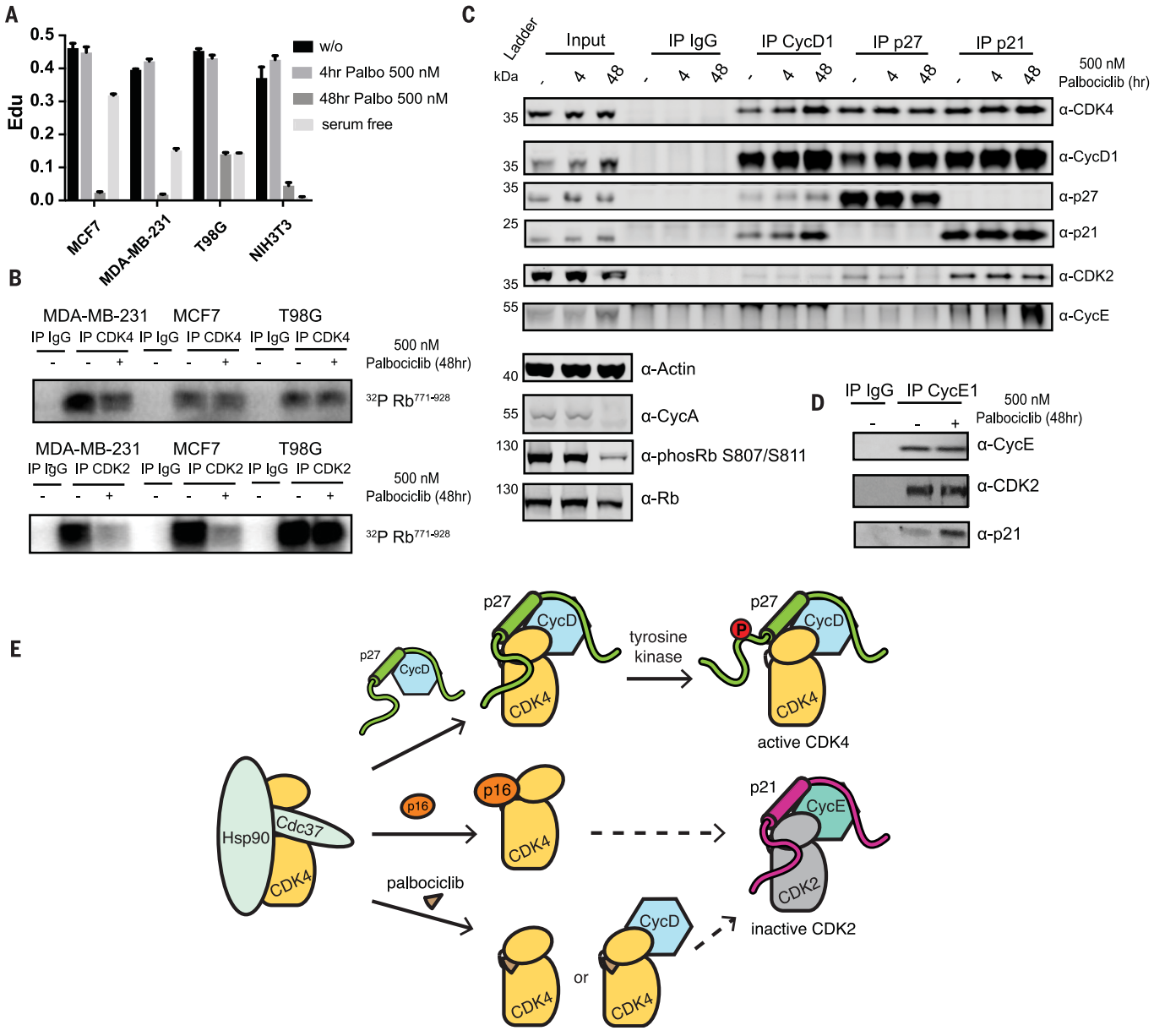



**Fig. 8. Palbociclib indirectly leads to down-regulation of CDK2 through p21.** (**A**) Cell cycle profiling of the cell lines used in this study upon treatment with 500 nM palbociclib for 4 or 48 hours or without drug treatment. Cells were assayed for EdU incorporation as a marker of S phase at the different time points after drug treatment. The fraction of cells showing EdU staining is reported. EdU incorporation in cells that were serum-starved (serum free) for 48 hours is also shown for comparison. (**B**) Cells were treated with palbociclib as indicated for 48 hours, and lysates were immunoprecipitated with CDK4 antiserum or a CDK2 antibody. Activity of the immunoprecipitated complexes was assayed as in Fig. 5, D and E. (**C**) CycD1-, p27-, and p21-associated complexes were immunoprecipitated from MCF7 cells treated with palbociclib for the indicated time, and proteins were detected by Western blot. (**D**) CycE1 was immunoprecipitated from MCF7 cells treated with palbociclib for 48 hours. (**E**) Model for trimer assembly and CDK4 inhibition by palbociclib. Like p16 family CDK4 inhibitors, palbociclib binds monomer CDK4 and indirectly leads to inactive CDK2 complexes. Although not observed in the breast cancer cells studied here, there may be some contexts in which palbociclib also targets CDK4 dimers.

other kinase therapeutic targets as well (*69*), and we note that Gleevec (imatinib) similarly inhibits an inactive conformation of the Abl kinase (*70*).

## Materials and methods

### *Recombinant protein expression*

Full-length human CDK4, CDK6, cycD1, p21, p27, and Brk were expressed and purified from Sf9 cells. CDK4, CDK6, p21, p27, and Brk were expressed as GST fusion proteins. CycD1 was coexpressed with other components untagged. Lysates were first purified by GS4B affinity chromatography. Proteins were then eluted from the resin and subject to Source 15Q (GE Healthcare) anion exchange chromatography. The elution fraction was then subjected to TEV protease cleavage overnight in 25 mM Tris pH 8.0, 200 mM NaCl, 1 mM DTT, 0.5 mM EDTA. The protein was then passed over GS4B affinity resin again to remove free GST, concentrated, and stored in a buffer containing 20 mM Tris pH 8.0, 200 mM NaCl, 1 mM DTT, and 20% glycerol. CDK2-CycA was expressed in *E. coli* and purified as described (*71*).

### *Phosphorylation of p21 and p27*

Recombinant p21 and p27 were expressed as a GST fusion in *E. coli* and purified as described above. Purified proteins were treated with 10%

GST-Brk (by mass) in a buffer containing 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$, and 1 mM ATP and incubated at 4°C for 24 hours. The phosphorylated p21 or p27 was purified by passing through GS4B affinity resin and eluted from a Superdex 75 column (GE Healthcare) in a buffer containing 25 mM Tris pH 8.0, 100 mM NaCl, and 1 mM DTT. The extent of phosphorylation was confirmed using electrospray mass spectrometry on a SCIEX X500 QTOF spectrometer.

### *Crystallization, data collection, structure determination, and model refinement*

For crystallization, human CDK4 residues 1-303 (ΔG45-G47, G43E, G44E), truncated cycD1 (16-267), p27 kinase-inhibitory domain ($p27^{KID}$; 25-93), and the p21 kinase-inhibitory domain ($p21^{KID}$; 14-81) were prepared as described above. The CDK4 contains mutations in a glycine-rich sequence that mimic CDK6 and were required for crystallization (*10*, *11*). The CDK4 complexes were prepared for crystallization by elution from a Superdex 200 column (GE Healthcare) in a buffer containing 10 mM Tris pH 8.0, 100 mM NaCl, and 1 mM DTT. The p27-CDK4-CycD1, p27(3E)-CDK4-CycD1, and p21-CDK4-CycD1 complexes were crystallized from a 10 mg/ml solution by microbatch method under Al's oil at 22°C. Rods formed after 3 days in 100 mM Tris, 10% PEG 8000, and 200 mM $MgCl_2$, pH 7.0. Crystals were cryoprotected in reservoir solution supplemented with 25% glycerol and cryo-cooled in liquid nitrogen.

The phosp27-CDK4-CycD1 complex was prepared for crystallization by mixing 3:1 molar ratio of phosphorylated p27 and CDK4-CycD1 dimer followed by elution from a Superdex 200 column in a buffer containing 10 mM Tris, 100 mM NaCl, and 1 mM DTT, pH 8.0. The complex was crystallized from a 10 mg/ml solution by sitting drop vapor diffusion method at 22°C. Rods formed after 3 days in 100 mM Tris, 17% PEG 3350, 100 mM $CaCl_2$, and 10 mM $MgCl_2$, pH 7.0. Crystals were cryo-cooled in the reservoir solution supplemented with 25% glycerol.

Data were collected at the Advanced Light Source, Laurence Berkeley National Laboratory at beamline 8.3.1. Diffraction spots were integrated using MOSFLM (*72*), and data were merged and scaled using Scala in the CCP4 software package (*73*, *74*). The structure of the p27-CDK4-CycD1 trimer was first determined by molecular replacement using Phaser and the CDK4-CycD1 dimer as a search model (PDB ID 2W9F) (*75*). A model for p27 was then built into the extra electron density with Coot (*76*), and the model was refined with Phenix (*77*). Models for the CDK4-CycD1 trimer complexes containing phosp27, p27(3E), or p21 were then built and refined starting from the unphosphorylated p27 structure.

### *Western blots and antibodies*

Details on the antibodies used and the procedures for Western blotting can be found in the supplementary materials.

### *Kinase assays*

Recombinant CDK4 (0.5 μM), CDK2 (0.5 μM), or CDK6 (1 μM) complexes were mixed with substrate (30 μM) in a kinase buffer containing 25 mM Tris pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 250 μM ATP (or as indicated), and 100 μCi of [$^{32}$P]γ-ATP. Substrate was diluted into the reaction buffer, and the reaction was initiated by addition of ATP. Reactions were quenched after 30 min by addition of SDS-PAGE loading buffer. Independent time-course experiments confirmed that phosphate addition is still linear with time beyond 45 min using our experimental conditions. SDS-PAGE gels were imaged with a Typhoon scanner and bands quantified using the ImageJ software package. For each assay, three replicates were performed. We note that our measured $K_i$ values for the inhibitors on K4D1/– (Fig. 5B and fig. S8A) are consistent with our ITC-measured $K_d$ value for palbociclib (Fig. 5C) but are higher than $IC_{50}$ values previously reported (*4*, *7*, *46*). We believe the difference is due to the higher ATP concentration in our assay, which we chose to be closer to the ATP $K_M$ for K4D1/–.

For the activity assays using immunoprecipitation of endogenous CDK4, CDK2, or p27 complexes, 1 mg of whole-cell extract was immunoprecipitated with 2 μg of anti-p27 antibody (SCBT DCS-72), 2 μg of anti-CDK2 antibody (SCBT D-12), or 20 μl of CDK4 antiserum to C-terminal peptide (gift of C. Sherr, St. Jude Children's Research Hospital) (*48*) and 25 μl of protein A/G PLUS beads (SCBT) for 2 hours at 4°C. The beads were then washed 3 times with 50 mM Tris pH 7, 150 mM NaCl, 1 mM DTT, and 5% glycerol. The final wash was performed in kinase buffer and the reaction was performed with the recombinant substrate while the enzyme complexes were on the beads.

For the activity assays using immunoprecipitated CycD1 complexes, MCF7 cells were lysed with lysis buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% NP-40, 10 mM DTT, 10% glycerol, 1× Halt protease inhibitor (Thermo Fisher), and 1 mM PMSF. 5 μg of anti-CycD1 (Invitrogen; MA5-12707) was incubated with 300 μg of the lysate overnight at 4°C. Normal mouse IgG1 (Cell Signaling Technology, 5415) was used as a control. Protein G–agarose beads were added to each of the IP samples and incubated up to 4 hours at 4°C. Protein immunocomplexes were washed 3 times with the lysis buffer and 2 times with kinase reaction buffer (50 mM HEPES-KOH pH 7.5, 20 mM $MgCl_2$, 1 mM DTT). Kinase reactions were carried out in 100 μl of kinase buffer in the presence of 2 mM ATP and 0.5 μg of recombinant Rb C-terminal domain as substrate by gently shaking at room temperature up to 1 hour. The resulting phosphorylated Rb protein was detected by Western blotting using anti-p-Rb (S780) antibody (Cell Signaling Technology, 9307L).

Activity assays with immunoprecipitated CycE1 complexes were performed similarly, but the lysis buffer contained 50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.1% Tween-20, 1× Halt protease inhibitor, and 1 mM PMSF. Immunoprecipitation was performed with an anti-CycE1 antibody (Santa Cruz Biotechnology; SC-377100) as described for CycD1. The kinase reactions were performed in 100 μl of kinase buffer (40 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 0.1 mg/ml BSA, 50 μM DTT) in the presence of 2 mM ATP and 0.5 μg of Rb protein as substrate by gently shaking at room temperature up to 1 hour. The resulting phosphorylated Rb protein was detected by Western blotting using anti-p-Rb (S807/S811) antibody (Cell Signaling Technology, 8516S).

### *XO44 labeling assays*

XO44 labeling was carried out as described (*49*) with some modifications. Live adherent cells were labeled by incubation with 2 μM XO44 at 37°C for 30 min. Cells grown in 15-cm dish format to be analyzed by gel filtration were harvested as described below. Cells grown in 6-well plate format (seeded at 500,000 cells per well and allowed to recover for ~20 hours) to be analyzed by Western blot were then washed twice with cold PBS, perturbed and lysed in place with 80 μl of lysis buffer containing 100 mM HEPES pH 7.5, 150 mM NaCl, 0.1% NP-40, 1× cOmplete EDTA-free protease inhibitor cocktail and 1× PhosSTOP phosphatase inhibitor cocktail (both Roche). Lysates were clarified by centrifuging at 20,000*g* for 20 min, then normalized to 2.5 mg/ml protein concentration using the Pierce Rapid Gold BCA assay (Thermo). 20.75 μl of lysate was treated with 4.25 μl of a master mix to give final concentrations of 1% SDS, 50 μM TAMRA-azide (in DMSO), 1 mM TCEP, 100 μM TBTA (in 1:4 DMSO:t-butyl alcohol) and 1 mM $CuSO_4$. The resulting mixture was incubated at room temperature for 1 hour before being quenched with 5 μl of 6× SDS loading buffer and analyzed by Western blot as described above. The TAMRA dye was used to add mass to XO44-labeled CDK4, so the XO44 modification is observable as a gel shift by using SDS-PAGE.

### *Gel filtration assay*

Gel filtration assays on crude cell lysates were performed as described (*36*) with modifications. Cells were seeded on 15-cm dishes at 10 million cells per dish and allowed to recover for ~20 hours before treatment with compounds as indicated and harvesting by washing with

PBS followed by trypsinizing, pelleting and two further washes with cold PBS. Pellets of ~25 million cells were lysed in buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 1% NP-40, 10% glycerol, 1 mM EDTA plus cOmplete EDTA-free protease inhibitor cocktail (Roche) on ice for 20 min. Lysates were then clarified by centrifuging at 18,000*g* for 2 × 15 min. In each case, 500 µl (5 mg) of lysate was incubated with DNase I (NEB, 1 unit/mg lysate) in 1× DNase 1 reaction buffer (NEB) for 20 min on ice and briefly clarified again by centrifuging for 5 min. Samples were then loaded on a Superdex200 Increase 10/300 GL column pre-equilibrated in 50 mM HEPES pH 7.5, 150 mM NaCl, 1% NP-40, 10% glycerol. The column was eluted for 1.25 column volumes at 0.35 ml/min and 0.5-ml fractions were collected starting at 5 ml. Fractions 10 to 26 were analyzed by Western blot as described above. Samples of fractions from XO44-treated cells were used in click reactions as described above. To benchmark purified protein complexes, 0.8 µg of p27-CDK4-CycD1 trimer was diluted in 500 µl of HEPES buffer for loading. Gel filtration and Western blot analysis were performed as described above.

### *LC-MS analysis of CDK4 covalent labeling*

Recombinant purified protein complexes (1 µM) were suspended in a buffer containing 20 mM HEPES, 150 mM NaCl, and 1 mM $MgCl_2$, pH 7.5, and treated with XO44 (10 µM) or DMSO (1% (v/v) DMSO final). After 30 min at room temperature, reactions were stopped by the addition of 10% (v/v) formic acid to a final concentration of 2% (v/v) formic acid. The extent of covalent labeling was assessed by LC-MS (Waters Xevo G2-XS QTof, ACQUITY UPLC Protein BEH C4 Column, 300 Å, 1.7 µm, 2.1 mm × 50 mm). Deconvolution of multiply charged ions was performed using Waters MassLynx software (version 4.1).

### *Palbociclib-biotin*

Details of the synthesis and characterization of palbo-biotin can be found in the supplementary materials. For palbo-biotin precipitations, 100 µM palbo-biotin was first incubated with streptactin beads in PBS for 30 min at room temperature. The beads were then washed 3× with PBS and resuspended in 1 mg of cell extract in 50 mM HEPES pH 7.0, 150 mM NaCl, 0.1% Tween-20, 1 mM DTT, 5% glycerol, 1× cOmplete EDTA-free protease inhibitor cocktail and incubated at 4°C for 2 hours. Complexes bound to streptactin beads were eluted using 2× SDS buffer and subjected to immunoblot analysis. For recombinant protein, the same protocol was followed except 2 µg of recombinant protein was added to the beads in IP buffer with 1% BSA. 10 µM excess free palbociclib was added to the extract or purified proteins where indicated.

### *Isothermal titration calorimetry*

Equilibrium dissociation constants were obtained using a Micro Cal VP-ITC at 15°C. For drug binding experiments, 40 µM purified CDK4-CycD1, phosp27($p27^{KID}$; 25-93)-CDK4-CycD1, and GST-CDK4 were dialyzed overnight in a buffer containing 25 mM Tris pH 8.0, 200 mM NaCl, and 5 mM BME. Proteins were titrated with 0.5 mM palbociclib dissolved in the same buffer. For measurement of p27 ($p27^{KID}$; 25-93) binding to GST-CDK4, proteins were eluted in separate runs from a Superdex 200 column equilibrated in a buffer containing 25 mM Tris pH 8.0, 200 mM NaCl, and 5 mM BME. 250 µM p27 was titrated into 40 µM GST-CDK4 in the absence and presence of 50 µM palbociclib.

## REFERENCES AND NOTES


1. T. Hunter, J. Pines, Cyclins and cancer. II: Cyclin D and CDK inhibitors come of age. *Cell* **79**, 573–582 (1994). doi: 10.1016/0092-8674(94)90543-6; pmid: 7954824
2. C. J. Sherr, Cancer cell cycles. *Science* **274**, 1672–1677 (1996). doi: 10.1126/science.274.5293.1672; pmid: 8939849
3. M. Malumbres, Cyclin-dependent kinases. *Genome Biol.* **15**, 122 (2014). doi: 10.1186/gb4184; pmid: 25180339
4. T. Otto, P. Sicinski, Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* **17**, 93–115 (2017). doi: 10.1038/nrc.2016.138; pmid: 28127048
5. C. J. Sherr, D. Beach, G. I. Shapiro, Targeting CDK4 and CDK6: From Discovery to Therapy. *Cancer Discov.* **6**, 353–367 (2016). doi: 10.1158/2159-8290.CD-15-0894; pmid: 26658964
6. R. A. Weinberg, The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995). doi: 10.1016/0092-8674(95)90385-2; pmid: 7736585
7. R. S. Finn *et al*., Palbociclib and Letrozole in Advanced Breast Cancer. *N. Engl. J. Med.* **375**, 1925–1936 (2016). doi: 10.1056/NEJMoa1607303; pmid: 27959613
8. R. S. Finn *et al*., The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): A randomised phase 2 study. *Lancet Oncol.* **16**, 25–35 (2015). doi: 10.1016/S1470-2045(14)71159-3; pmid: 25524798
9. P. D. Jeffrey *et al*., Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376**, 313–320 (1995). doi: 10.1038/376313a0; pmid: 7630397
10. P. J. Day *et al*., Crystal structure of human CDK4 in complex with a D-type cyclin. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4166–4170 (2009). doi: 10.1073/pnas.0809645106; pmid: 19237565
11. T. Takaki *et al*., The structure of CDK4/cyclin D3 has implications for models of CDK activation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4171–4176 (2009). doi: 10.1073/pnas.0809674106; pmid: 19237555
12. H. Matsushime *et al*., Identification and properties of an atypical catalytic subunit (p34PSK-J3/cdk4) for mammalian D type G1 cyclins. *Cell* **71**, 323–334 (1992). doi: 10.1016/0092-8674(92)90360-0; pmid: 1423597
13. L. Anders *et al*., A systematic screen for CDK4/6 substrates links FOXM1 phosphorylation to senescence suppression in cancer cells. *Cancer Cell* **20**, 620–634 (2011). doi: 10.1016/j.ccr.2011.10.001; pmid: 22094256
14. M. Kitagawa *et al*., The consensus motif for phosphorylation by cyclin D1-Cdk4 is different from that for phosphorylation by cyclin A/E-Cdk2. *EMBO J.* **15**, 7060–7069 (1996). doi: 10.1002/j.1460-2075.1996.tb01097.x; pmid: 9003781
15. A. K. Konstantinidis, R. Radhakrishnan, F. Gu, R. N. Rao, W. K. Yeh, Purification, characterization, and kinetic mechanism of cyclin D1-CDK4, a major target for cell cycle regulation. *J. Biol. Chem.* **273**, 26506–26515 (1998). doi: 10.1074/jbc.273.41.26506; pmid: 9756886
16. Z. A. Knight, K. M. Shokat, Features of selective kinase inhibitors. *Chem. Biol.* **12**, 621–637 (2005). doi: 10.1016/j.chembiol.2005.04.011; pmid: 15975507
17. W. Pan, S. Cox, R. H. Hoess, R. H. Grafström, A cyclin D1/cyclin-dependent kinase 4 binding site within the C domain of the retinoblastoma protein. *Cancer Res.* **61**, 2885–2891 (2001). pmid: 11306463
18. B. R. Topacio *et al*., Cyclin D-Cdk4,6 Drives Cell-Cycle Progression via the Retinoblastoma Protein's C-Terminal Helix. *Mol. Cell* **74**, 758–770.e4 (2019). doi: 10.1016/j.molcel.2019.03.020; pmid: 30982746
19. K. A. Verba *et al*., Atomic structure of Hsp90-Cdc37-Cdk4 reveals that Hsp90 traps and stabilizes an unfolded kinase. *Science* **352**, 1542–1547 (2016). doi: 10.1126/science.aaf5023; pmid: 27339980
20. S. T. Hallett *et al*., Differential Regulation of G1 CDK Complexes by the Hsp90-Cdc37 Chaperone System. *Cell Rep.* **21**, 1386–1398 (2017). doi: 10.1016/j.celrep.2017.10.042; ]pmid: 29091774
21. D. H. Brotherton *et al*., Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d. *Nature* **395**, 244–250 (1998). doi: 10.1038/26164; pmid: 9751051
22. A. A. Russo, L. Tong, J.-O. Lee, P. D. Jeffrey, N. P. Pavletich, Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* **395**, 237–243 (1998). doi: 10.1038/26155; pmid: 9751050
23. M. Matsuoka, J. Y. Kato, R. P. Fisher, D. O. Morgan, C. J. Sherr, Activation of cyclin-dependent kinase 4 (cdk4) by mouse MO15-associated kinase. *Mol. Cell. Biol.* **14**, 7265–7275 (1994). doi: 10.1128/MCB.14.11.7265; pmid: 7935441
24. A. Ray, M. K. James, S. Larochelle, R. P. Fisher, S. W. Blain, p27Kip1 inhibits cyclin D-cyclin-dependent kinase 4 by two independent modes. *Mol. Cell. Biol.* **29**, 986–999 (2009). doi: 10.1128/MCB.00898-08; pmid: 19075005
25. C. J. Sherr, J. M. Roberts, CDK inhibitors: Positive and negative regulators of G1-phase progression. *Genes Dev.* **13**, 1501–1512 (1999). doi: 10.1101/gad.13.12.1501; pmid: 10385618
26. E. R. Lacy *et al*., p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat. Struct. Mol. Biol.* **11**, 358–364 (2004). doi: 10.1038/nsmb746; pmid: 15024385
27. J. Martín-Caballero, J. M. Flores, P. García-Palencia, M. Serrano, Tumor susceptibility of p21(Waf1/Cip1)-deficient mice. *Cancer Res.* **61**, 6234–6238 (2001). pmid: 11507077
28. M. L. Fero *et al*., A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in p27(Kip1)-deficient mice. *Cell* **85**, 733–744 (1996). doi: 10.1016/S0092-8674(00)81239-8; pmid: 8646781
29. T. Sperka, J. Wang, K. L. Rudolph, DNA damage checkpoints in stem cells, ageing and cancer. *Nat. Rev. Mol. Cell Biol.* **13**, 579–590 (2012). doi: 10.1038/nrm3420; pmid: 22914294
30. A. A. Russo, P. D. Jeffrey, A. K. Patten, J. Massagué, N. P. Pavletich, Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* **382**, 325–331 (1996). doi: 10.1038/382325a0; pmid: 8684460
31. M. Cheng *et al*., The p21(Cip1) and p27(Kip1) CDK 'inhibitors' are essential activators of cyclin D-dependent kinases in murine fibroblasts. *EMBO J.* **18**, 1571–1583 (1999). doi: 10.1093/emboj/18.6.1571; pmid: 10075928
32. S. W. Blain, E. Montalvo, J. Massagué, Differential interaction of the cyclin-dependent kinase (Cdk) inhibitor p27Kip1 with cyclin A-Cdk2 and cyclin D2-Cdk4. *J. Biol. Chem.* **272**, 25863–25872 (1997). doi: 10.1074/jbc.272.41.25863; pmid: 9325318
33. T. J. Soos *et al*., Formation of p27-CDK complexes during the human mitotic cell cycle. *Cell Growth Differ.* **7**, 135–146 (1996). pmid: 8822197
34. M. K. James, A. Ray, D. Leznova, S. W. Blain, Differential modification of p27Kip1 controls its cyclin D-cdk4 inhibitory activity. *Mol. Cell. Biol.* **28**, 498–510 (2008). doi: 10.1128/MCB.02171-06; pmid: 17908796
35. M. D. Larrea *et al*., Phosphorylation of p27Kip1 regulates assembly and activation of cyclin D1-Cdk4. *Mol. Cell. Biol.* **28**, 6462–6472 (2008). doi: 10.1128/MCB.02300-07; pmid: 18710949
36. B. B. McConnell, F. J. Gregory, F. J. Stott, E. Hara, G. Peters, Induced expression of p16(INK4a) inhibits both CDK4- and CDK2-associated kinase activity by reassortment of cyclin-CDK-inhibitor complexes. *Mol. Cell. Biol.* **19**, 1981–1989 (1999). doi: 10.1128/MCB.19.3.1981; pmid: 10022885
37. J. LaBaer *et al*., New functional activities for the p21 family of CDK inhibitors. *Genes Dev.* **11**, 847–862 (1997). doi: 10.1101/gad.11.7.847; pmid: 9106657
38. P. Patel *et al*., Brk/Protein tyrosine kinase 6 phosphorylates p27KIP1, regulating the activity of cyclin D-cyclin-dependent kinase 4. *Mol. Cell. Biol.* **35**, 1506–1522 (2015). doi: 10.1128/MCB.01206-14; pmid: 25733683

39. N. R. Brown, M. E. Noble, J. A. Endicott, L. N. Johnson, The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nat. Cell Biol.* **1**, 438–443 (1999). doi: 10.1038/15674; pmid: 10559988
40. B. A. Schulman, D. L. Lindstrom, E. Harlow, Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10453–10458 (1998). doi: 10.1073/pnas.95.18.10453; pmid: 9724724
41. B. Nolen, S. Taylor, G. Ghosh, Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* **15**, 661–675 (2004). doi: 10.1016/j.molcel.2004.08.024; pmid: 15350212
42. D. J. Wood, J. A. Endicott, Structural insights into the functional diversity of the CDK-cyclin family. *Open Biol.* **8**, 180112 (2018). doi: 10.1098/rsob.180112; pmid: 30185601
43. S. W. Blain, Switching cyclin D-Cdk4 kinase activity on and off. *Cell Cycle* **7**, 892–898 (2008). doi: 10.4161/cc.7.7.5637; pmid: 18414028
44. M. Tsytlonok *et al*., Dynamic anticipation by Cdk2/Cyclin A-bound p27 mediates signal integration in cell cycle regulation. *Nat. Commun.* **10**, 1676 (2019). doi: 10.1038/s41467-019-09446-w; pmid: 30976006
45. M. Grimmler *et al*., Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases. *Cell* **128**, 269–280 (2007). doi: 10.1016/j.cell.2006.11.047; pmid: 17254966
46. P. L. Toogood *et al*., Discovery of a potent and selective inhibitor of cyclin-dependent kinase 4/6. *J. Med. Chem.* **48**, 2388–2406 (2005). doi: 10.1021/jm049354h; pmid: 15801831
47. P. Chen *et al*., Spectrum and Degree of CDK Drug Interactions Predicts Clinical Performance. *Mol. Cancer Ther.* **15**, 2273–2281 (2016). doi: 10.1158/1535-7163.MCT-16-0300; pmid: 27496135
48. H. Matsushime *et al*., D-type cyclin-dependent kinase activity in mammalian cells. *Mol. Cell. Biol.* **14**, 2066–2076 (1994). doi: 10.1128/MCB.14.3.2066; pmid: 8114738
49. Q. Zhao *et al*., Broad-Spectrum Kinase Profiling in Live Cells with Lysine-Targeted Sulfonyl Fluoride Probes. *J. Am. Chem. Soc.* **139**, 680–685 (2017). doi: 10.1021/jacs.6b08536; pmid: 28051857
50. S. Polier *et al*., ATP-competitive inhibitors block protein kinase recruitment to the Hsp90-Cdc37 system. *Nat. Chem. Biol.* **9**, 307–312 (2013). doi: 10.1038/nchembio.1212; pmid: 23502424
51. J. L. Green *et al*., Direct CDKN2 Modulation of CDK4 Alters Target Engagement of CDK4 Inhibitor Drugs. *Mol. Cancer Ther.* **18**, 771–779 (2019). doi: 10.1158/1535-7163.MCT-18-0755; pmid: 30837298
52. J. L. Dean, C. Thangavel, A. K. McClendon, C. A. Reed, E. S. Knudsen, Therapeutic CDK4/6 inhibition in breast cancer: Key mechanisms of response and failure. *Oncogene* **29**, 4018–4032 (2010). doi: 10.1038/onc.2010.154; pmid: 20473330
53. W. R. Wiedemeyer *et al*., Pattern of retinoblastoma pathway inactivation dictates response to CDK4/6 inhibition in GBM. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11501–11506 (2010). doi: 10.1073/pnas.1001613107; pmid: 20534551
54. J. W. Harper, G. R. Adami, N. Wei, K. Keyomarsi, S. J. Elledge, The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993). doi: 10.1016/0092-8674(93)90499-G; pmid: 8242751
55. A. Kato, H. Takahashi, Y. Takahashi, H. Matsushime, Inactivation of the cyclin D-dependent kinase in the rat fibroblast cell line, 3Y1, induced by contact inhibition. *J. Biol. Chem.* **272**, 8065–8070 (1997). doi: 10.1074/jbc.272.12.8065; pmid: 9065480
56. M. H. Ladha, K. Y. Lee, T. M. Upton, M. F. Reed, M. E. Ewen, Regulation of exit from quiescence by p27 and cyclin D1-CDK4. *Mol. Cell. Biol.* **18**, 6605–6615 (1998). doi: 10.1128/MCB.18.11.6605; pmid: 9774675
57. H. Toyoshima, T. Hunter, p27, a novel inhibitor of G1 cyclin-Cdk protein kinase activity, is related to p21. *Cell* **78**, 67–74 (1994). doi: 10.1016/0092-8674(94)90573-8; pmid: 8033213
58. K. Polyak *et al*., p27Kip1, a cyclin-Cdk inhibitor, links transforming growth factor-beta and contact inhibition to cell cycle arrest. *Genes Dev.* **8**, 9–22 (1994). doi: 10.1101/gad.8.1.9; pmid: 8288131
59. H. Zhang, G. J. Hannon, D. Beach, p21-containing cyclin kinases exist in both active and inactive states. *Genes Dev.* **8**, 1750–1758 (1994). doi: 10.1101/gad.8.15.1750; pmid: 7958854
60. H. Wang *et al*., The metabolic function of cyclin D3-CDK6 kinase in cancer cell survival. *Nature* **546**, 426–430 (2017). doi: 10.1038/nature22797; pmid: 28607489
61. J. Zhang *et al*., Cyclin D-CDK4 kinase destabilizes PD-L1 via cullin 3-SPOP to control cancer immune surveillance. *Nature* **553**, 91–95 (2018). doi: 10.1038/nature25015; pmid: 29160310
62. B. Zhao, K. Burgess, PROTACs suppression of CDK4/6, crucial kinases for cell cycle regulation in cancer. *Chem. Commun.* **55**, 2704–2707 (2019). doi: 10.1039/C9CC00163H; pmid: 30758029
63. A. E. Schade, M. G. Oser, H. E. Nicholson, J. A. DeCaprio, Cyclin D-CDK4 relieves cooperative repression of proliferation and cell cycle gene expression by DREAM and RB. *Oncogene* **38**, 4962–4976 (2019). doi: 10.1038/s41388-019-0767-9; pmid: 30833638
64. M. R. Ramsey *et al*., Expression of p16Ink4a compensates for p18Ink4c loss in cyclin-dependent kinase 4/6-dependent tumors and tissues. *Cancer Res.* **67**, 4732–4741 (2007). doi: 10.1158/0008-5472.CAN-06-3437; pmid: 17510401
65. R. J. Young *et al*., Loss of CDKN2A expression is a frequent event in primary invasive melanoma and correlates with sensitivity to the CDK4/6 inhibitor PD0332991 in melanoma cell lines. *Pigment Cell Melanoma Res.* **27**, 590–600 (2014). doi: 10.1111/pcmr.12228; pmid: 24495407
66. H. Jiang, H. S. Chou, L. Zhu, Requirement of cyclin E-Cdk2 inhibition in p16(INK4a)-mediated growth suppression. *Mol. Cell. Biol.* **18**, 5284–5290 (1998). doi: 10.1128/MCB.18.9.5284; pmid: 9710613
67. I. Reynisdóttir, K. Polyak, A. Iavarone, J. Massagué, Kip/Cip and Ink4 Cdk inhibitors cooperate to induce cell cycle arrest in response to TGF-beta. *Genes Dev.* **9**, 1831–1845 (1995). doi: 10.1101/gad.9.15.1831; pmid: 7649471
68. J. Moser, I. Miller, D. Carter, S. L. Spencer, Control of the Restriction Point by Rb and p21. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8219–E8227 (2018). doi: 10.1073/pnas.1722446115; pmid: 30111539
69. C. J. Novotny *et al*., Overcoming resistance to HER2 inhibitors through state-specific kinase binding. *Nat. Chem. Biol.* **12**, 923–930 (2016). doi: 10.1038/nchembio.2171; pmid: 27595329
70. T. Schindler *et al*., Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**, 1938–1942 (2000). doi: 10.1126/science.289.5486.1938; pmid: 10988075
71. D. A. McGrath *et al*., Structural basis of divergent cyclin-dependent kinase activation by Spy1/RINGO proteins. *EMBO J.* **36**, 2251–2262 (2017). doi: 10.15252/embj.201796905; pmid: 28666995
72. A. G. Leslie, The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006). doi: 10.1107/S0907444905039107; pmid: 16369093
73. Collaborative Computational Project 4, The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994). doi: 10.1107/S0907444994003112; pmid: 15299374
74. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006). doi: 10.1107/S0907444905036693; pmid: 16369096
75. A. J. McCoy *et al*., Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). doi: 10.1107/S0021889807021206; pmid: 19461840
76. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
77. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702


## ACKNOWLEDGMENTS

**Funding:** Supported by NIH grants R01CA211878 (E.S.K. and A.K.W.), R01 CA190408 and U01 CA19924 (K.M.S.), and R01 GM124148 and R01 CA228413 (S.M.R.); Tobacco-Related Disease Research Program of the University of California grant 28IR-0046 (S.M.R.); the Samuel Waxman Cancer Research Foundation (K.M.S.); and NIH fellowships F31 CA206244 (K.Z.G.) and F30 CA239476 (K.L.). Data collection at the ALS Beamline 8.3.1 is supported by the UC Office of the President, Multicampus Research Programs and Initiatives grant MR-15-328599 and the Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation. **Author contributions:** K.Z.G. and S.M.R. designed the structural studies and enzyme analysis; K.Z.G., J.W.S., K.L., K.J.B., and K.M.S. designed the XO44 and palbociclib-biotin experiments; K.Z.G., A.K.W., E.S.K., K.M.S., and S.M.R. designed the breast cancer cell experiments; K.Z.G., T.U.W., and K.L.B. purified protein; K.Z.G. crystallized protein and determined the structures with help from S.T.; K.Z.G. and S.M.R. performed enzyme and binding assays; J.W.S., K.L., and K.J.B. performed the XO44 experiments; K.L. synthesized palbociclib-biotin; K.Z.G., J.W.S., and V.K. performed immunoprecipitation experiments; and K.Z.G., J.W.S., A.K.W., E.S.K., K.M.S., and S.M.R. wrote the manuscript. **Competing interests:** S.M.R. and K.Z.G. are inventors on patent application 62/663,914 submitted by the Regents of the University of California that covers an engineered p27 protein. Kin of K.L. hold stock in and are employed by Pharmaron, a contract research organization. **Data and materials availability:** All data and materials are publicly available. Structure factors and coordinates for the crystal structures are available in the Protein Data Bank under accession codes 6P8E, 6P8F, 6P8G, and 6P8H.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/eaaw2106/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 and S2
Reference (*78*)

View/request a protocol for this paper from *Bio-protocol*.

29 November 2018; accepted 5 November 2019
10.1126/science.aaw2106

RESEARCH ARTICLE SUMMARY

MICROBIOTA

# Depletion of microbiome-derived molecules in the host using *Clostridium* genetics

**Chun-Jun Guo, Breanna M. Allen, Kamir J. Hiam, Dylan Dodd, Will Van Treuren, Steven Higginbottom, Kazuki Nagashima, Curt R. Fischer, Justin L. Sonnenburg, Matthew H. Spitzer*, Michael A. Fischbach***

**INTRODUCTION:** The gut microbiota produce hundreds of molecules that are present at high concentrations in circulation and whose levels vary widely among humans. These molecules are a promising starting point for studying the mechanisms of microbiome-host interactions; the few that have been characterized in detail have potent immune or metabolism modulatory activity and are ligands for a G protein–coupled receptor or nuclear hormone receptor. However, in most cases, molecule production has not been linked to specific bacterial strains or metabolic pathways and unraveling the contribution of each molecule to host biology remains difficult.

**RATIONALE:** A general system to knock out the production of microbiome-derived molecules in host tissues would enable interrogation of the mechanisms by which they modulate host biology and disease processes. Such a system has been elusive owing to limitations in the genetic manipulability of bacterial species from the microbiome. *Clostridium* and its relatives, the anaerobic Firmicutes, have been especially difficult to manipulate. Because the anaerobic Firmicutes are the source of many molecules in this pool, their genetic intractability is a central challenge in studying microbiome-derived molecules.

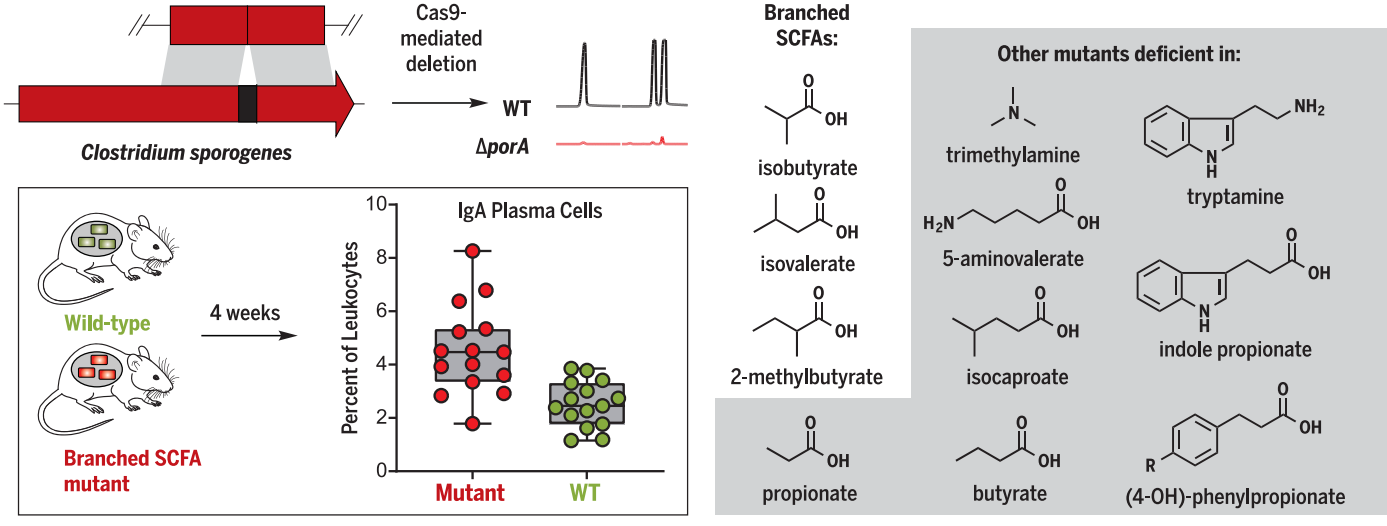
**RESULTS:** We describe a CRISPR-Cas9–based method for reliably constructing markerless deletions in a model commensal *Clostridium*: *Clostridium sporogenes*. Our system is recyclable, enabling the construction of multiple mutated strains. We demonstrate the utility of this method by using it to knock out the production of 10 *C. sporogenes*–derived molecules: trimethylamine, 5-aminovalerate, tryptamine, indole propionate, isovalerate, 2-methylbutryate, isobutyrate, isocaproate, propionate, and butyrate; we validated each knockout by verifying the absence of the corresponding metabolite in culture extracts by liquid chromatography–mass spectrometry or gas chromatography–mass spectrometry. We then colonized germ-free mice with wild-type *C. sporogenes* or one of five metabolite-deficient mutants, showing that the products of these pathways accumulate in host tissues but are depleted one at a time in mice colonized by the corresponding pathway mutant. By comparing mice colonized by wild-type *C. sporogenes* versus a mutant deficient in the production of the branched short-chain fatty acids isobutyrate, 2-methylbutyrate, and isovalerate, we discovered a previously unknown immunoglobulin A (IgA) plasma cell–modulatory activity of these abundant microbiome-derived molecules.

**CONCLUSION:** New genetic systems for abundant bacterial species in the microbiome will enable a variety of mechanistic questions to be addressed. Among these, our method opens the door to interrogating, in a controlled way, the role of highly abundant chemicals from the microbiome in host biology. In addition, it could be useful in future efforts of two varieties. First, its scope could be expanded to other currently inaccessible Firmicutes, such as butyrate producers from *Clostridium* clusters IV and XIVa (e.g., *Faecalibacterium prausnitzii*) and the bile acid metabolizers *C. scindens* and *C. hylemonae*. Second, it could be used to control the molecular output of the microbiome by editing the genomes of metabolite-producing Firmicutes. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aav1282

The list of author affiliations is available in the full article online.
*Corresponding author. Email: fischbach@fischbachgroup.org (M.A.F.); matthew.spitzer@ucsf.edu (M.H.S.)
Cite this article as C.-J. Guo *et al.*, *Science* **366**, eaav1282 (2019). DOI: 10.1126/science.aav1282




**Deciphering the contribution of microbiota-derived molecules to host biology remains difficult.** To help address this challenge, we introduce a new method for constructing markerless deletions in a model gut *Clostridium*, using it to knock out the production of 10 *C. sporogenes*–derived molecules (right). By colonizing germ-free mice with wild-type (WT) *C. sporogenes* versus a mutant (Δ*porA*) deficient in the production of branched short-chain fatty acids (SCFAs) (top left), we reveal an immune modulatory function of these molecules (bottom left). Error bars indicate standard deviation.

RESEARCH ARTICLE

MICROBIOTA

# Depletion of microbiome-derived molecules in the host using *Clostridium* genetics

Chun-Jun Guo[1,2], Breanna M. Allen[3,4], Kamir J. Hiam[3,4], Dylan Dodd[5,6], Will Van Treuren[6,4], Steven Higginbottom[6,4], Kazuki Nagashima[1], Curt R. Fischer[1,4], Justin L. Sonnenburg[6,4], Matthew H. Spitzer[3,4]*, Michael A. Fischbach[1,4]*

**The gut microbiota produce hundreds of molecules that are present at high concentrations in the host circulation. Unraveling the contribution of each molecule to host biology remains difficult. We developed a system for constructing clean deletions in *Clostridium* spp., the source of many molecules from the gut microbiome. By applying this method to the model commensal organism *Clostridium sporogenes*, we knocked out genes for 10 *C. sporogenes*–derived molecules that accumulate in host tissues. In mice colonized by a *C. sporogenes* for which the production of branched short-chain fatty acids was knocked out, we discovered that these microbial products have immunoglobulin A–modulatory activity.**

Gut bacteria produce hundreds of diffusible molecules that are notable for four reasons: (i) Most have no host source, so their levels are determined predominantly or exclusively by the microbiome. (ii) Many get into the bloodstream, so they can access peripheral tissues. (iii) They often reach concentrations that approach or exceed what a typical drug reaches, and the concentration range can be large—more than an order of magnitude in many cases—so they have the potential to cause biological differences among humans. (iv) Several of these molecules are known to be ligands for key host receptors; additional compounds from this category are candidate ligands for, for example, G protein–coupled receptors (GPCRs) and nuclear hormone receptors that play an important role in the host immune and metabolic systems (*1*). Thus, the gut microbiome is a prolific endocrine organ, but its output is not well understood.

The biological activities of most of these molecules remain unknown. One reason is that there has not been a general method for "toggling" one or more of them on or off in the host, akin to a gene knockout experiment in a model organism. Such a method would open the door to interrogating—and ultimately controlling—one of the most concrete contributions gut bacteria make to host biology.

Previous efforts that have sought to study an individual microbiome-derived molecule in the setting of host colonization have used two main strategies: First, a compound is administered by injection or gavage, which can offer insights into mechanism of action but suffers from the lack of a clean background (i.e., existing physiologic levels of the molecule of interest) and the possible effects of differences in route and timing of administration relative to the native context of gut bacterial production. Second, a bacterial species that produces the molecule is added or removed, which has the advantage of a more native-like context but makes it difficult to distinguish between molecule-induced phenotypes and other biological activities of the organism (*2*–*4*).

The most precise format for interrogating a microbiome-derived molecule is to compare two organisms that differ only in its production. Such an experiment has two threshold requirements: (i) knowledge of the metabolic genes for the molecule of interest and (ii) the ability to perform genetics in a robustly producing strain. This approach has been successful in *Bacteroides* (*5*, *6*), *Escherichia coli* (*7*), and *Lactobacillus* (*8*), but a key technical barrier limiting its generalization is that many of the known highly abundant gut-derived molecules are produced by *Clostridium* and its relatives, which have been difficult to manipulate genetically. We recently reported the use of a group II intron, a ribozyme-encoding mobile element (*9*), to mutate a single pathway in *Clostridium sporogenes* (*10*), but this insertional mutagenesis system performs unpredictably and cannot be used to make strains that carry multiple mutations. Here, we address these challenges by developing a highly efficient CRISPR-Cas9–based system for constructing single and multiple mutants in a model gut-resident *Clostridium* species.

[1]Department of Bioengineering and ChEM-H, Stanford University, Stanford, CA 94305, USA. [2]Jill Roberts Institute for Research in Inflammatory Bowel Disease, Department of Medicine, Weill Cornell Medicine, NY 10021, USA. [3]Graduate Program in Biomedical Sciences, Departments of Otolaryngology and Microbiology and Immunology, Helen Diller Family Comprehensive Cancer Center, Parker Institute for Cancer Immunotherapy, University of California, San Francisco, San Francisco, CA 94143, USA. [4]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA. [5]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA. [6]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA.
*Corresponding author. Email: fischbach@fischbachgroup.org (M.A.F.); matthew.spitzer@ucsf.edu (M.H.S.)

### Selection of *C. sporogenes* as a model gut *Clostridium*

We chose *C. sporogenes* ATCC 15579 (hereafter *C. sporogenes*) as a model gut commensal from the anaerobic Firmicutes for three reasons: This strain has long been known as a robust producer of highly abundant small molecules (*11*–*13*), it stably colonizes germ-free mice (*10*), and it is a commensal or mutualist (i.e., neither a pathogen nor a pathobiont). We began by performing metabolic profiling experiments to systematically determine the set of highly abundant small molecules produced by this strain in vitro. As shown in Fig. 1, *C. sporogenes* produces 10 molecules that are highly abundant and either primarily or exclusively derived from the gut microbiota: 5-aminovalerate, trimethylamine, tryptamine, indole propionate and other aryl propionates, isobutyrate, 2-methylbutyrate, isovalerate, isocaproate, propionate, and butyrate, confirming previous studies (*10*–*13*).

### Prediction and computational analysis of metabolic pathways

Next, we sought to predict the genes responsible for producing each molecule. Metabolic genes for trimethylamine (*14*), tryptamine (*13*), and indole propionate (*10*) have been discovered using genetics in *C. sporogenes* or another gut bacterium. Pathways for the remaining seven molecules had not been validated genetically in the gut microbiota. We made predictions for each one on the basis of three sources of evidence (Fig. 1): (i) pathways validated in nonmicrobiome organisms [e.g., the pathway for 5-aminovalerate production from the terrestrial isolate *Clostridium sticklandii* (*15*)]; (ii) biochemical studies that implicate an enzyme superfamily, which enabled us to search for orthologs in *C. sporogenes* [e.g., 2-hydroxyisocaproate dehydratases (*16*), which led us to a predicted cluster for isocaproate]; and (iii) a metagenomic analysis of butyrate gene clusters (*17*), which yielded a predicted gene cluster for butyrate production in *C. sporogenes*.

We determined whether the metabolic genes we predicted are widely distributed in the human gut microbiome and actively transcribed during colonization. We used MetaQuery (*18*) to measure the abundance of the key genes in each pathway (colored red in fig. S1) in >2000 publicly available human gut metagenomes. Every gene, except *tdcA*, was present in >95% of the stool metagenomes, indicating that the predicted pathways are cosmopolitan (minimum abundance = 1 copy per 1000 cells) (fig. S2). Because the mere presence of a gene in a metagenome does not imply that it is transcribed, we determined the transcript abundance

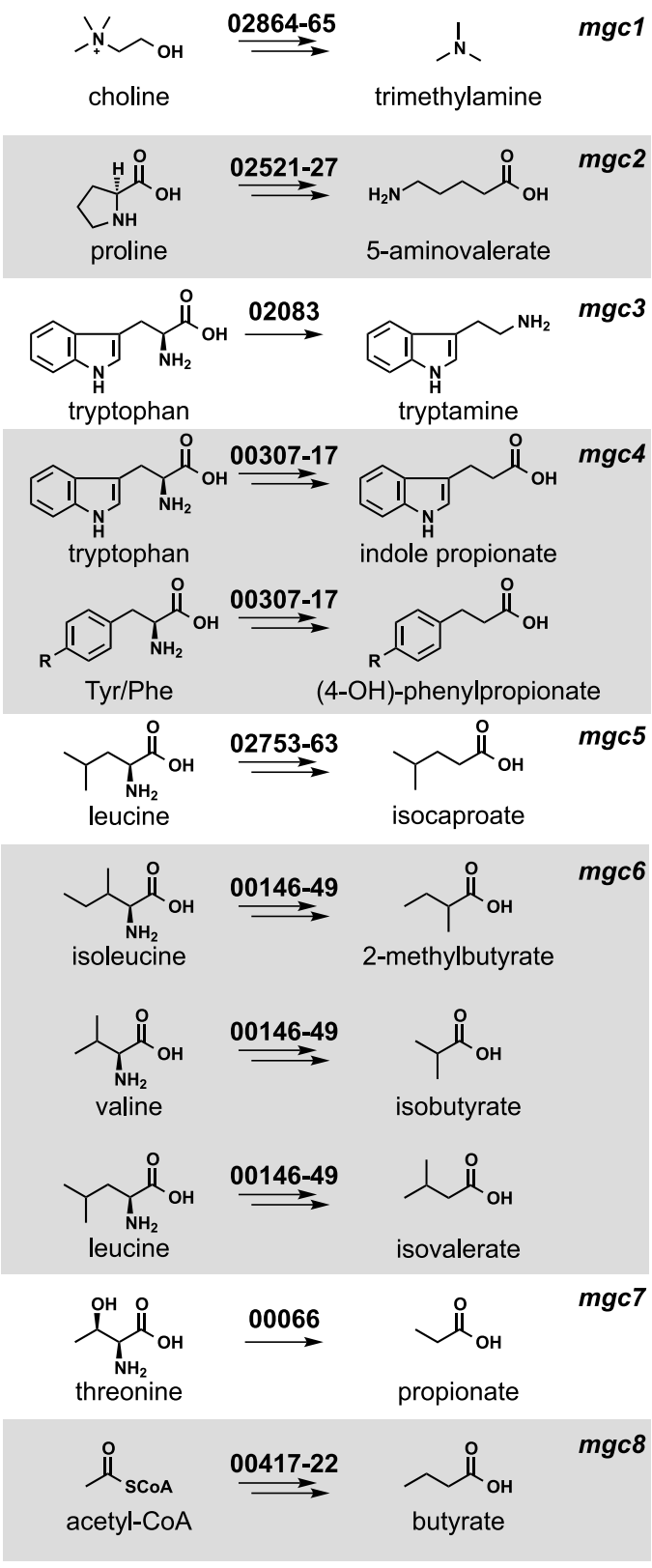



**Fig. 1. Metabolic pathways from *C. sporogenes* ATCC 15579 examined in this study.** Each pathway generates a microbiome-derived metabolite present at high abundance in the host's circulation. The prefix "*mgc*" stands for "metabolic gene cluster." Genes that comprise each pathway are shown above the corresponding arrows; the numbers indicate a locus tag suffix for *C. sporogenes*, where the prefix is "CLOSPO_."

of the key metabolic genes (including close homologs from non–*C. sporogenes* genomes) by recruiting reads from nine publicly available RNA-sequencing datasets derived from stool samples of healthy subjects (*19*). This analysis revealed that multiple homologs of each pathway are highly transcribed in the host (fig. S2). For example, *porA* and its homologs—ALIPUT_00387 in *Alistipes putredinis* DSM 17216, BACSTE_01839 from *Bacteroides stercoris* ATCC 43183, and BVU_2313 from *Bacteroides vulgatus* ATCC 8482—are transcribed in at least one sample. Taken together, these data suggest that the predicted pathways are widely distributed and actively transcribed in the host.

### Development of a CRISPR-Cas9–based genetic system for *C. sporogenes*

Next, we tested our metabolic pathway predictions by constructing mutants in each of them, along with the known pathways in *C. sporogenes* for tryptamine (*13*), indole propionate (*10*), and trimethylamine (*14*) (Fig. 1). The genetic system we used previously (*10*), which is based on a group II intron, had two important limitations: Because the intron's targeting mechanism is not well understood, generating one insertional mutant typically requires testing several targeting sequences; moreover, this process regularly fails to yield a mutant. After multiple attempts, we were not able to recycle the antibiotic resistance marker using ClosTron in order to create strains with multiple mutations; our experience is consistent with a previous report in the literature (*9*).

Reasoning that a dependable, markerless, recyclable genetic system for *Clostridium* would open the door to more systematic studies of microbiome metabolism, we developed a CRISPR-Cas9–based genome-editing system for *C. sporogenes*. *Clostridium* species have been notoriously difficult to modify genetically, in part because of inefficient homologous recombination (*9*). We postulated that a Cas9-induced double-strand break would help select for a rare homologous recombination event. To this end, we constructed a single vector that includes all the essential components of a bacterial CRISPR-Cas9 system—the Cas9 gene, a guide RNA (gRNA), and a 1.5- to 2.5-kb repair template—and transferred it by conjugation into *C. sporogenes*. However, we did not obtain viable colonies, even after multiple rounds of vector design modifications (see supplementary text for more detail).

Having observed that the conjugation efficiency for *C. sporogenes* is greatly diminished for plasmids >10 kb, we redesigned the system by splitting its components into two separate vectors: one that contains the gRNA and repair template and another that harbors Cas9 under the control of a ferredoxin promoter (Fig. 2 and fig. S3). When we introduced these

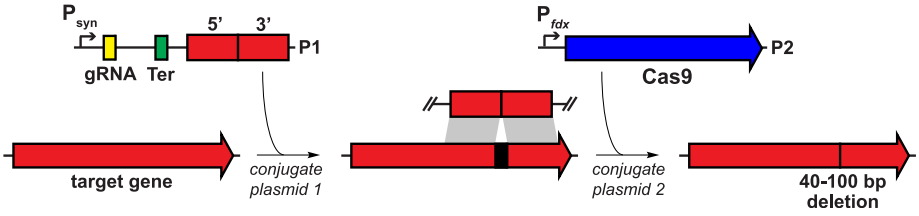



**Fig. 2. Development of a CRISPR-Cas9–based genetic system for *C. sporogenes*.** Schematic view of the genetic system. In the first step, plasmid P1 is introduced by conjugation into *C. sporogenes*. P1 contains a gRNA expressed under the control of $P_{syn}$, a synthetic promoter generated using PePPER (*31*); Ter, a terminator sequence from the *C. sporogenes* 16*S* ribosomal RNA gene; and a ~1.5- to 2.0-kb repair template. In the second step, plasmid P2 is introduced by conjugation. P2 consists of the Cas9 gene from *Streptococcus pyogenes* expressed under the control of $P_{fdx}$, the promoter from the *C. sporogenes* ferredoxin gene. Key steps in the development of the method were to introduce the genome-editing components (Cas9, gRNA, and repair template) sequentially on two plasmids and to lengthen the donor-acceptor cocultivation step of the second conjugal transfer from 24 to 72 hours.

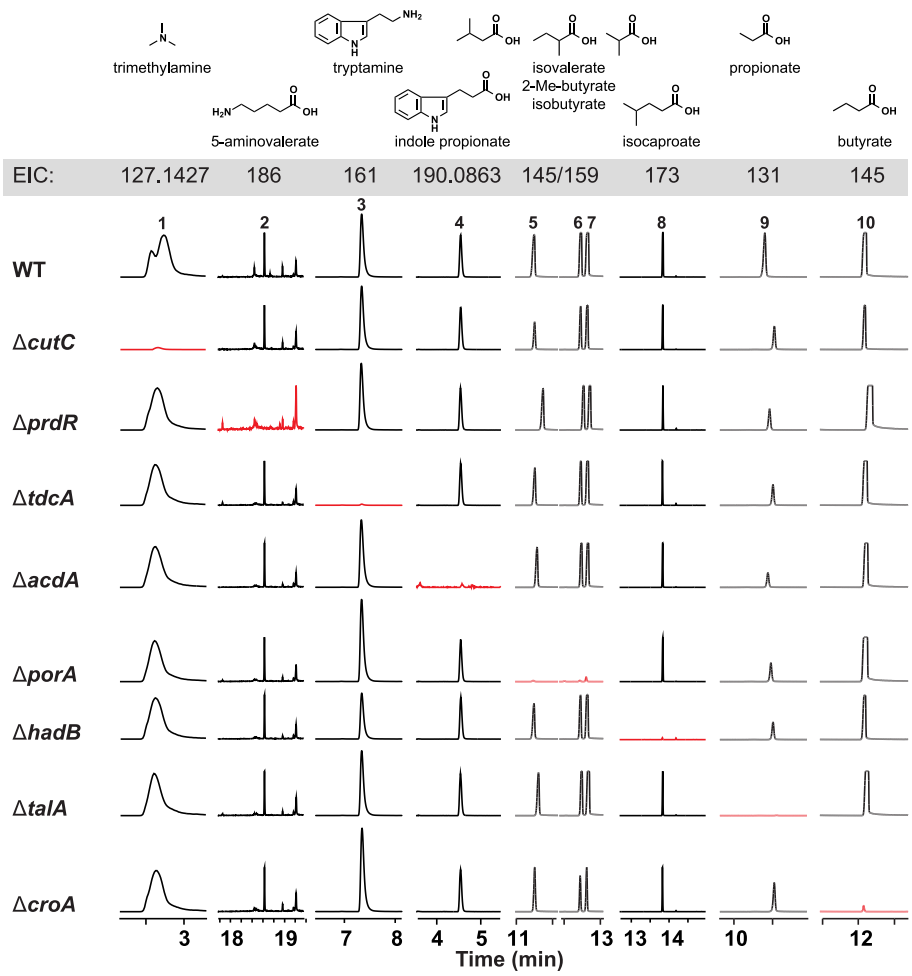



**Fig. 3. *C. sporogenes* mutants exhibit specific loss of metabolite production in vitro.** WT and mutant strains of *C. sporogenes* were cultured individually with the pathway substrates, and metabolites were assayed by LC-MS and GC-MS. Extracted ion chromatogram windows corresponding to each of the pathway products are shown to compare the metabolic output of each strain; ion counts (*y* axis) are on the same scale within a column. Full traces are displayed in fig. S5. Traces in red show the metabolite whose production is blocked by the mutation indicated at the beginning of each row. Each mutant is deficient in the production of the corresponding pathway product but proficient in the production of all other pathway products. EIC, extracted-ion chromatogram.

plasmids sequentially, we failed to get any viable colonies after introducing the second vector that harbors the Cas9 coding sequence. Reasoning that the efficiency of the second conjugation step could be the source of failure, we lengthened the donor-acceptor cocultivation step of the second conjugal transfer from 24 to 72 hours; this optimized protocol yielded reproducible, high-efficiency mutations at multiple test loci (see supplementary text for a more detailed description of the method's development).

### Constructing mutants to validate *C. sporogenes* pathways in vitro

We used the Cas9-based genetic system to construct deletion mutants in each of the 10 pathways. In each case, our repair template effected the removal of an 80– to 150–base pair (bp) portion of the targeted gene in the *C. sporogenes* chromosome (Fig. 2 and fig. S4); for simplicity, the resulting mutants [e.g., Δ*porA* (coding sequence 330-409)] are referred to simply as Δ*porA* (see table S4 for a list of deleted regions). We cultured each strain in vitro and analyzed culture extracts by liquid chromatography–mass spectrometry (LC-MS) or gas chromatography (GC)–MS (depending on the analyte), yielding the following conclusions: (i) The production of each of the 10 molecules is blocked by the corresponding pathway mutant (Fig. 3 and fig. S5), validating our prediction set; an important exception is detailed below in (iii). (ii) Deleting one pathway does not appreciably alter the production of other molecules nor does it affect growth rate (fig. S6), indicating that these pathways are dispensable and function independently in vitro. (iii) Our predicted genetic locus for the branched short-chain fatty acids (SCFAs) isobutyrate, 2-methylbutyrate, and isovalerate was incorrect. In other organisms [e.g., *Streptomyces avermitilis* (*20*)], the branched-chain α-ketoacid dehydrogenase complex (BCKDH) converts branched-chain amino acids (BCAAs) into branched SCFAs. To our surprise, deletion of the BCKDH gene *CLOSPO_03305* did not affect branched SCFA production. The Δ*porA* mutant—a gene we previously showed to be involved in the oxidative catabolism of aromatic amino acids (*10*)—was deficient in the production of all three branched SCFAs (Fig. 3). PorA is a member of the pyruvate:ferredoxin oxidoreductase (PFOR) superfamily; like the BCKDH, PFOR enzymes are thiamine-pyrophosphate dependent, but they harbor an array of iron-sulfur clusters for electron transfer in place of lipoate and flavin and reduce ferredoxin or flavodoxin instead of nicotinamide adenine dinucleotide (oxidized form) ($NAD^+$) (*21*). Although PFOR superfamily members are known to use pyruvate in anaerobic bacteria (generating acetate),

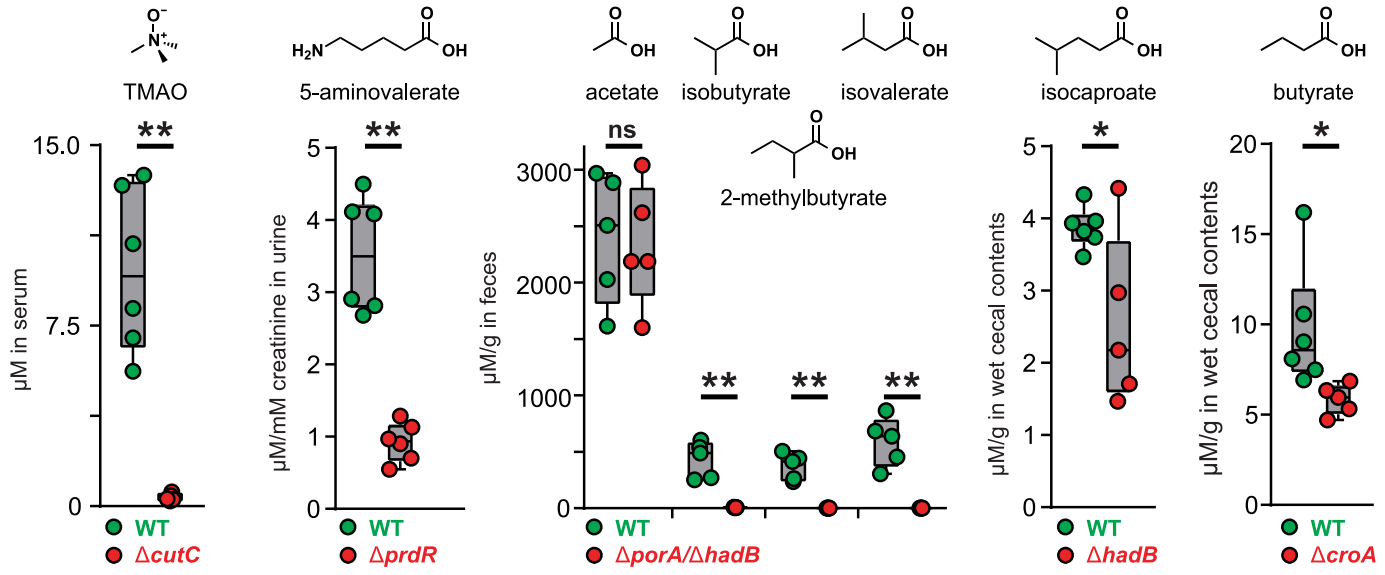



**Fig. 4. Genetic depletion of metabolites in vivo by colonization with WT versus mutant strains of *C. sporogenes*.** Germ-free mice were monocolonized with WT *C. sporogenes* or the Δ*cutC*, Δ*prdR*, Δ*porA*/Δ*hadB*, Δ*hadB*, or Δ*croA* mutants ($n \geq 5$ mice per group). Metabolite levels were altered in the host by mutation of the corresponding pathway. Error bars indicate standard deviation. *$P < 0.05$, **$P < 0.01$, and ns is not significant.

their use of branched-chain α-keto acids derived from BCAAs constitutes a noncanonical pathway for branched SCFA production, having only been previously observed in Archaea (*22*, *23*). We found multiple *porA* homologs in the genomes of other human gut isolates that are highly transcribed in human stool metatranscriptomes, suggesting that this pathway could be a substantial contributor to the host branched SCFA pool (Figs. 1 and 3 and fig. S2).

### In vivo modulation of *C. sporogenes*–derived molecules

Having validated our target pathways in vitro, we set out to determine whether we could use these mutants to knock out the production of each pathway product in the host. For this experiment, we used a subset of five mutants. The first four were Δ*cutC*, Δ*prdR*, Δ*croA*, and Δ*hadB*; for the fifth, we took advantage of the markerless nature of our CRISPR-based genetic system to construct a Δ*porA*/Δ*hadB* double mutant, with the goal of eliminating the production of all four branched SCFAs (isobutyrate, 2-methylbutyrate, and isovalerate via Δ*porA* and isocaproate via Δ*hadB*). We monocolonized germ-free mice with wild-type (WT) *C. sporogenes* and the five mutants (Fig. 4); after 4 weeks, we sacrificed the mice and measured the concentration of each molecule in serum, urine, cecal contents, and fecal pellets. WT *C. sporogenes* and the mutant strains colonized germ-free mice at comparable levels (fig. S7). We drew three observations from these data: (i) Each metabolite (or its host metabolic product) was substantially reduced in the corresponding mutant (Fig. 4). (ii) The WT and mutant differences were smaller for isocaproate and butyrate as a result of two factors: low production by WT *C. sporogenes* relative to the native level of each molecule (usually mM) and a background level of the metabolite in mutant-colonized mice, possibly owing to a source of contaminating molecule in the chow. (iii) The WT and mutant difference was especially large for the branched SCFAs isobutyrate, 2-methylbutyrate, and isovalerate, which form a pool of >2 mM in the cecal contents of WT *C. sporogenes*–associated mice that falls to near baseline in the Δ*porA*/Δ*hadB*–associated animals. Overall, these data validate the utility and generality of using *C. sporogenes* mutants to deplete microbiome-derived molecules in the host, and they highlight that the presence of a pathway in an organism can lead to production levels that vary from native to orders of magnitude below native. Thus, the choice of a producer organism that can support the biosynthesis of native metabolite levels depends on unknown factors beyond the mere presence of a corresponding pathway.

### Immune modulatory properties of branched SCFAs

In light of our ability to knock out the production of the *C. sporogenes*–derived small molecules in vivo, we returned to our original motivation for establishing the new genetic system: to study the role of microbiome-derived molecules in mediating microbe-host interactions. We chose the branched SCFAs for two reasons: (i) Like the conventional SCFAs acetate, propionate, and butyrate, they are highly abundant in the cecum and colon (concentrations in the mM range) (*24*); but unlike the SCFAs, which are known to modulate the host immune response via GPCRs GPR41 and 43 (*1*), very little is known about the biology of the branched SCFAs. (ii) *C. sporogenes* produces them robustly; they constitute a pool of >2 mM (Fig. 4). Hypothesizing that the branched SCFAs—like conventional SCFAs—might modulate the host immune response, we colonized germ-free mice with WT *C. sporogenes* and the Δ*porA*/Δ*hadB* mutant (Fig. 5A). After 5 weeks, we sacrificed the mice, isolated immune cells from the small intestine and mesenteric lymph nodes, and analyzed them by mass cytometry. WT- and Δ*porA*/Δ*hadB*-colonized mice were similar by broad metrics of immune function (e.g., total numbers of $CD4^+$ and $CD8^+$ T cells, B cells, and innate immune cells; percentages of effector cell subpopulations; and cytokine levels). However, the Δ*porA*/Δ*hadB*-colonized mice had an increased number of immunoglobulin A (IgA)–producing plasma cells (Fig. 5, B and C, and figs. S8 and S9) and increased levels of IgA bound to the surface of a variety of innate immune cells: neutrophils, eosinophils, macrophages, plasmacytoid and conventional dendritic cells, and classical and nonclassical monocytes (Fig. 5, D and E). These results suggest a previously unrecognized role for the branched SCFAs in modulating IgA-related immune cells. There is no significant difference in the levels of free IgA and *Clostridium*-specific IgA in fecal pellets from WT-colonized versus Δ*porA*/Δ*hadB*-colonized mice (fig. S10). Thus, the branched SCFAs regulate the frequency of $IgA^+$ plasma cells in the small intestine without resulting in the accumulation of IgA in feces, in contrast to another

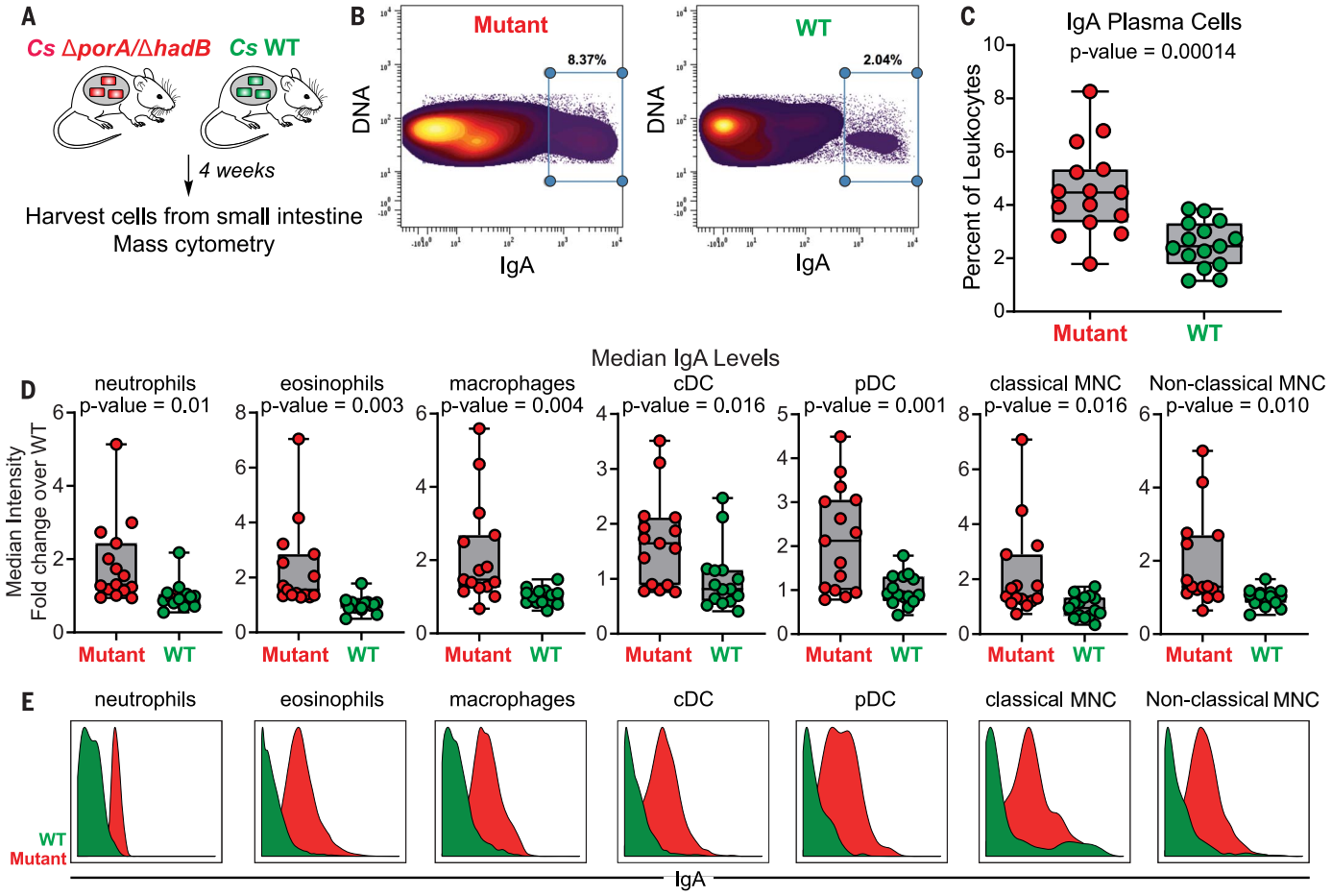



**Fig. 5. In vivo modulation of branched SCFAs reveal their IgA-related modulatory activity.** (**A**) Schematic of germ-free mouse monocolonization experiment. *Cs*, *C. sporogenes*. (**B** to **E**) Small intestinal lamina propria cells were analyzed by mass cytometry ($n$ = 15 mice per group; three biological replicates of $n$ = 5 mice per group). The frequency of $IgA^{hi}$ cells of total live immune cells are shown in (B). In (C), IgA plasma cells were quantified as a percent of live immune cells ($CD45^{+}$) after excluding cells expressing common markers of other lineages (Ly6G, Siglec-F, B220, CD3, F4/80, CD64, CD11c, CD90, CD115, NK1.1, CD49b, and FcεR1α). In (D) and (E), levels of surface-bound IgA were quantified on the indicated immune cell populations. Statistical analysis was performed using a Wilcoxon rank-sum test for all comparisons. Error bars indicate standard deviation. cDC, conventional dendritic cells; pDC, plasmacytoid dendritic cells; MNC, monocytes.

*C. sporogenes*–derived metabolite, indole propionic acid, which affects IgA levels in the lumen (*10*). These data highlight the compartmentalized nature of the IgA response and suggest multiple points of microbiome-mediated control.

## Discussion

The hundreds of microbiome-derived molecules that accumulate in circulation represent one of the most concrete modes of communication between the host and its microbiota. Little progress has been made in systematically studying their effects on host biology, owing in part to the absence of a method that enables the selective depletion of one member of this pool. One approach to addressing this challenge is to colonize germ-free mice with a WT gut bacterial species versus a metabolite-deficient mutant. Given the outsize role of *Clostridium* and related anaerobic Firmicutes in generating this pool of highly abundant metabolites, the lack of a reliable genetic system for commensal strains of *Clostridium* has been a key impediment to generalizing this approach. The system we introduce here is a first step toward that goal; it validates that genetics can be performed in a microbiome-derived *Clostridium* species rapidly, reliably, and without the need for a marker, and it demonstrates the utility of WT-mutant pairs for interrogating the host effects of microbiome-derived molecules. Given the versatility of Cas9, the key factors for generalizing this system to other *Clostridium* species are the ability to get DNA into a strain and the availability of replication origins for the plasmids that deliver the mutagenesis and repair elements. An alternative strategy would be to deliver Cas9 and the gRNA as a purified ribonucleoprotein, forgoing the need for plasmid-borne elements. Both strategies could expand the scope of Cas9-mediated mutagenesis to important but previously inaccessible Firmicutes, such as butyrate producers from *Clostridium* clusters IV and XIVa (e.g., *Faecalibacterium prausnitzii*) (*25*), the bile acid metabolizers *Clostridium scindens* and *Clostridium hylemonae* (*26*), and the leanness-associated bacterium *Christensenella minuta* (*27*).

*C. sporogenes* is a prolific producer of amino acid metabolites, and our in vitro studies show that we can predictably block each pathway. However, one of the most notable observations is the difference in the concentration of each pathway product in the context of host colonization. 5-Aminovalerate, indole propionate, isobutyrate, 2-methylbutyrate, isovalerate, propionate, and trimethylamine-*N*-oxide (TMAO) are produced robustly in vivo, whereas isocaproate and butyrate are generated at 10- to 100-fold below the native concentration range. By highlighting that the mere presence of a pathway does not imply that it will function robustly in the host, our data raise questions about computational approaches that

predict metabolic function on the basis of gene or transcript levels (*28*, *29*), and they suggest the importance of understanding pathway regulation and substrate availability under conditions of colonization by a complex native community.

The conventional SCFAs acetate, propionate, and butyrate modulate host immune function and induce regulatory T cells via a pair of GPCRs, GPR41 and 43. Although they are also predominantly microbiome-derived and present at high concentration (*24*), much less is known about the branched SCFAs isobutyrate, 2-methylbutyrate, and isovalerate, which are produced by a distinct pathway from a different set of precursors, BCAAs. Isovalerate was recently shown to be a ligand for Olfr558, a GPCR in enterochromaffin cells that controls the secretion of serotonin (*30*). Our data uncover a role for the branched SCFA pathway in modulating IgA-related immune cells.

## Materials and methods summary

### *Constructing* C. sporogenes *mutants*

The coding sequence of Cas9 was assembled with pMTL83153 using Instant Sticky-end Ligase (NEB). The gRNA and repair templates were assembled with pMTL82254. The repair template consists of two 700- to 1200-bp sequences flanking the ~100-bp *C. sporogenes* genome sequence targeted for Cas9 excision. To construct a *C. sporogenes* mutant, pMTL82254 harboring the gRNA and repair template was introduced into *C. sporogenes* via *E. coli* conjugation. Next, a second conjugation (~72 hours) was performed to introduce pMTL83153 harboring the Cas9 coding sequence. *C. sporogenes* colonies typically appeared after 36 to 48 hours. Genomic DNA was isolated from at least 16 colonies, and mutations were verified by sequencing.

### *Quantification of* C. sporogenes *metabolites using LC-MS or GC-MS*

A single colony of WT or mutant *C. sporogenes* was inoculated in TYG (tryptone–yeast extract–glucose) broth at 37°C for 48 hours in an anaerobic chamber. The conversion of choline to trimethylamine (TMA) was quantified using a previously published protocol (*14*). The serum TMAO level was quantified using Agilent 6570 LC-QQQ by comparing its area under the curve (AUC) with that of d9-TMAO. The metabolites 5-aminovalerate, indole propionate, propionate, butyrate, isovalerate, 2-methylbutyrate, isobutyrate, and isocaproate were detected and quantified using previously published protocols (*3*, *10*). The creatinine concentration of each urine sample was measured using a Creatinine Assay Kit (ab204537).

### *Germ-free mouse experiments*

All experiments were performed using germ-free Swiss Webster mice (male, 6 to 10 weeks of age, $n$ = 5 or 6 mice per group) originally from Taconic Biosciences. Germ-free mice were maintained in gnotobiotic isolators and monocolonized with an overnight culture of either WT or mutant *C. sporogenes* in TYG medium [200 μl; ~$1 \times 10^7$ colony forming units (CFU)]. For quantifying *C. sporogenes*–derived small molecules in vivo, mice were maintained on standard chow (LabDiet 5K67). After 4 weeks of colonization, mice were euthanized humanely by $CO_2$ asphyxiation. Blood was collected by cardiac puncture, and serum was prepared using serum separator tubes from BD (cat. #365967). For mass cytometry analysis, mice were maintained on a high-protein chow (Teklad TD.90018).

### *Preparation of single-cell suspensions from the small intestine for mass cytometry*

Small intestines were harvested from animals and placed in cold RPMI (Roswell Park Memorial Institute) medium. Intestines were filleted open and washed two times in phosphate-buffered saline (PBS). Tissues were cut, washed, and digested in complete RPMI supplemented with Liberase (Roche) and DNAse I (Sigma-Aldrich). Digestion was quenched with complete RPMI, and cells were washed with PBS and labeled with cisplatin viability dye for mass cytometry analysis. The viability stain was quenched, and the cells were washed twice before fixation with 1.5% paraformaldehyde (Electron Microscopy Sciences). Cells were then washed twice and stored at −80°C for subsequent mass cytometry analysis. A summary of all mass cytometry antibodies, reporter isotopes, and concentrations used for analysis can be found in table S3, and a detailed protocol for mass cytometry analysis is available in the supplementary materials.

## REFERENCES AND NOTES


1. M. G. Rooks, W. S. Garrett, Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* **16**, 341–352 (2016). doi: 10.1038/nri.2016.42; pmid: 27231050
2. P. M. Smith *et al*., The microbial metabolites, short-chain fatty acids, regulate colonic $T_{reg}$ cell homeostasis. *Science* **341**, 569–573 (2013). doi: 10.1126/science.1241165; pmid: 23828891
3. Y. Furusawa *et al*., Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446–450 (2013). doi: 10.1038/nature12721; pmid: 24226770
4. L. Zhao *et al*., Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science* **359**, 1151–1156 (2018). doi: 10.1126/science.aao5774; pmid: 29590046
5. A. S. Devlin *et al*., Modulation of a circulating uremic solute via rational genetic manipulation of the gut microbiota. *Cell Host Microbe* **20**, 709–715 (2016). doi: 10.1016/j.chom.2016.10.021; pmid: 27916477
6. S. K. Mazmanian, C. H. Liu, A. O. Tzianabos, D. L. Kasper, An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107–118 (2005). doi: 10.1016/j.cell.2005.05.007; pmid: 16009137
7. K. A. Romano *et al*., Metabolic, epigenetic, and transgenerational effects of gut bacterial choline consumption. *Cell Host Microbe* **22**, 279–290.e7 (2017). doi: 10.1016/j.chom.2017.07.021; pmid: 28844887
8. C. M. Thomas *et al*., Histamine derived from probiotic *Lactobacillus reuteri* suppresses TNF via modulation of PKA and ERK signaling. *PLOS ONE* **7**, e31951 (2012). doi: 10.1371/journal.pone.0031951; pmid: 22384111
9. J. T. Heap, O. J. Pennington, S. T. Cartman, G. P. Carter, N. P. Minton, The ClosTron: A universal gene knock-out system for the genus *Clostridium*. *J. Microbiol. Methods* **70**, 452–464 (2007). doi: 10.1016/j.mimet.2007.05.021; pmid: 17658189
10. D. Dodd *et al*., A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* **551**, 648–652 (2017). doi: 10.1038/nature24661; pmid: 29168502
11. S. R. Elsden, M. G. Hilton, J. M. Waller, The end products of the metabolism of aromatic amino acids by clostridia. *Arch. Microbiol.* **107**, 283–288 (1976). doi: 10.1007/BF00425340; pmid: 1275638
12. S. R. Elsden, M. G. Hilton, Volatile acid production from threonine, valine, leucine and isoleucine by clostridia. *Arch. Microbiol.* **117**, 165–172 (1978). doi: 10.1007/BF00402304; pmid: 678022
13. B. B. Williams *et al*., Discovery and characterization of gut microbiota decarboxylases that can produce the neurotransmitter tryptamine. *Cell Host Microbe* **16**, 495–503 (2014). doi: 10.1016/j.chom.2014.09.001; pmid: 25263219
14. A. Martínez-del Campo *et al*., Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *mBio* **6**, e00042-15 (2015). doi: 10.1128/mBio.00042-15; pmid: 25873372
15. U. C. Kabisch *et al*., Identification of D-proline reductase from *Clostridium sticklandii* as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a proprotein. *J. Biol. Chem.* **274**, 8445–8454 (1999). doi: 10.1074/jbc.274.13.8445; pmid: 10085076
16. J. Kim, M. Hetzel, C. D. Boiangiu, W. Buckel, Dehydration of (*R*)-2-hydroxyacyl-CoA to enoyl-CoA in the fermentation of α-amino acids by anaerobic bacteria. *FEMS Microbiol. Rev.* **28**, 455–468 (2004). doi: 10.1016/j.femsre.2004.03.001; pmid: 15374661
17. M. Vital, A. C. Howe, J. M. Tiedje, Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio* **5**, e00889 (2014). doi: 10.1128/mBio.00889-14; pmid: 24757212
18. S. Nayfach, M. A. Fischbach, K. S. Pollard, MetaQuery: A web server for rapid annotation and quantitative analysis of specific genes in the human gut microbiome. *Bioinformatics* **31**, 3368–3370 (2015). doi: 10.1093/bioinformatics/btv382; pmid: 26104745
19. L. A. David *et al*., Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014). doi: 10.1038/nature12820; pmid: 24336217
20. C. D. Denoya *et al*., A second branched-chain alpha-keto acid dehydrogenase gene cluster (*bkdFGH*) from *Streptomyces avermitilis*: Its relationship to avermectin biosynthesis and the construction of a *bkdF* mutant suitable for the production of novel antiparasitic avermectins. *J. Bacteriol.* **177**, 3504–3511 (1995). doi: 10.1128/jb.177.12.3504-3511.1995; pmid: 7768860
21. E. Chabrière *et al*., Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate. *Nat. Struct. Biol.* **6**, 182–190 (1999). doi: 10.1038/5870; pmid: 10048931
22. J. Heider, X. Mai, M. W. Adams, Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. *J. Bacteriol.* **178**, 780–787 (1996). doi: 10.1128/jb.178.3.780-787.1996; pmid: 8550513
23. K. Ma, A. Hutchins, S. J. Sung, M. W. Adams, Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9608–9613 (1997). doi: 10.1073/pnas.94.18.9608; pmid: 9275170
24. J. H. Cummings, E. W. Pomare, W. J. Branch, C. P. Naylor, G. T. Macfarlane, Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28**, 1221–1227 (1987). doi: 10.1136/gut.28.10.1221; pmid: 3678950
25. S. E. Pryde, S. H. Duncan, G. L. Hold, C. S. Stewart, H. J. Flint, The microbiology of butyrate formation in the human colon. *FEMS Microbiol. Lett.* **217**, 133–139 (2002). doi: 10.1111/j.1574-6968.2002.tb11467.x; pmid: 12480096
26. J. M. Ridlon, S. C. Harris, S. Bhowmik, D.-J. Kang, P. B. Hylemon, Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* **7**, 22–39 (2016). doi: 10.1080/19490976.2015.1127483; pmid: 26939849
27. J. K. Goodrich *et al*., Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014). doi: 10.1016/j.cell.2014.09.053; pmid: 25417156

28. M. G. I. Langille *et al*., Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013). doi: 10.1038/nbt.2676; pmid: 23975157
29. J. Kaminski *et al*., High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLOS Comput. Biol.* **11**, e1004557 (2015). doi: 10.1371/journal.pcbi.1004557; pmid: 26682918
30. N. W. Bellono *et al*., Enterochromaffin cells are gut chemosensors that couple to sensory neural pathways. *Cell* **170**, 185–198.e16 (2017). doi: 10.1016/j.cell.2017.05.034; pmid: 28648659
31. A. de Jong, H. Pietersma, M. Cordes, O. P. Kuipers, J. Kok, PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics* **13**, 299 (2012). doi: 10.1186/1471-2164-13-299; pmid: 23512433

## ACKNOWLEDGMENTS

We are deeply indebted to members of the Fischbach group for helpful suggestions and comments on the manuscript. **Funding:** This work was supported by NSF Graduate Research Fellowship DGE-114747 (W.V.T.). This work was also supported by an HHMI-Simons Faculty Scholar Award (M.A.F.); a Fellowship for Science and Engineering from the David and Lucile Packard Foundation (M.A.F.); an Investigators in the Pathogenesis of Infectious Disease award from the Burroughs Wellcome Foundation (M.A.F.); an award from BASF (M.A.F.); an award from the Leducq Foundation (M.A.F.); NIH grants DK110174 (M.A.F.), DK113598 (M.A.F.), DK101674 (M.A.F. and J.L.S.), OD023056 (M.H.S.), OD018040 (UCSF, to acquire the mass cytometer used in this study), DK110335 (D.D.), DK085025 (J.L.S.), and 1DP2HD101401-01 (C.-J.G.); the Stanford Microbiome Therapies Initiative (M.A.F., J.L.S., and D.D.); the Chan Zuckerberg Biohub (M.A.F., M.H.S., and J.L.S.); Stanford's ChEM-H Institute (C.R.F.); the Parker Institute for Cancer Immunotherapy (M.H.S.); the Human Frontier Science Program LT000493/2018-L (K.N.); and the Fellowship of Astellas Foundation for Research on Metabolic Disorders (K.N.). **Author contributions:** Conceptualization: C.-J.G., D.D., J.L.S., M.H.S., and M.A.F.; Investigation: C.-J.G., B.M.A., K.J.H., W.V.T., S.H., K.N., and C.R.F.; Writing – original draft: C.-J.G. and M.A.F.; Writing – review and editing: B.M.A., K.J.H., D.D., W.V.T., S.H., K.N., C.R.F., J.L.S., and M.H.S.; Funding acquisition and supervision: J.L.S., M.H.S., and M.A.F. **Competing interests:** M.A.F. is a co-founder and director of Federation Bio. J.L.S. is a co-founder of Novome Biotechnologies and of January, Inc. He serves on the scientific advisory board of Second Genome and Kaleido Biosciences. **Data and materials availability:** Data described in the paper are present in the main text or archived in the supplementary materials.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/eaav1282/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S6
References (*32*–*42*)

View/request a protocol for this paper from *Bio-protocol*.

16 August 2018; resubmitted 14 August 2019
Accepted 23 October 2019
10.1126/science.aav1282

RESEARCH ARTICLE SUMMARY

MICROBIOTA

# A metagenomic strategy for harnessing the chemical repertoire of the human microbiome

**Yuki Sugimoto*, Francine R. Camacho*, Shuo Wang, Pranatchareeya Chankhamjon, Arman Odabas, Abhishek Biswas, Philip D. Jeffrey, Mohamed S. Donia†**

**INTRODUCTION:** The human microbiome has been correlated with several health and disease conditions, but the molecular mechanisms underlying these correlations remain largely unexplored. Biologically active small molecules that are produced by the human microbiome offer an important route for exploring these mechanisms because they often mediate important microbe-microbe and microbe-host interactions. In bacterial genomes, small-molecule biosynthetic genes are usually encoded in distinct clusters known as biosynthetic gene clusters (BGCs), which enables scientists to use computational tools for recognizing them and predicting their products. Here, we present a hybrid strategy that uses computational and synthetic biology tools for discovering microbiome-encoded small molecules.

**RATIONALE:** Previous efforts to discover small-molecule BGCs from the human microbiome relied mainly on analyzing genomic data of sequenced bacterial isolates. Although this approach has revealed the enormous and largely untapped diversity of microbiome-encoded BGCs, it fails to report on the biosynthetic potential of members of the human microbiome that have not yet been cultured or isolated: the majority of species in metagenomic sequencing data. Therefore, we sought to develop a computational algorithm that discovers small-molecule BGCs directly in complex metagenomic sequencing data of the human microbiome: metagenomic identifier of biosynthetic gene clusters (*MetaBGC*). First, high-performance probabilistic models for identifying homologs of a biosynthetic enzyme of interest are built specifically for use with complex metagenomic datasets (*MetaBGC-Build*). Next, these models are used to identify biosynthetic genes in thousands of metagenomic datasets of the human microbiome at the single-read level (*MetaBGC-Identify*). Finally, identified biosynthetic reads are quantified in the entire cohort of samples (*MetaBGC-Quantify*) and clustered into biosynthetic read bins on the basis of their abundance profiles across samples (*MetaBGC-Cluster*). To evaluate the utility of this approach, we used it to discover BGCs for type II polyketides, a clinically relevant class of small molecules, directly from metagenomic sequencing data of the human microbiome.

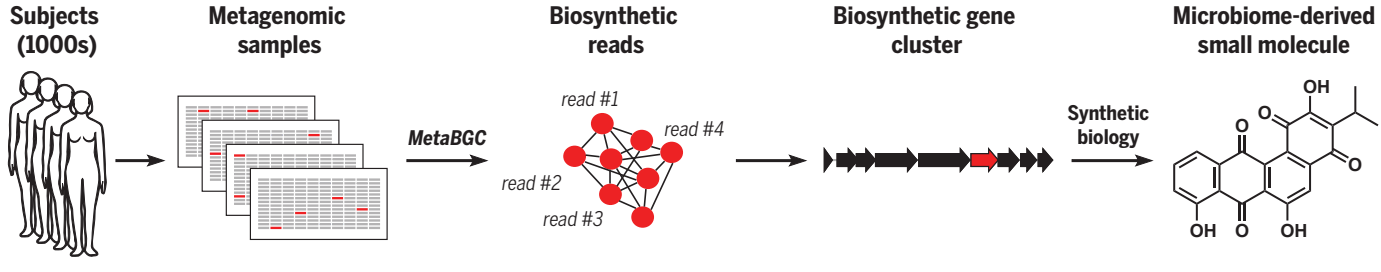
**RESULTS:** We applied *MetaBGC* to 3203 metagenomic samples of the human microbiome originating from Western (subjects from the United States, Spain, and Denmark) and non-Western (subjects from China and Fiji) populations and from every major human body site (gut, mouth, skin, and vagina). Overall, we discovered 13 complete BGCs that potentially encode type II polyketides; eight of these were encoded by diverse bacterial isolates of the human microbiome in a strain-specific manner and five could not be assigned to any sequenced species. Type II polyketide BGCs are found in three major human body sites, gut, mouth, and skin, and at least six of them are transcribed under host-colonization conditions and widely distributed in different human populations (e.g., 46% of healthy subjects from the United States encode at least one BGC in their gut, oral, or skin microbiome). Next, we selected two of the identified BGCs for experimental characterization, one from the oral microbiome and another from the gut microbiome. We used a synthetic biology strategy in which metagenomically discovered BGCs are genetically engineered and expressed in various heterologous hosts without the need for cultivation of the native producer. Using this strategy, we successfully purified and solved the structures of five new type II polyketide molecules as the products of the two characterized BGCs. Finally, we show that two of the discovered molecules exert strong antibacterial activities against members of the human microbiome that occupy the same niche as their producer, implying a possible role in microbe-microbe competition.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aax9176

**CONCLUSION:** We developed a hybrid strategy that combines computational and experimental techniques for discovering and characterizing small-molecule BGCs directly from complex datasets of the human microbiome. Using this strategy, we discovered that a clinically relevant class of molecules, type II polyketides, are widely encoded in the human microbiome and that human microbiome–derived polyketides resemble in structure and biological activity clinically used ones. Our approach is generally applicable to other classes of small molecules and can be used to systematically unveil the chemical potential of the human microbiome—a goal that is useful for both mechanistic microbiome explorations and drug discovery. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: donia@princeton.edu
Cite this article as Y. Sugimoto *et al.*, *Science* **366**, eaax9176 (2019). DOI: 10.1126/science.aax9176




**Small-molecule discovery from the human microbiome.** A hybrid computational and synthetic biology approach was developed to discover small molecules encoded by the human microbiome. Biosynthetic gene clusters were discovered at the single metagenomic read level from thousands of samples using a new algorithm, *MetaBGC*, and expressed in a synthetic biology platform to yield previously unknown small molecules.

RESEARCH ARTICLE

MICROBIOTA

# A metagenomic strategy for harnessing the chemical repertoire of the human microbiome

Yuki Sugimoto[1]*, Francine R. Camacho[2]*, Shuo Wang[3], Pranatchareeya Chankhamjon[1], Arman Odabas[1], Abhishek Biswas[1,4], Philip D. Jeffrey[1], Mohamed S. Donia[1]†

**Extensive progress has been made in determining the effects of the microbiome on human physiology and disease, but the underlying molecules and mechanisms governing these effects remain largely unexplored. Here, we combine a new computational algorithm with synthetic biology to access biologically active small molecules encoded directly in human microbiome–derived metagenomic sequencing data. We discover that members of a clinically used class of molecules are widely encoded in the human microbiome and that they exert potent antibacterial activities against neighboring microbes, implying a possible role in niche competition and host defense. Our approach paves the way toward a systematic unveiling of the chemical repertoire encoded by the human microbiome and provides a generalizable platform for discovering molecular mediators of microbiome-host and microbiome-microbiome interactions.**

The human microbiome harbors thousands of bacterial species, varies in composition between different sites of the human body and between individuals, and has been correlated with several diseases (*1*–*5*). Recently, scientists have begun to examine these correlations at a mechanistic level, often by identifying and characterizing microbiome-derived small molecules that are responsible for a specific phenotype (*6*, *7*). These molecules may mediate their effects directly, by targeting human cells or receptors (microbe-host interactions) (*8*–*14*), or indirectly, by affecting other members of the microbiome in a competitive or collaborative manner (microbe-microbe interactions) (*15*–*18*). With the importance of microbiome-derived small molecules becoming apparent, there is a dire need to develop systematic approaches for discovering and characterizing them.

We previously undertook a systematic approach to describe the general capacity of the microbiome to produce small molecules (*15*) by identifying the biosynthetic gene cluster (BGC) repertoire encoded in genomes of thousands of bacterial strains isolated originally from humans and characterizing the structure and biological activity of a subset of their products (*9*, *15*). Although this approach revealed an enormous and largely unexplored capacity of the microbiome to produce small molecules, it suffered from one major limitation: Its starting point relied on analyzing assembled genomes from readily cultured and sequenced isolates of the human microbiome. Therefore, this approach misses BGCs encoded in sequenced clinical samples but not in genomes of previously isolated bacteria. This limitation is noteworthy for three main reasons.

First, most large culturing efforts have focused on human gut microbiome samples, especially those from healthy and Western populations, resulting in a relatively limited representation of reference genomes from other human body sites and from diseased and non-Western cohorts (*19*–*22*). By contrast, thousands of metagenomic sequencing datasets have been produced from all human body sites (*1*, *2*), for cohorts of several human diseases (*3*–*5*, *23*–*25*), and from non-Western populations (*26*, *27*). Second, recent metagenomic binning studies revealed thousands of new bacterial species in the human microbiome, representing rare, not yet cultivated or not yet sequenced members (*28*–*30*). These new species expanded not only the taxonomical space of the human microbiome, but also its functional capacity beyond what has been previously observed in the genomes of cultured isolates. Finally, BGCs are commonly encoded on mobile elements as strain-specific traits (*15*) and capturing them would require deep sampling within single species, which is not often the goal in culturing efforts. It is evident that our previous approach, which relied solely on discovering BGCs from genomes of cultured isolates, is limited in its ability to fully represent the biosynthetic potential of the collective human microbiome.

[1]Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. [2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA. [3]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA. [4]Research Computing, Office of Information Technology, Princeton University, Princeton, NJ 08544, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: donia@princeton.edu

### A read-based algorithm for the detection of BGCs in metagenomic data

Realizing the above limitation, we set out to develop an algorithm that allows for the identification of small-molecule BGCs directly in human microbiome–derived metagenomic sequencing data without the need for bacterial cultivation or sequencing; we called this algorithm *MetaBGC* (metagenomic identifier of biosynthetic gene clusters). Our algorithm needed to fulfill three main criteria: (i) be able to detect BGCs de novo from metagenomic sequencing data without any prior knowledge of bacterial species in the sample and therefore show no bias toward sequenced isolates; (ii) be able to detect new BGCs at the single metagenomic read level [~100 base pairs (bp)] and therefore show no bias toward abundant or easy-to-assemble species that often dominate the metagenomic assemblies; and (iii) be computationally efficient without sacrificing sensitivity so that it can be used to profile thousands of metagenomic samples simultaneously.

A simple and fast method for detecting homologs of a given protein family is through profile hidden Markov models (pHMMs) (*31*, *32*). In pHMMs, the probability of finding a given amino acid, insertion, or deletion is calculated at each position of an alignment of interest (training set) and then used to construct a probability profile. New sequences (search set) are then scored on the basis of their fit to that profile. pHMMs are typically constructed from full-length proteins in the training set, whereas most Illumina-based metagenomic reads are ~100 bp in length (~33 amino acids). When used in a search set, these short sequences are aligned locally to different regions of their respective pHMMs. Depending on the sequence complexity and conservation at these local regions, varying specificities and sensitivities may be obtained across a given pHMM. Therefore, we sought to adapt pHMMs for use in metagenomic applications by developing what we refer to here as segmented pHMMs (spHMMs). spHMMs are built on 30–amino acid segments of aligned, full-length protein homologs, resulting in probability profiles with lengths that match those of the sequences in the search set. Because of the small size of each segment, spHMMs are able to distinguish regions of the aligned proteins that are of high complexity from repetitive or low-complexity ones that are commonly found in sequences, and regions of high conservation among homologs from ones that are more variable. This differentiation would result in models of varying performances depending on their intervals (Fig. 1A). Models are evaluated using a synthetic dataset and spHMMs that perform poorly (with a high false-positive, high false-negative, or low true-positive rate) are eliminated. This first module of the

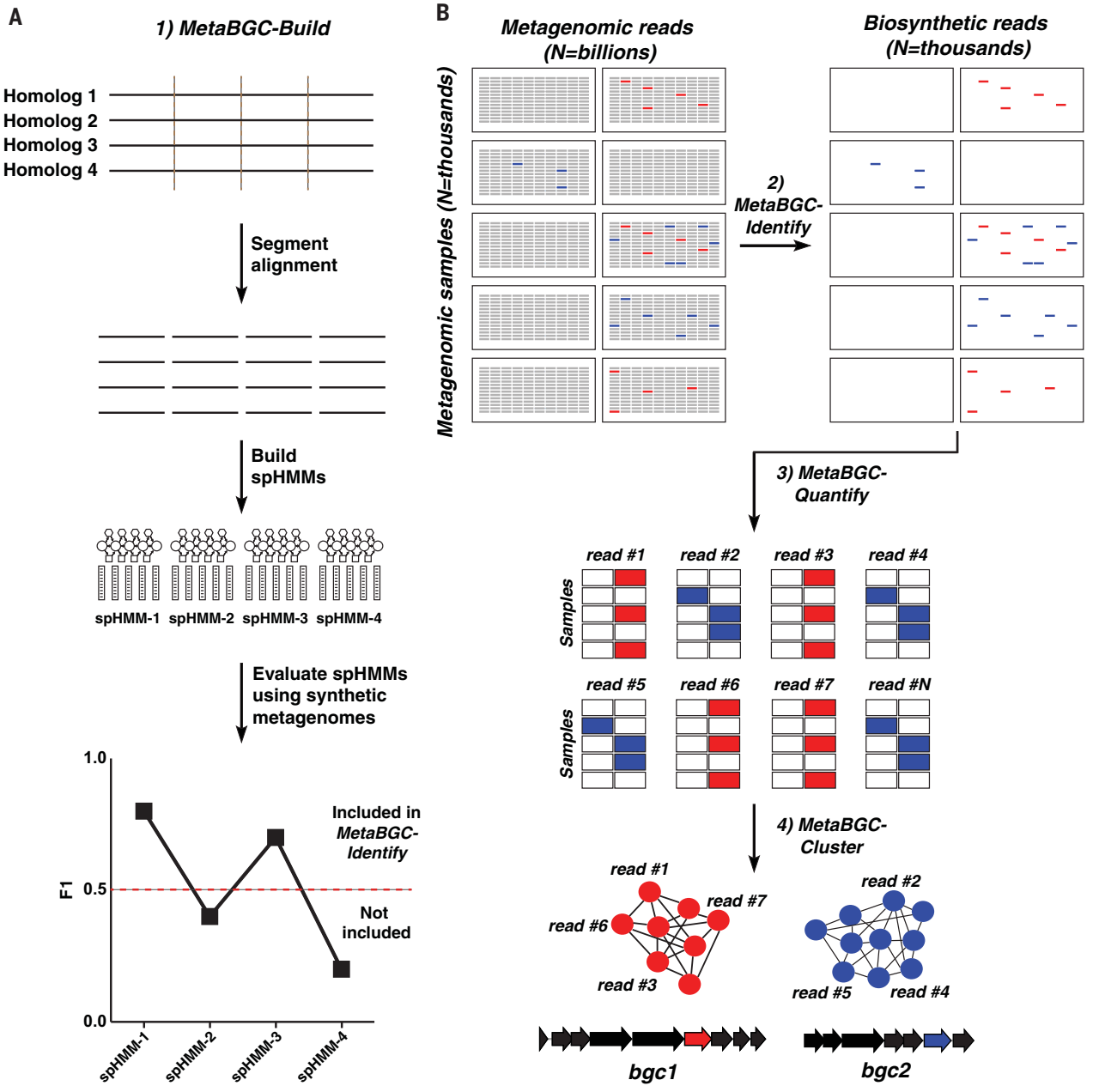



**Fig. 1. Overview of *MetaBGC*.** (**A**) The first step of *MetaBGC* is the development of high-performance spHMMs that are specific for a given class of BGCs (*MetaBGC-Build*). Homologs of the protein family of interest are aligned and the alignment is segmented into 30–amino acid fragments with a 10–amino acid window shift. spHMMs are built using the segmented alignments, which are then evaluated using synthetic metagenomes composed of a predominantly negative background and spiked with positive BGCs. spHMMs with F1 scores ≥0.5 form the basis of *MetaBGC-Identify* and those with F1 scores <0.5 are excluded. (**B**) The remaining three steps of *MetaBGC* are as follows: (i) *MetaBGC-identify* detects biosynthetic reads from complex metagenomic datasets of the human microbiome using spHMMs [as described in (A)]; (ii) unique (nonredundant) biosynthetic reads are then quantified in all samples (*MetaBGC-Quantify*) and an abundance profile is generated for each read; and (iii) nonredundant biosynthetic reads are finally clustered on the basis of their abundance profiles (*MetaBGC-Cluster*) to produce read "bins" that originate from specific BGCs. For example, biosynthetic reads colored in red or blue are detected from 10 metagenomic samples, quantified in the same samples, and finally clustered to produce two distinct bins (one red and one blue) that originate from *bgc1* or *bgc2*, respectively. Nonbiosynthetic reads are colored in gray.

algorithm is named *MetaBGC-Build*. Only high-performance spHMMs are selected for the next step and used to search metagenomic reads from clinical samples (*MetaBGC-Identify*).

Reads identified by the selected spHMMs (scored above a defined threshold) are then deemed "biosynthetic" and passed on to a third module of the algorithm: *MetaBGC-Quantify* (Fig. 1B). In this module, biosynthetic reads are de-replicated and quantified in all samples of the entire cohort(s) and an abundance matrix is generated for all nonredundant reads against all samples. Because reads that originate from a single BGC should have an even coverage across metagenomic samples, we devised a clustering strategy to produce "bins" of identified reads on the basis of their abundance profiles in different samples. In this strategy, we use DBSCAN (density-based spatial clustering of applications with noise) (*33*) to cluster reads with similar abundance

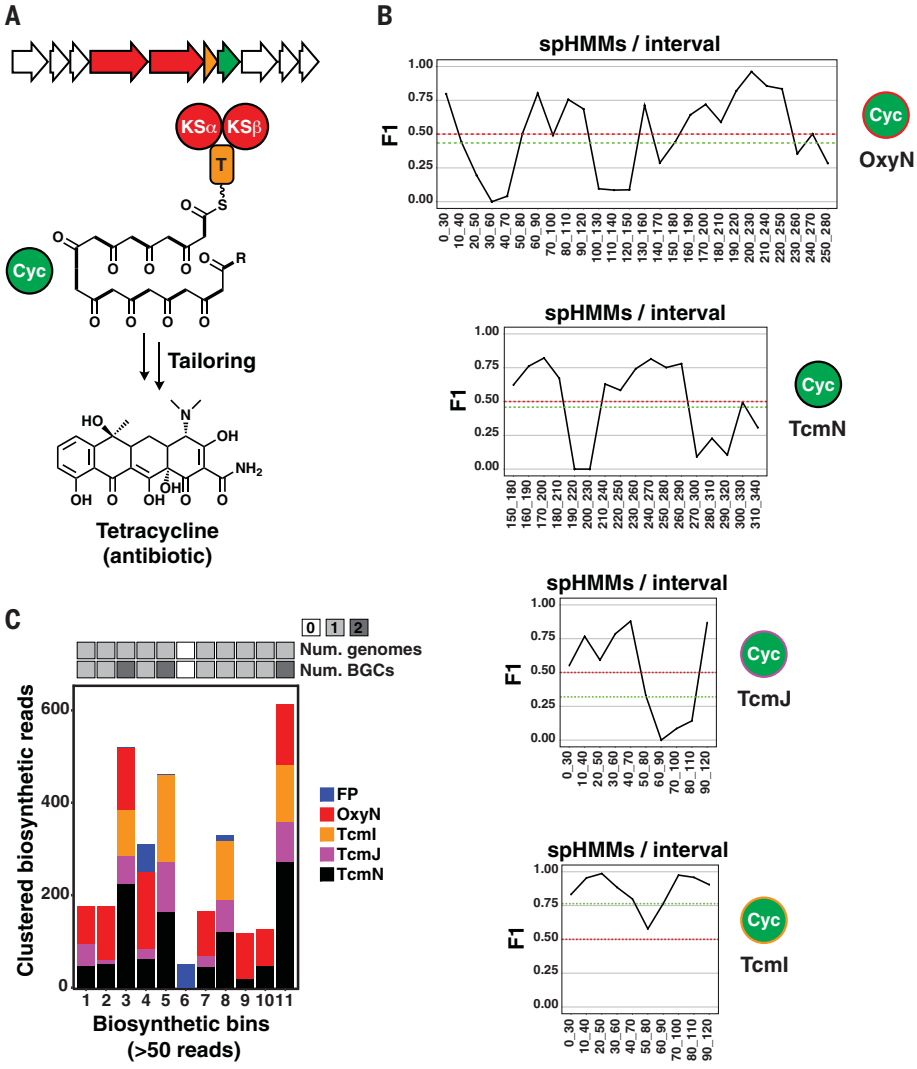



**Fig. 2. Using *MetaBGC* to detect TII-PKS BGCs in synthetic metagenomic samples.** (**A**) Typical organization of a TII-PKS BGC composed of genes encoding for two ketosynthases (KSα and KSβ) responsible for the elongation and chain-length determination of the growing polyketide chain, respectively; a thiolation domain (T), on which the growing chain is attached; and at least one type of cyclase/aromatase (Cyc), which is responsible for the cyclization of the linear chain. Through additional tailoring reactions, more elaborate structures are formed, such as the clinically used antibiotic tetracycline. (**B**) Four types of cyclases or aromatases are specifically found in TII-PKS BGCs and are chosen as protein families for *MetaBGC*. For each of the four protein families, segmented alignments corresponding to 30–amino acid intervals are used to build spHMMs. Each spHMM is then evaluated using synthetic metagenomes designed to simulate several expected compositions in human microbiome samples (synthetic dataset 1). The F1 score (*y* axis) of each spHMM per interval (*x* axis) is calculated, which is indicative of its accuracy (considering both precision and recall). Values shown represent the maximum F1 score after testing various spHMM score thresholds (fig. S8). Note that some spHMMs have very low F1 scores, whereas others have high ones (only spHMMs with F1 scores ≥0.5 were included in *MetaBGC-Identify*). The F1 score of each unsegmented cyclase pHMM is shown in green. (**C**) Stacked bar graph showing 11 bins (>50 nonredundant biosynthetic reads each) that *MetaBGC* produced when applied to the 140 synthetic metagenomes simulated in this study (synthetic dataset 1). Colors indicate the types of cyclases that the biosynthetic reads belong to within each bin. Blue indicates false-positive reads (FP). The heatmap on top indicates the number of BGCs represented per bin and the number of genomes from which these BGCs originate. Most bins represent a single BGC, and when two BGCs are represented by the same bin, they are expectedly derived from the same spiked genome. Results from synthetic dataset 2 can be found in figs. S12 to S14.

profiles across different metagenomic samples into distinct bins on the basis of their pairwise Pearson correlation distance (see the supplementary materials). This fourth module is called *MetaBGC-Cluster*. This final strategy not only decreases the total number of hits that need to be analyzed, but also presents an opportunity for a targeted assembly of the reads that originate from the same BGC and end up in the same bin. In summary, our spHMM-based algorithm identifies, quantifies, and clusters microbiome-derived BGCs at the single metagenomic read level (Fig. 1B).

### Test case for *MetaBGC*: Type II polyketide synthase BGCs

To evaluate the utility of this approach, we focused on one class of small-molecule BGCs that has never been reported from members of the human microbiome: Type II polyketide synthases (TII-PKS BGCs). TII-PKS BGCs are relatively uncommon in bacterial genomes and almost always encode small molecules with interesting biological activities (including the clinically used anticancer drug doxorubicin and the clinically used antibiotic drug tetracycline) (*34*, *35*). To first test whether TII-PKS BGCs can be identified at all in human metagenomic sequencing data, we performed de novo metagenomic assemblies on 759 samples from the Human Microbiome Project 1 (HMP-1-1) (*1*) and used a common BGC identification tool, antiSMASH, to detect TII-PKS BGCs in assembled scaffolds >5000 bp (*36*). Using this strategy, we identified six new TII-PKS BGCs in three body sites (mouth, gut, and skin) (*bgc1* to *bgc6*; see below), which indicated that this class of molecules is indeed encoded in human-derived metagenomes despite not having been reported from common isolates of the human microbiome. Motivated by this finding, we proceeded toward adapting our *MetaBGC* strategy for the systematic discovery and quantification of TII-PKS BGCs directly in human-derived metagenomic sequencing data, at the single-read level and without relying on metagenomic assemblies. This strategy would largely eliminate the contingency of discovery success on the abundance of the encoding organism or the ability to properly assemble its genome from complex metagenomic samples.

Four essential enzymes are universally present in TII-PKS BGCs: two ketosynthases (KSα and KSβ), which are responsible for the elongation of the polyketide chain and chain length determination, respectively; an acyl carrier protein (ACP), also known as a thiolation domain (T), on which the growing chain is attached through a thioester bond; and at least one of four types of cyclases or aromatases (OxyN, TcmN, TcmI, and TcmJ types), which are responsible for the final cyclization/aromatization of the polyketide chain through a series of

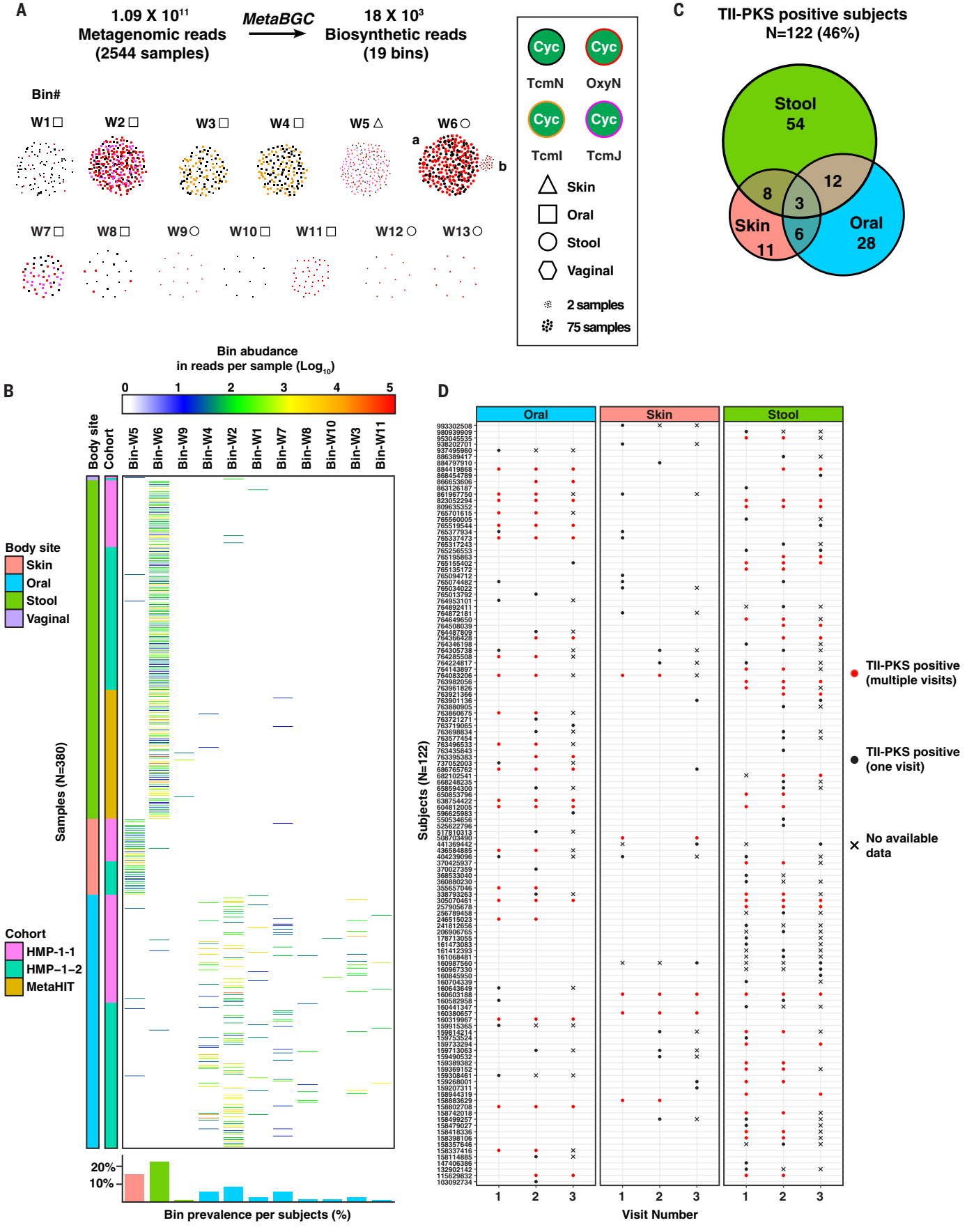

A
1.09 X 10^11
Metagenomic reads
(2544 samples)
MetaBGC
18 X 10^3
Biosynthetic reads
(19 bins)
Bin#
W1 □
W2 □
W3 □
W4 □
W5 △
W6 ○
a
b
W7 □
W8 □
W9 ○
W10 □
W11 □
W12 ○
W13 ○
Cyc
TcmN
OxyN
TcmI
TcmJ
Skin
Oral
Stool
Vaginal
2 samples
75 samples
C
TII-PKS positive subjects
N=122 (46%)
Stool
54
12
8
3
Skin
11
6
Oral
28
B
Bin abudance
in reads per sample (Log10)
0
1
2
3
4
5
Body site
Cohort
Bin-W5
Bin-W6
Bin-W9
Bin-W4
Bin-W2
Bin-W1
Bin-W7
Bin-W8
Bin-W10
Bin-W3
Bin-W11
Body site
Skin
Oral
Stool
Vaginal
Samples (N=380)
Cohort
HMP-1-1
HMP–1–2
MetaHIT
20%
10%
Bin prevalence per subjects (%)
D
Oral
Skin
Stool
Subjects (N=122)
Visit Number
1
2
3
TII-PKS positive
(multiple visits)
TII-PKS positive
(one visit)
No available
data

**Fig. 3. Testing *MetaBGC* on human microbiome–derived metagenomic samples from three large cohorts.** (**A**) *MetaBGC* results from profiling 2544 metagenomic samples of three cohorts (HMP-1-1, HMP-1-2, and MetaHIT). A selection of bins ($n$ = 13) corresponding to TII-PKS BGCs is shown, for which each point represents a nonredundant biosynthetic read that was detected by *MetaBGC*. For all other bins, see fig. S16. The color, shape, and size of each point indicate the type of cyclase to which each read belongs, the body site from which each read originated, and the number of metagenomic samples where each read was detected following the key on the right. All bins harbor biosynthetic reads from one BGC. Bins W6a and W6b were separated using a refined clustering strategy (see supplementary materials). (**B**) Heatmap indicating the abundance of each of 11 bins that correspond to complete TII-PKS BGCs (data table S6) in 380 metagenomic samples. For a similar heatmap of all other bins, see fig. S17. Bin abundance is calculated as the sum of abundances for all biosynthetic reads in a given bin (in $\log_{10}$) and is indicated on a gradient color scale as shown above (with a minimum bin abundance of 10 reads). The body site and cohort that each sample belongs to are indicated following the color code on the left. The bar graph shown at the bottom indicates the prevalence of each corresponding biosynthetic bin in subjects from the three cohorts analyzed here. Percentages are shown out of the total number of subjects with available samples at a given body site (mouth, 219 subjects; gut, 646 subjects; skin, 161 subjects). (**C**) Venn diagram indicating the number of HMP-1-1 and HMP-1-2 subjects that are positive for complete TII-PKS BGCs in one, two, or three body sites (with a minimum bin abundance of 10 reads). (**D**) Longitudinal analysis of HMP-1-1 and HMP-1-2 subjects. Samples from up to three visits of the same subject are deemed TII-PKS–positive (dot, with a minimum bin abundance of 10 reads) or TII-PKS–negative (no dot). Red dots indicate that the subject had the same TII-PKS bin at subsequent visits (longitudinal consistency); black dots indicate that the subject had a TII-PKS bin at only one visit. X indicates visits for which no metagenomic data are available.

aldol condensation reactions (Fig. 2A) (*35*, *37*). Although KS and ACP domains exist in other BGCs (e.g., iterative, TI-PKS, and TIII-PKS BGCs), cyclase domains are relatively specific to TII-PKS BGCs and can be used as a proxy for identifying them using *MetaBGC* (*37*). Therefore, we aligned selected, diverse homologs of each of the four types of TII-PKS cyclases and built spHMMs for all intervals at a window size of 30 amino acids and a window shift of 10 amino acids (see the supplementary materials, data table S1, and figs. S1 to S5).

To evaluate the performance of *MetaBGC*, we applied it to a carefully designed synthetic metagenomic dataset. We simulated 140 metagenomic samples containing 42 (low-diversity) or 126 (high-diversity) human microbiome–derived genomes that contain no TII-PKS BGCs. We then spiked in simulated reads from 10 diverse bacterial genomes that harbor in total 13 TII-PKS BGCs, none of which was part of the spHMM training sets (see the supplementary materials, fig. S6, and data table S2). Overall, this synthetic dataset (synthetic dataset 1) was constructed to simulate several conditions that are expected in human microbiome samples: very low abundance of a given BGC (~1× coverage), medium abundance (~10× coverage), several BGCs per sample, and no BGCs per sample (fig. S7). We then computed the F1 scores (harmonic mean between precision and recall) to individually evaluate each spHMM of all four cyclases. As expected, *MetaBGC-Build* revealed two types of spHMMs for each alignment: high-performance spHMMs (F1 scores ≥0.5) and low-performance spHMMs (F1 scores <0.5) (Fig. 2B). After eliminating low-performance models and tuning spHMM scores of the high-performance ones (fig. S8), we reached a final set of 40 spHMMs to be used in the *MetaBGC-Identify* module. Moreover, as expected, the number of reads detected by these models for a given cyclase positively correlates with the coverage of the spiked genomes harboring the same cyclase (fig. S9). In summary, we used synthetic metagenomic data to evaluate the performance of each of the cyclase spHMMs in *MetaBGC* and to select and tune the best-performing ones for next steps.

To evaluate the remaining components of *MetaBGC*, we subjected the identified "biosynthetic reads" from the 40 high-performance models to the quantification and clustering modules and analyzed the resulting bins for composition. In total, *MetaBGC* produced 11 bins with >50 nonredundant biosynthetic reads (range 51 to 616 reads, average 278 reads per bin), 10 of which were true-positive bins, and the smallest of them (bin 6, 51 reads) contained only false-positive reads (fig. S10). Collectively, the 10 true-positive bins harbored reads from all 37 spiked cyclases (100% cyclase recovery rate). Next, we investigated how many BGCs were represented in each of the 10 true-positive bins. Seven out of the 10 bins corresponded to reads that originated exclusively from one of seven nonredundant BGCs, whereas the remaining three bins harbored reads from two BGCs each (Fig. 2C). Each pair of BGCs that shared a bin was encoded in the same spiked genome and thus had the same coverage and representation in the metagenomic dataset and was clustered together by *MetaBGC-Cluster*. In conclusion, *MetaBGC* recovered reads from all cyclases of all spiked BGCs and correctly clustered them into their BGC- and genome-specific bins. Moreover, *MetaBGC* produced only 4% false-positive reads after clustering, 40% of which were binned together into a single, easily identifiable false-positive bin (Fig. 2C).

To test whether a wider distribution of coverages for the TII-PKS–positive genomes would affect the performance of *MetaBGC*, we simulated a new synthetic dataset (synthetic dataset 2) in which the proportion of each TII-PKS–positive and -negative genome in a given sample is sampled from a log-normal distribution and then normalized such that the sum of proportions equals one (data table S2 and fig. S11). Despite a wide range of coverages for the TII-PKS–positive genomes (0 to >1000×) in this new dataset, *MetaBGC* performed in the same manner as before: overall, it produced 12 bins with >50 nonredundant biosynthetic reads in each (10 true- and two false-positive ones), 95% true-positive reads after binning, 100% cyclase recovery rate, and 100% correct binning of cyclase reads into their corresponding BGCs and genomes (figs. S12 to S14). These promising results illustrate the validity of *MetaBGC* for the identification, quantification, and clustering of TII-PKS BGCs in human-derived metagenomes.

### Tuning of *MetaBGC* using metagenomic data from three large cohorts

After evaluating *MetaBGC* on simulated datasets of the human microbiome, we sought to tune its performance on clinically derived metagenomic data. We applied it to 2544 human-derived metagenomic samples from three large studies [the Human Microbiome Project (HMP-1-1 and HMP-1-2) and the Metagenomics of the Human Intestinal Tract Consortium, (MetaHIT) (*1*, *2*, *38*)]. These samples originated from all major human body sites (skin and airways, 293 samples; gut, 872 samples; vagina, 215 samples; and mouth, 1164 samples) and collectively harbored $1.09 \times 10^{11}$ reads. By comparing *MetaBGC* results with the National Center for Biotechnology Information (NCBI) nonredundant protein sequence database, 10 of our spHMM score cutoffs were further fine-tuned to eliminate false-positive hits resulting from genomes commonly observed in the metagenomic datasets but not initially included in our synthetic dataset, and one model was eliminated (see the supplementary materials and data table S1). Our final, most tuned algorithm detected $18 \times 10^3$ reads ($1.66 \times 10^{-7}$ hit rate), which clustered into 19 bins of at least 10 nonredundant biosynthetic reads (data tables S3 to S5).

As expected, six of these bins mapped to our initial set of six BGCs (*bgc1* to *bgc6*) and 13 appeared to originate from new ones (Fig. 3 and figs. S15 to S17). Only one bin, bin-W6, appeared to harbor two groups of reads with different abundance profiles (the first group

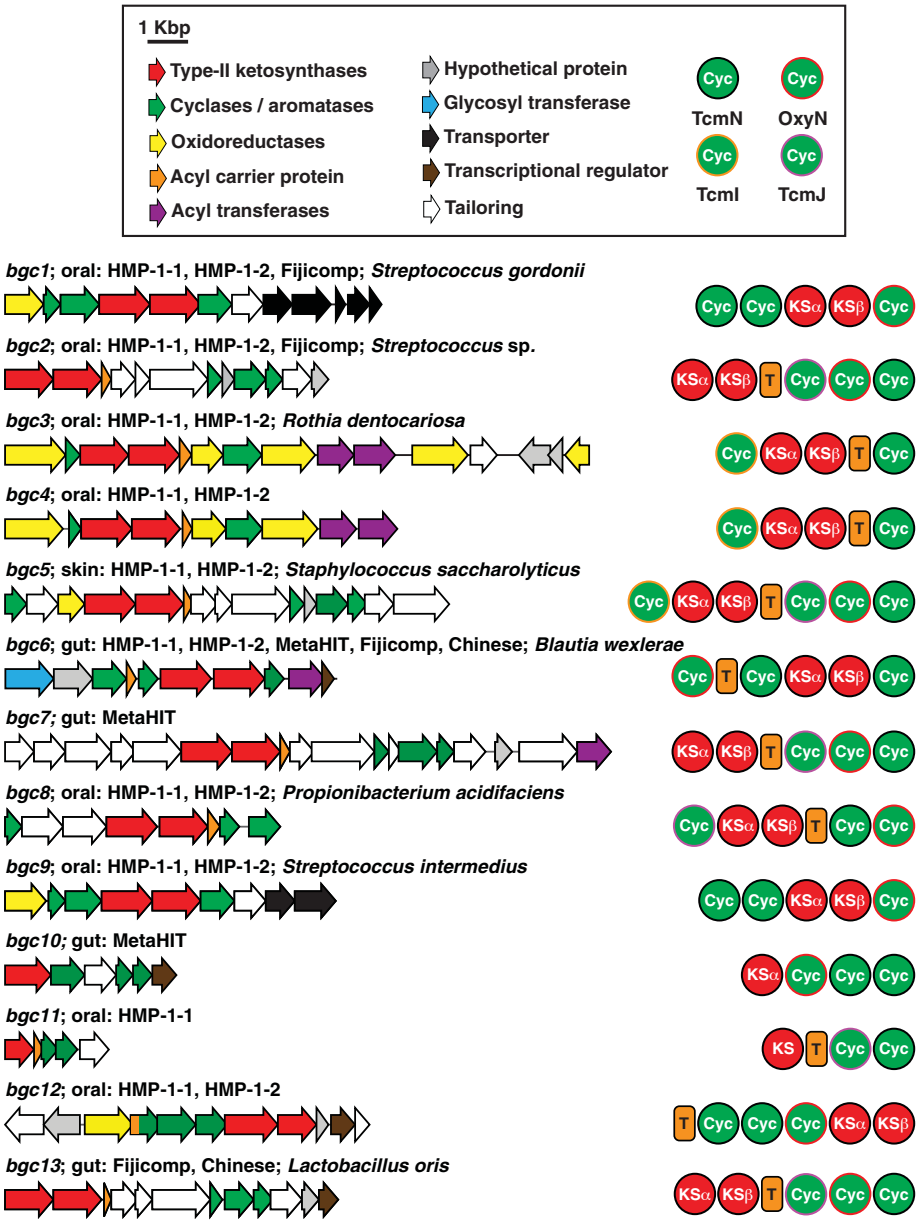



**Fig. 4. Genetic organization of human-derived TII-PKS BGCs.** Colored arrows in each BGC indicate genes following the color code on top. Typical TII-PKS biosynthetic domains (KSα, KSβ, T, and the four possible types of Cyc) are shown to the right of each BGC. The main body site where each BGC was found, the subject cohort(s) in which it was discovered, and the bacterial species with the closest match to it are also indicated. Note that several BGCs are only found in metagenomic data and have no species designation. For more details, see data table S6.

was found on average in 72 stool samples, whereas the second was found only in two stool samples), which were easily separated into bins W6a and W6b using more stringent clustering parameters (Fig. 3A, supplementary materials, and data table S5). Various assembly approaches on 14 selected samples that have the highest abundance of the 14 new bins provided 10 complete BGCs and four partial ones. Of the new, complete ones, six were clearly TII-PKS BGCs that have not been previously described (Fig. 4 and data table S6). Four of the six newly discovered TII-PKS BGCs (*bgc9* to *bgc12*) could not be identified by antiSMASH (*36*), even after fully assembling them, further highlighting the sensitivity of our read-based discovery approach. For all complete TII-PKS BGCs, whenever a cyclase read was detected by an spHMM for a given interval, this read was mapped correctly to the same interval in the fully assembled cyclase gene (see the supplementary materials, data tables S7 to S10, and fig. S18). The remaining six completely assembled bins do not represent TII-PKS BGCs but instead represent several new architectures of OxyN-containing BGCs with no previously characterized function (named here: nonpolyketide cyclase BGCs or NPCs) (fig. S15).

### TII-PKS BGCs are widespread in both Western and non-Western cohorts

We then used *MetaBGC-Quantify* to study the representation of all TII-PKS bins that produced complete BGCs in the 2544 human samples (660 subjects). With a cutoff bin abundance of 10 reads per sample (see the supplementary materials), we detected at least one TII-PKS BGC in 122 (46%) of HMP-1-2 and HMP-1-2 subjects at any body site (total of 265 subjects). Of these TII-PKS–positive subjects, 93 (76.2%) harbored TII-PKS-BGCs in the oral, gut, or skin microbiome, whereas 26 (21.3%) and three (2.5%) harbored them at two or all three body sites, respectively (Fig. 3, B and C, figs. S16 and S17, and data table S4). By analyzing longitudinal samples that originated from the same subject over different visits (collected a few months apart), we found that 52% (oral), 52% (gut), and 19% (skin) of subjects who were positive for a given TII-PKS bin at one visit were positive for the same bin at a subsequent one (Fig. 3D and data table S11). Finally, in the MetaHIT cohort in which only samples from the gut microbiome were available, 73 (18.5%) of the 395 subjects harbored at least one TII-PKS bin (Fig. 3B and data table S4).

After the analysis of the HMP-1-1, HMP-1-2, and MetaHIT cohorts (subjects from the United States, Denmark, and Spain), we investigated whether TII-PKS BGCs are also widespread in non-Western cohorts. To answer this question, we analyzed two additional cohorts (Fijicomp, a collection of 434 oral and fecal samples from Fijian subjects, and a Chinese cohort of 225 fecal samples) (3, 26). From $28 \times 10^9$ total reads, *MetaBGC* detected $10 \times 10^3$ reads ($3.6 \times 10^{-7}$ hit rate), which were clustered into 19 bins of >10 nonredundant biosynthetic reads (figs. S19 and S20 and data tables S3 to S5). Six bins mapped to TII-PKS BGCs and NPCs previously identified in the Western cohorts; two bins contained only false-positive reads and 11 appeared to be new ones. Of these, assemblies of 11 samples enabled the discovery of one complete and four partial TII-PKS BGCs and five NPCs (Fig. 4, fig. S15, and data table S6). Compared with the Western cohorts, 15 (8%) and 56 (25%) of the Fijian and Chinese subjects, respectively, harbored at least one TII-PKS BGC in their gut microbiome, whereas 38 (13%) of the Fijian subjects harbored a TII-PKS BGC in either their gut or oral microbiome and four (1.4%) of them were positive in both body sites (figs. S19 and S20

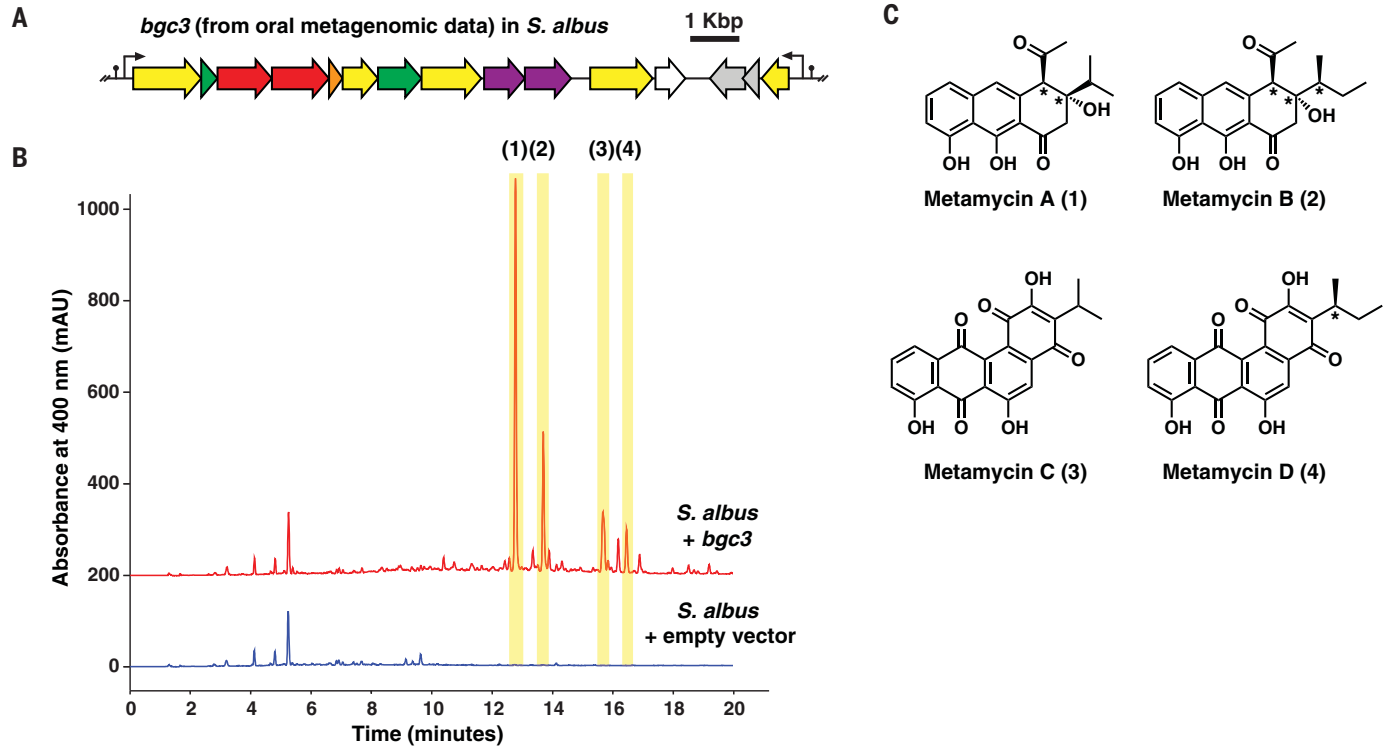



**Fig. 5. Experimental characterization of *bgc3*.** (**A**) The DNA sequence of *bgc3*, discovered from human oral metagenomic data, was codon optimized for *Streptomyces* sp. and synthesized, assembled, and cloned into the heterologous host *S. albus*. Terminator (black dots) and constitutive promoter (black arrows) sequences were engineered to control the expression of *bgc3* in *S. albus* (see the supplementary materials). (**B**) HPLC-MS analysis of a chemical extract from the culture of *S. albus*::*bgc3* (red) compared with one from *S. albus* with an empty vector control (blue). An HPLC chromatogram at the absorbance of 400 nm is shown for both samples, indicating four *bgc3*-specific peaks produced by *S. albus*::*bgc3* and not the control (**1** to **4**, highlighted in yellow). (**C**) Molecular structure of metamycins A to D (**1** to **4**), the products of *bgc3*. Asterisks indicate the relative configuration at the chiral carbons.

and data table S4). Taken together, these results indicate that TII-PKS BGCs are widespread in the human microbiome of different populations at three major body sites (mouth, gut, and skin) and are harbored by consistent colonizers of the human body (stable over several months).

Because *MetaBGC* infers the abundance of a given TII-PKS BGC from quantifying metagenomic reads derived solely from its cyclase genes, we sought to further verify the validity of this inference. For all complete TII-PKS BGCs discovered in a given cohort, we mapped metagenomic reads from all samples in the same cohort to the entire length of the BGC using an independent read recruitment method and investigated whether TII-PKS genes other than the four types of cyclases could be detected (see the supplementary materials). In >91% of the instances where a sample was deemed positive by *MetaBGC* for a given TII-PKS BGC, reads that mapped to both cyclase and noncyclase genes within the same BGC were detected in the sample (data tables S12 to S16). In addition, a strong positive correlation was observed between the bin abundances calculated by *MetaBGC* and the independently calculated abundances for the entire BGCs expressed in RPKM (reads per kilobase pair per million sequenced reads) values: the Pearson correlation coefficient was 0.85 and the *p*-value was $2.2 \times 10^{-16}$ (data tables S12 to S16). These results further validate *MetaBGC* as a sensitive, read-based algorithm for the discovery and quantification of TII-PKS BGCs in metagenomic samples of the human microbiome.

### TII-PKS BGCs are encoded by diverse members of the human microbiome and expressed under host colonization conditions

Altogether, we discovered 13 complete TII-PKS BGCs directly from metagenomic datasets of the human microbiome, each of which was found predominantly in one body site (mouth, skin, or gut) (Figs. 3B and 4). To gain more insight into the members of the human microbiome with genomes that encode these BGCs, we searched the 13 BGCs against databases of reference genomes in NCBI and the Integrated Microbial Genomes and Microbiomes (IMG) (*39*) and successfully mapped eight of them to previously sequenced bacterial genomes (the remaining five had no matches; see the supplementary materials). We found that TII-PKS BGCs are encoded by a diverse suite of Firmicutes (*Streptococcus* sp., *Staphylococcus* sp., *Lactobacillus* sp., and *Blautia* sp.) and Actinobacteria (*Propionibacterium* sp., and *Rothia* sp.) (Fig. 4 and data table S6). The mapped BGCs have strain-specific representation even among extensively sampled genera: *bgc1* and *bgc9* exist in only one strain each out of ~3000 sequenced *Streptococcus* isolates, *bgc5* exists in only one strain out of >5500 sequenced *Staphylococcus* isolates, and *bgc13* exists in only one strain out of ~1000 sequenced *Lactobacillus* isolates. These results indicate that human microbiome–derived TII-PKS BGCs are encoded by both sequenced and not-yet-sequenced members of the human microbiome in a strain-specific manner, further emphasizing the importance of discovering these pathways directly from metagenomic sequencing data.

To determine whether the discovered TII-PKS BGCs are indeed expressed in the human body, we mapped publicly available metatranscriptomic data from different human oral and fecal samples to the 13 complete TII-PKS BGCs discovered here (see the supplementary materials). Overall, we observed in vivo transcription of at least five oral BGCs (*bgc1*, *bgc2*, *bgc3*, *bgc8*, and *bgc9*) and one gut BGC (*bgc6*) in at least two different samples each (fig. S21 and data table S17). These results suggest that TII-PKS BGCs are not only encoded in the human microbiome, but

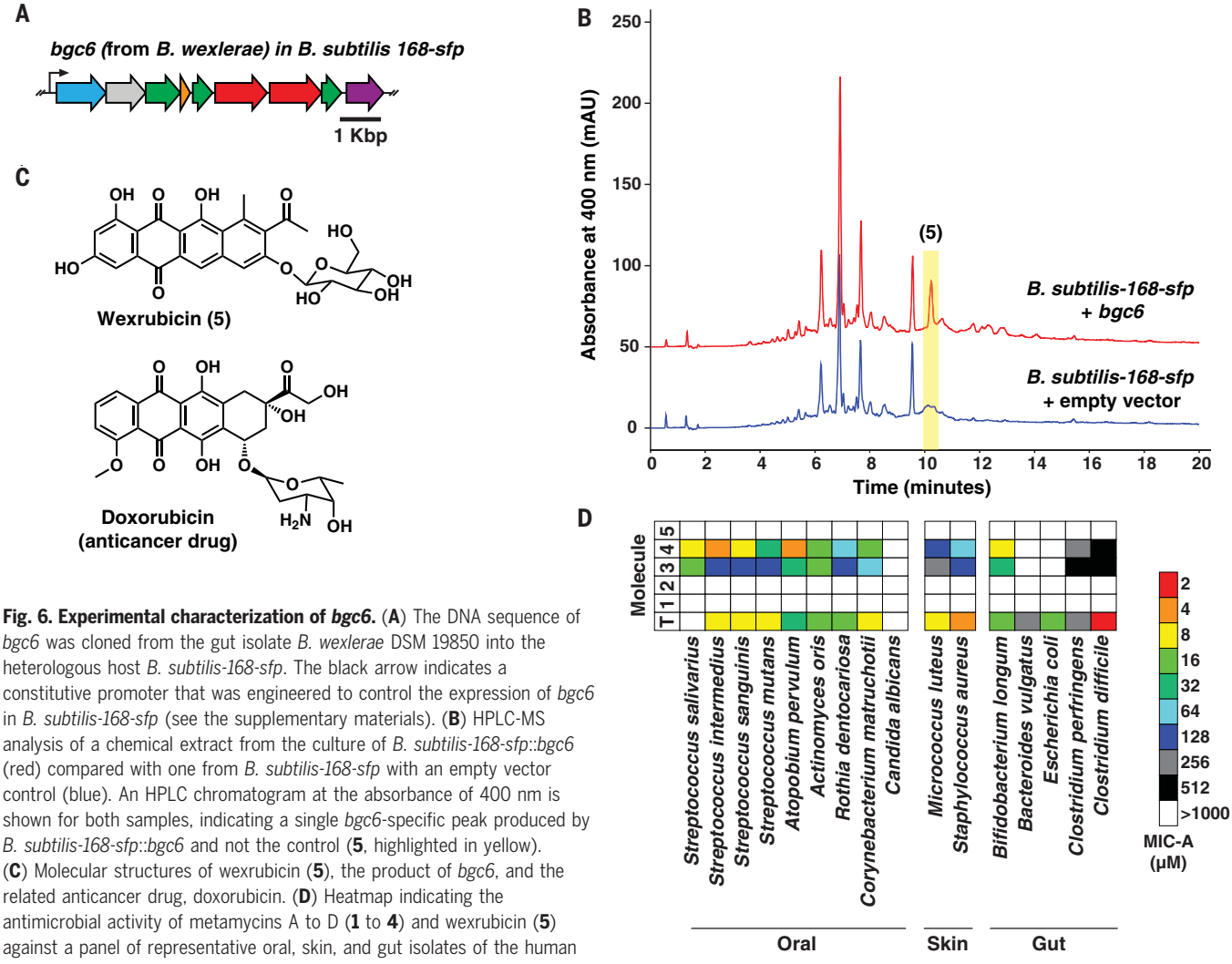



**Fig. 6. Experimental characterization of *bgc6*.** (**A**) The DNA sequence of *bgc6* was cloned from the gut isolate *B. wexlerae* DSM 19850 into the heterologous host *B. subtilis-168-sfp*. The black arrow indicates a constitutive promoter that was engineered to control the expression of *bgc6* in *B. subtilis-168-sfp* (see the supplementary materials). (**B**) HPLC-MS analysis of a chemical extract from the culture of *B. subtilis-168-sfp*::*bgc6* (red) compared with one from *B. subtilis-168-sfp* with an empty vector control (blue). An HPLC chromatogram at the absorbance of 400 nm is shown for both samples, indicating a single *bgc6*-specific peak produced by *B. subtilis-168-sfp*::*bgc6* and not the control (**5**, highlighted in yellow). (**C**) Molecular structures of wexrubicin (**5**), the product of *bgc6*, and the related anticancer drug, doxorubicin. (**D**) Heatmap indicating the antimicrobial activity of metamycins A to D (**1** to **4**) and wexrubicin (**5**) against a panel of representative oral, skin, and gut isolates of the human microbiome. The activity is measured in micromolar as the minimum inhibitory concentration on agar (MIC-A). The antibiotic tetracycline (T) was tested in the same manner, and its activity was compared with that of the newly discovered polyketides. Note that only metamycin C (**3**) and metamycin D (**4**) exert antimicrobial activities against several isolates, with metamycin D activity being similar to that of tetracycline in several cases.

are also expressed under host colonization conditions.

### Experimental characterization of TII-PKS BGCs from the oral and gut microbiome

To determine whether human microbiome–derived TII-PKS BGCs produce similar molecules (in structure and biological activity) to their previously characterized counterparts from environmental bacteria, we selected two of the in vivo–expressed BGCs for experimental characterization: an oral one (*bgc3*) and a gut one (*bgc6*). Because a native strain harboring *bgc3* had not yet been isolated at the beginning of this work, we undertook a synthetic biology strategy for its characterization. We synthesized the coding sequence of *bgc3* optimized for the heterologous expression in *Streptomyces* sp., a widely used host for the expression of small-molecule BGCs (including TII-PKS BGCs) (*40*). *bgc3* was synthesized in two overlapping fragments, with strong promoters driving the expression of its two potential operons (see the supplementary materials and Fig. 5A). The fragments were then assembled into an *Escherichia coli*–*Streptomyces*–yeast shuttle vector by transformation-associated recombination in *Saccharomyces cerevisiae* and finally integrated into a phage attachment site on the chromosome of *Streptomyces albus* by bacterial conjugation (see the supplementary materials) (*41*). We then compared the organic extracts of cultures of recombinant *S. albus*::*bgc3* with those of empty vector controls using high-performance liquid chromatography coupled with mass spectrometry (HPLC-MS), revealing the presence of several *bgc3*-specific peaks (Fig. 5B).

To further characterize the *bgc3* products, we scaled up the fermentation of *S. albus*::*bgc3* (27 L) and then isolated, purified, and solved the structure of four new type II polyketides from its organic extract (compounds **1** to **4**, named metamycins A to D, respectively; Fig. 5), using a combination of HPLC–high-resolution tandem MS (HPLC-HR-MS/MS), 1D and 2D nuclear magnetic resonance (NMR), and x-ray crystallography (see the supplementary text, figs. S22 to S37, and tables S1 and S2). Metamycins A (**1**) and B (**2**) share the same aromatic, tricyclic backbone (C16), but differ in their starter unit at C16 (a *sec*-butyl moiety in **2** instead of an isopropyl one in **1**). Metamycins C (**3**) and D (**4**) share a longer, tetracyclic aromatic backbone (C18) and similarly differ in their starter units.

The second BGC that we characterized was *bgc6* (Fig. 6A). This BGC is of special interest to us because of its variable prevalence in human fecal samples from different cohorts,

from 7% in subjects from Fiji, to 17% in subjects from Denmark or Spain, to 23% in subjects from China, and finally to 28% in subjects from the United States (data table S4). In addition, *bgc6* is encoded in the genome of the gut isolate *Blautia wexlerae* DSM 19850 from the class Clostridia. This is unusual because TII-PKS BGCs are rarely encoded in anaerobic bacteria and only two anaerobes are known to produce aromatic polyketides (*42*–*44*). To characterize the product of *bgc6*, we amplified its coding sequence from the genomic DNA of *B. wexlerae* DSM 19850, cloned it in an *E. coli*–*Bacillus subtilis* shuttle vector under a strong constitutive promoter, and integrated it into the genome of *B. subtilis* 168-sfp by natural transformation and double-crossover homologous recombination to yield *B. subtilis*-168-*sfp*::*bgc6* (see the supplementary materials). *B. subtilis* was chosen in this case because it belongs to the same phylum as *B. wexlerae* (Firmicutes), harbors an average guanine-cytosine (GC) content in its genomic DNA that is similar to that of *B. wexlerae* (GC content is 41.5% in *B. wexlerae* and 43.5% in *B. subtilis*-168), and has been previously used as a heterologous host for the expression of type I polyketides (e.g., 6-deoxyerythronolide B), so it should not be limited in primary polyketide substrates such as acetate and malonate (*45*).

We then compared the organic extracts of cultures of *B. subtilis-168-sfp*::*bgc6* and an empty vector control using HPLC-MS, revealing a single *bgc6*-specific peak (Fig. 6B). Next, we scaled up the *B. subtilis*-168-*sfp*::*bgc6* culture (31L) and then isolated, purified, and elucidated the structure of a single new molecule (compound **5**, named wexrubicin) using a combination of HPLC-HR-MS/MS and NMR (see supplementary text, figs. S38 to S44, and table S3). Wexrubicin has a tetracyclic, anthracycline ring system (C21) connected with a β-glucose moiety at C4, which is consistent with an encoded glycosyl transferase in *bgc6* (Fig. 6C).

### Biological activity of type II polyketides encoded in the human microbiome

The discovery of **1** to **5** not only shows that the human microbiome encodes previously undescribed type II polyketides, but also that some of these molecules resemble to a great extent clinically used drugs. Wexrubicin (**5**), for example, falls into the group of anthracycline polyketides, which includes the clinically used anticancer drugs doxorubicin and danuribicin, as well as the antitumor antibiotic molecules tetracenomycin and elloramycin (*46*, *47*). Conversely, metamycins A and B (**1**, **2**), and C and D (**3**, **4**) resemble the previously isolated antibiotics setomimycin from *S. pseudovenezuelae* and oviedomycin from *S. antibioticus*, respectively (*48*, *49*). To determine whether microbiome-derived polyketides exert the same biological activity as their closely related derivatives, we tested them in two types of assays: cytotoxicity against mammalian cell lines and antimicrobial activity against selected bacteria and fungi (see the supplementary materials). Compared with doxorubicin as a positive control, none of the tested molecules showed notable cytotoxicity against HeLa cell lines (fig. S45).

In antimicrobial assays, however, metamycins C and D (**3**, **4**) showed strong inhibitory activity against several Gram-positive bacteria at a magnitude similar to that of the clinically used antibiotic tetracycline in some cases. These molecules were most potent against oral isolates of *Streptococcus*, *Atopobium*, *Actinomyces*, *Rothia*, and *Corynebacterium* sp. (Fig. 6D). Being the products of an oral BGC (*bgc3*), these results suggest that metamycins C and D (**3**, **4**) may play a role in niche competition in the oral cavity or in host protection against pathogens. This is consistent with the fact that *bgc3* is expressed in human supragingival plaque samples during early biofilm formation, as determined by metatranscriptomic analysis (fig. S21 and data table S17). Because no cytotoxicity or antibacterial activity was detected for metamycins A and B (**1**, **2**) or wexrubicin (**5**), further studies are needed to provide insights into their biological role.

## Discussion

Although substantial efforts have been spent on documenting and characterizing diverse small molecules encoded by human-associated bacteria, no aromatic polyketides have been previously described from the human microbiome. Here, we used a new computational strategy for the discovery of small-molecule BGCs directly from human microbiome–derived metagenomic sequencing data, revealing the unexpectedly wide distribution of BGCs encoding for this relatively rare class of molecules in the human microbiome (~50% of subjects from the United States carry at least one BGC of this class). We then combined this in silico approach with a synthetic biology strategy to heterologously express the identified BGCs and directly discover their chemical products. The structural diversity observed in the products of two of the 13 BGCs discovered here, and the fact that they resemble in structure or biological activity clinically used drugs, clearly motivate further functional investigations into this class. These investigations will not only serve as an important avenue for elucidating microbiome-host interactions at the molecular level, but they will also serve as an unprecedented resource for drug discovery from within the human body.

Our computational strategy relies on repurposing and tailoring established probabilistic models for the use with short-read metagenomic data, which achieves high-sensitivity and high-specificity detection of target protein families. We focus on not one but four distinct protein families in this study as a proof of principle (TII-PKS cyclases and aromatases), which demonstrates the generality of our method. The same approach could be easily adapted to any protein family of interest, including ones that are specific to other types of BGCs. To illustrate this point, we applied our same strategy to two new protein families that are specific to BGCs that encode two additional types of small molecules (siderophores and lanthipeptides). In both cases, we were successful in separating low- and high-performance spHMMs and in using the four modules of *MetaBGC* using synthetic dataset 2 (see the supplementary materials, data table S18, and figs. S46 to S49).

The performance and limitations of *MetaBGC* depend on several factors. First, to be used as a discovery and quantification method for a given BGC, it is important that spHMMs are built for a protein family that is both specific for and universally found in the BGC of interest. Second, a relatively high-performance, full-size pHMM needs to be generated before segmentation, which relies in turn on the availability of sufficient numbers and a proper alignment of sequenced homologs for the protein family of interest. Third, a carefully designed synthetic dataset needs to be generated to reflect as much as possible the expected complexity in the real metagenomic datasets of interest and to evaluate the spHMMs and tailor their score cutoffs. Finally, several flexible parameters need to be optimized on the basis of the specific study goals, including the size of the spHMM segments (depending on the read length of the search data), the F1 and spHMM score cutoffs (depending on how conservative the discovery goals are), the minimum number of nonredundant biosynthetic reads to define a new bin, the minimum bin abundance to define the "presence" or "absence" of a given bin in a metagenomic sample (depending on how sensitive the study is designed to be), and, finally, the minimum Pearson correlation distance used in the clustering step, which dictates the stringency of the binning. We have provided example settings for these parameters in our current study, which can be used as starting points for new applications and further optimized as needed.

In conclusion, we present a general strategy that is useful for quickly profiling metagenomic datasets from large clinical cohorts and prioritizing candidate BGCs for experimental characterization. More broadly, we see this as a systematic strategy for unveiling the chemical repertoire encoded by the human microbiome, a much-needed step for understanding its role in human health and disease.

## Materials and methods summary

An expanded materials and methods section can be found in the supplementary materials.

### Development of MetaBGC

*MetaBGC* is composed of four modules. In *MetaBGC-Build*, spHMMs are developed for every protein family of interest and evaluated using a synthetic metagenomic dataset that harbors reads from two groups of bacterial genomes: genomes that encode diverse members of the protein family of interest (but not included in the spHMMs) and genomes that lack any. In *MetaBGC-Identify*, high-performance spHMMs are used to search billions of translated metagenomic reads from thousands of metagenomic samples to identify "biosynthetic reads." In *MetaBGC-Quantify*, identified biosynthetic reads are de-replicated and quantified in all samples of a given cohort and an abundance profile matrix is built for all reads in all samples. Finally, in *MetaBGC-Cluster*, biosynthetic reads are clustered into "biosynthetic read bins" on the basis of their abundance profiles across samples. These bins are then used to calculate "biosynthetic read bin abundance" in each sample and as seeds for targeted or untargeted BGC assembly efforts from the sample with the highest abundance of a given bin. For exact details about the development of *MetaBGC*, see the supplementary materials.

### *Using* MetaBGC *to search metagenomic sequencing data of the human microbiome for TII-PKS BGCs*

High-performance spHMMs for four types of cyclases and aromatases commonly found in TII-PKS BGCs were developed and used to search quality-filtered and translated metagenomic sequencing reads from five large cohorts (HMP-1-1, HMP-1-2, MetaHIT, Chinese, and Fijicomp). After quantification, clustering, and assembly of biosynthetic read bins, complete TII-PKS BGCs were annotated and two were selected for experimental characterization. For details about the complete set of analyses performed on the five cohorts and the discovered TII-PKS BGCs, see the supplementary materials.

### *Experimental characterization of two human microbiome–derived TII-PKS BGCs*

Out of the 13 identified TII-PKS BGCs, we selected two, *bgc3* and *bgc6*, for experimental characterization on the basis of their profiles in metagenomic and metatranscriptomic datasets. The metagenomic DNA sequence for *bgc3* was codon optimized, synthesized, and engineered for heterologous expression in *S. albus*, and that of *bgc6* was amplified from genomic DNA isolated from *B. wexlerae* DSM 19850 and engineered for heterologous expression in *B. subtilis*. Heterologous expression lines harboring the BGCs or empty vector controls were cultured, chemically extracted, and analyzed using HPLC-MS. Large-scale cultivation experiments of the expression lines were used to provide enough materials for the isolation and purification of the five type II polyketide molecules reported here. Structural elucidation of the purified molecules was performed using a combination of NMR, x-ray crystallography, and MS. Biological activity of the purified molecules was assessed using cytotoxicity assays against HeLa cells and minimum inhibitory concentration (MIC)-on-agar assays against several pathogenic and commensal bacteria. For additional details, see the supplementary materials.

## REFERENCES AND NOTES


1. Human Microbiome Project Consortium, Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012). doi: 10.1038/nature11234; pmid: 22699609
2. J. Lloyd-Price *et al*., Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* **550**, 61–66 (2017). doi: 10.1038/nature23889; pmid: 28953883
3. J. Qin *et al*., A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012). doi: 10.1038/nature11450; pmid: 23023125
4. A. M. Thomas *et al*., Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat. Med.* **25**, 667–678 (2019). doi: 10.1038/s41591-019-0405-7; pmid: 30936548
5. J. Wirbel *et al*., Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* **25**, 679–689 (2019). doi: 10.1038/s41591-019-0406-6; pmid: 30936547
6. M. S. Donia, M. A. Fischbach, Small molecules from the human microbiota. *Science* **349**, 1254766 (2015). doi: 10.1126/science.1254766; pmid: 26206939
7. A. Milshteyn, D. A. Colosimo, S. F. Brady, Accessing bioactive natural products from the human microbiome. *Cell Host Microbe* **23**, 725–736 (2018). doi: 10.1016/j.chom.2018.05.013; pmid: 29902438
8. L. J. Cohen *et al*., Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* **549**, 48–53 (2017). doi: 10.1038/nature23874; pmid: 28854168
9. C. J. Guo *et al*., Discovery of reactive microbiota-derived metabolites that inhibit host proteases. *Cell* **168**, 517–526.e18 (2017). doi: 10.1016/j.cell.2016.12.021; pmid: 28111075
10. G. Schneditz *et al*., Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13181–13186 (2014). doi: 10.1073/pnas.1403274111; pmid: 25157164
11. M. Wlodarska *et al*., Indoleacrylic acid produced by commensal peptostreptococcus species suppresses inflammation. *Cell Host Microbe* **22**, 25–37.e6 (2017). doi: 10.1016/j.chom.2017.06.007; pmid: 28704649
12. Y. Bhattarai *et al*., Gut microbiota-produced tryptamine activates an epithelial G-protein-coupled receptor to increase colonic secretion. *Cell Host Microbe* **23**, 775–785.e5 (2018). doi: 10.1016/j.chom.2018.05.004; pmid: 29902441
13. Z. Wang *et al*., Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011). doi: 10.1038/nature09922; pmid: 21475195
14. M. R. Wilson *et al*., The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, eaar7785 (2019). doi: 10.1126/science.aar7785; pmid: 30765538
15. M. S. Donia *et al*., A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014). doi: 10.1016/j.cell.2014.08.032; pmid: 25215495
16. T. Nakatsuji *et al*., Antimicrobials from human skin commensal bacteria protect against *Staphylococcus aureus* and are deficient in atopic dermatitis. *Sci. Transl. Med.* **9**, eaah4680 (2017). doi: 10.1126/scitranslmed.aah4680; pmid: 28228596
17. A. Zipperer *et al*., Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511–516 (2016). doi: 10.1038/nature18634; pmid: 27466123
18. C. G. Buffie *et al*., Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517**, 205–208 (2015). doi: 10.1038/nature13828; pmid: 25337874
19. Human Microbiome Project Consortium, A framework for human microbiome research. *Nature* **486**, 215–221 (2012). doi: 10.1038/nature11209; pmid: 22699610
20. H. P. Browne *et al*., Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543–546 (2016). doi: 10.1038/nature17645; pmid: 27144353
21. J. C. Lagier *et al*., Culture of previously uncultured members of the human gut microbiota by culturomics. *Nat. Microbiol.* **1**, 16203 (2016). doi: 10.1038/nmicrobiol.2016.203; pmid: 27819657
22. S. C. Forster *et al*., A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat. Biotechnol.* **37**, 186–192 (2019). doi: 10.1038/s41587-018-0009-7; pmid: 30718869
23. K. R. Chng *et al*., Whole metagenome profiling reveals skin microbiome-dependent susceptibility to atopic dermatitis flare. *Nat. Microbiol.* **1**, 16106 (2016). doi: 10.1038/nmicrobiol.2016.106; pmid: 27562258
24. W. Zhou *et al*., Longitudinal multi-omics of host-microbe dynamics in prediabetes. *Nature* **569**, 663–671 (2019). doi: 10.1038/s41586-019-1236-x; pmid: 31142858
25. J. M. Fettweis *et al*., The vaginal microbiome and preterm birth. *Nat. Med.* **25**, 1012–1021 (2019). doi: 10.1038/s41591-019-0450-2; pmid: 31142849
26. I. L. Brito *et al*., Mobile genes in the human microbiome are structured from global to individual scales. *Nature* **535**, 435–439 (2016). doi: 10.1038/nature18927; pmid: 27409808
27. S. A. Smits *et al*., Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* **357**, 802–806 (2017). doi: 10.1126/science.aan4834; pmid: 28839072
28. A. Almeida *et al*., A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499–504 (2019). doi: 10.1038/s41586-019-0965-1; pmid: 30745586
29. S. Nayfach, Z. J. Shi, R. Seshadri, K. S. Pollard, N. C. Kyrpides, New insights from uncultivated genomes of the global human gut microbiome. *Nature* **568**, 505–510 (2019). doi: 10.1038/s41586-019-1058-x; pmid: 30867587
30. E. Pasolli *et al*., Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**, 649–662.e20 (2019). doi: 10.1016/j.cell.2019.01.001; pmid: 30661755
31. P. Baldi, Y. Chauvin, T. Hunkapiller, M. A. McClure, Hidden Markov models of biological primary sequence information. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1059–1063 (1994). doi: 10.1073/pnas.91.3.1059; pmid: 8302831
32. S. R. Eddy, Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998). doi: 10.1093/bioinformatics/14.9.755; pmid: 9918945
33. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, paper presented at the *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, 1996.
34. J. Kim, G. S. Yi, PKMiner: A database for exploring type II polyketide synthases. *BMC Microbiol.* **12**, 169 (2012). doi: 10.1186/1471-2180-12-169; pmid: 22871112
35. C. Hertweck, A. Luzhetskyy, Y. Rebets, A. Bechthold, Type II polyketide synthases: Gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**, 162–190 (2007). doi: 10.1039/B507395M; pmid: 17268612
36. T. Weber *et al*., antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43** (W1), W237–W243 (2015). doi: 10.1093/nar/gkv437; pmid: 25948579
37. M. E. Hillenmeyer, G. A. Vandova, E. E. Berlew, L. K. Charkoudian, Evolution of chemical diversity by coordinated gene swaps in type II polyketide gene clusters. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 13952–13957 (2015). doi: 10.1073/pnas.1511688112; pmid: 26499248
38. H. B. Nielsen *et al*., Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014). doi: 10.1038/nbt.2939; pmid: 24997787
39. A. Checcucci, A. Mengoni, The integrated microbial genome resource of analysis. *Methods Mol. Biol.* **1231**, 289–295 (2015). doi: 10.1007/978-1-4939-1720-4_18; pmid: 25343872
40. R. H. Baltz, Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes. *J. Ind. Microbiol. Biotechnol.* **43**, 343–370 (2016). doi: 10.1007/s10295-015-1682-x; pmid: 26364200
41. X. Tang *et al*., Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining.

*ACS Chem. Biol.* **10**, 2841–2849 (2015). doi: 10.1021/acschembio.5b00658; pmid: 26458099
42. A. C. Letzel, S. J. Pidot, C. Hertweck, A genomic approach to the cryptic secondary metabolome of the anaerobic world. *Nat. Prod. Rep.* **30**, 392–428 (2013). doi: 10.1039/C2NP20103H; pmid: 23263685
43. G. Shabuer *et al*., Plant pathogenic anaerobic bacteria use aromatic polyketides to access aerobic territory. *Science* **350**, 670–674 (2015). doi: 10.1126/science.aac9990; pmid: 26542569
44. S. Pidot, K. Ishida, M. Cyrulies, C. Hertweck, Discovery of clostrubin, an exceptional polyphenolic polyketide antibiotic from a strictly anaerobic bacterium. *Angew. Chem. Int. Ed.* **53**, 7856–7859 (2014). doi: 10.1002/anie.201402632; pmid: 24827417
45. J. Kumpfmüller *et al*., Production of the polyketide 6-deoxyerythronolide B in the heterologous host *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **100**, 1209–1220 (2016). doi: 10.1007/s00253-015-6990-6; pmid: 26432460
46. H. Drautz, P. Reuschenbach, H. Zähner, J. Rohr, A. Zeeck, Metabolic products of microorganisms. 225. Elloramycin, a new anthracycline-like antibiotic from *Streptomyces olivaceus*. Isolation, characterization, structure and biological properties. *J. Antibiot.* **38**, 1291–1301 (1985). doi: 10.7164/antibiotics.38.1291; pmid: 3840789
47. J. Rohr, A. Zeeck, Structure-activity relationships of elloramycin and tetracenomycin C. *J. Antibiot.* **43**, 1169–1178 (1990). doi: 10.7164/antibiotics.43.1169; pmid: 2211379
48. S. Omura *et al*., A new antibiotic, setomimycin, produced by a strain of *Streptomyces*. *J. Antibiot.* **31**, 1091–1098 (1978). pmid: 721705
49. C. Méndez *et al*., Oviedomycin, an unusual angucyclinone encoded by genes of the oleandomycin-producer *Streptomyces antibioticus* ATCC11891. *J. Nat. Prod.* **65**, 779–782 (2002). doi: 10.1021/np010555n; pmid: 12027768
50. Y. Sugimoto *et al*., donia-lab/MetaBGC: v1.0.0, Version 1.0.0, Zenodo (2019); https://doi.org/10.5281/zenodo.3383399.

## ACKNOWLEDGMENTS

We thank M. Cahn for assistance with metagenomic data analysis; B. Javdan, S. Rath, and A. Korennykh for assistance with mammalian cell culture; the crystallography core facility at the Department of Molecular Biology, Princeton University; S. Chatterjee for general assistance; and the rest of the Donia laboratory for useful discussions. **Funding:** Funding for this project was provided by an NIH Director's New Innovator Award (ID: 1DP2AI124441) to M.S.D., the Pew Biomedical Scholars Program to M.S.D., and a Focused Research Team on Precision Antibiotics Award by the School of Engineering and Applied Science at Princeton University through the generosity of Helen Shipley Hunt *71. **Author contributions:** M.S.D., Y.S., and F.R.C. designed the study. Y.S., F.R.C., S.W., P.C., A.O., A.B., P.D.J., and M.S.D. performed experiments and analyzed the data. M.S.D., F.C., and Y.S. wrote the manuscript with input from the rest of the authors. **Competing interests:** M.S.D. is a member of the scientific advisory board of DeepBiome Therapeutics and a consultant for Flagship Pioneering. **Data and materials availability:** All data are available in the main text or the supplementary materials. The code for *MetaBGC* is archived at Zenodo (*50*). Crystallography data were deposited at the Cambridge Crystallographic Data Centre under accession numbers CCDC 1878185, CCDC 1878186, and CCDC 1878187.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/eaax9176/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S49
Tables S1 to S3
References
Data Tables S1 to S18
View/request a protocol for this paper from *Bio-protocol*.

4 May 2019; accepted 23 September 2019
Published online 3 October 2019
10.1126/science.aax9176

RESEARCH ARTICLE

ANCIENT ATMOSPHERE

# Stepwise Earth oxygenation is an inherent property of global biogeochemical cycling

Lewis J. Alcott*, Benjamin J. W. Mills, Simon W. Poulton

Oxygenation of Earth's atmosphere and oceans occurred across three major steps during the Paleoproterozoic, Neoproterozoic, and Paleozoic eras, with each increase having profound consequences for the biosphere. Biological or tectonic revolutions have been proposed to explain each of these stepwise increases in oxygen, but the principal driver of each event remains unclear. Here we show, using a theoretical model, that the observed oxygenation steps are a simple consequence of internal feedbacks in the long-term biogeochemical cycles of carbon, oxygen, and phosphorus, and that there is no requirement for a specific stepwise external forcing to explain the course of Earth surface oxygenation. We conclude that Earth's oxygenation events are entirely consistent with gradual oxygenation of the planetary surface after the evolution of oxygenic photosynthesis.

Oxygenation of Earth's surface environment is thought to have occurred across three broad steps (Fig. 1). The Great Oxidation Event (GOE) occurred around 2.4 to 2.2 billion years ago (Ga) and saw atmospheric $O_2$ rise from trace levels to more than $10^{-5}$ of the present atmospheric level (PAL) (*1*). The following ~1 billion years (Gyr) of the Proterozoic eon likely sustained atmospheric $O_2$ levels of ~$10^{-3}$ to $10^{-1}$ PAL (*2*). Partial oxygenation of the surface ocean persisted throughout the Proterozoic (*3*), but deeper waters remained dominantly anoxic (*4*). The Neoproterozoic Oxygenation Event (NOE) occurred between ~800 and 540 million years ago (Ma) and is generally believed to have resulted in atmospheric $O_2$ levels of 0.1 to 0.5 PAL, as well as the first oxygenation of the deep ocean (*5*). However, there was considerable variability in the temporal and spatial extent of deep-ocean oxygenation at this time, including the possibility of pulsed oceanic oxic events (*6*). Evidence for periodic deep-water anoxia remains frequent up until the mid-Paleozoic, when a final major rise in atmospheric $O_2$ concentration occurred around 450 to 400 Ma (*7*). This Paleozoic Oxygenation Event (POE) appears to have elevated atmospheric $O_2$ to present-day levels and established a dominantly oxygenated deep ocean, which persisted throughout the Mesozoic and Cenozoic eras.

These major oxygenation steps are intertwined with the evolution of progressively more complex life-forms. The first eukaryotes evolved either after the GOE or during the run-up to the event when $O_2$ began to rise (*8*), whereas the NOE was coincident with major eukaryote diversification and the evolution of the first animals (*9*), followed by the Cambrian explosion, during which animals began to dominate marine ecosystems. The POE was accompanied by a major increase in animal body size, more diverse and specialized predators, and the evolution of vascular land plants (*7*). However, determining causality between rises in marine and atmospheric $O_2$ levels and the evolution of the biosphere is complex, and there is considerable debate over the role of $O_2$ in driving biological evolution versus the role of life in bioengineering $O_2$ to higher levels (*10*, *11*).

Tectonic evolution has also been considered as a potential driver of the stepwise transitions in Earth surface oxygenation. Changes to plate tectonics have been linked to the GOE through, for example, a change in the fraction of subaerial volcanism (*12*) or the composition of the crust (*13*). Some, but not all, supercontinent formation times correspond to oxygenation events (*14*), as do emplacement times of some large igneous provinces (LIPs), which are proposed to have driven ocean oxygenation through delivery of the limiting nutrient, phosphate (*15*).

However, the geologically rapid yet ultimately rare nature of Earth's oxygenation events does not clearly correspond to either tectonic or evolutionary processes. For example, mantle dynamics and the supercontinent cycle are unlikely to produce large-scale changes on time scales of the order of less than ~100 million years (Myr), whereas LIP emplacements are far more common than major rises in $O_2$. Looking to biological innovations, the time scale between the origination of a domain or kingdom of life and its rise to global ecological dominance may also be hundreds of million years (e.g., Eukarya) (*16*). Furthermore, the oscillations in ocean redox (reduction–oxidation) that are apparent during the NOE are difficult to explain through a sequence of tectonic or biological "switches" acting on the system (*17*).

It is therefore possible that Earth's stepwise oxygenation was not the product of individual trigger events and may instead be explained by some inherent property of global biogeochemical feedbacks. This hypothesis has wide implications for the evolution of life on Earth and other planets, and there have therefore been a number of attempts to explain the known stepwise $O_2$ trajectory as a feature of Earth's internal dynamics: For example, it has been shown that atmospheric feedbacks might have promoted the GOE (*18*, *19*). However, no study has provided a sound theoretical basis that can explain the trajectory and timing of marine and atmospheric oxygenation over Earth's history without relying on either external trigger events or arbitrary switches in the model itself (such as assuming a transition to greater nutrient availability when $O_2$ crosses a threshold) (*20*, *21*).

### Biogeochemical feedbacks

Here we identify a set of feedbacks that exist between the global P, C, and O cycles, which are capable of driving rapid shifts in ocean and atmospheric $O_2$ levels without requiring any stepwise change in either tectonics or the evolution of the biosphere. These feedbacks reproduce the observed three-step

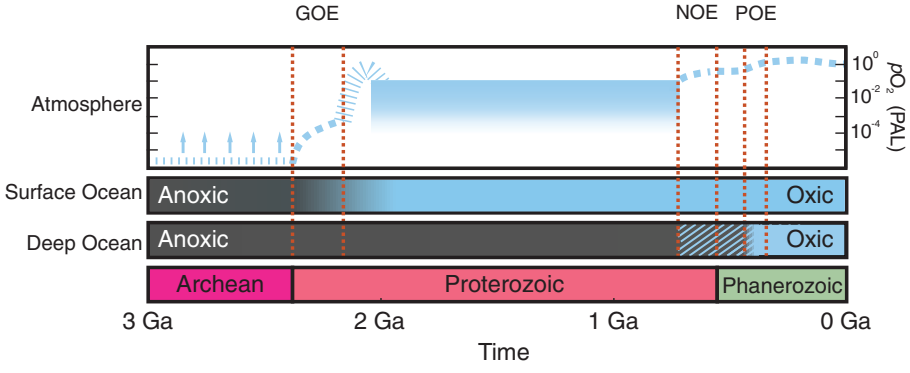



**Fig. 1. Redox history of Earth.** Atmospheric $O_2$ based on (*47*). Crosshatching indicates variable deep-ocean redox between the start of the NOE and the POE. See text for a summary and related references.

School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK.
*Corresponding author. Email: eelja@leeds.ac.uk

oxygenation pattern when driven solely by a gradual shift from reducing to oxidizing surface conditions over time.

Phosphorus (P) is generally considered the ultimate limiting nutrient for marine productivity over geological time scales (*22*), and P bioavailability exerts a key control on the long-term rate of $O_2$ production through oxygenic photosynthesis and organic carbon ($C_{org}$) burial. In the modern ocean, bioavailable P is fixed in the sediments via three primary pathways. Organic-bound P ($P_{org}$) is buried with sinking organic matter, iron-bound P ($P_{Fe}$) forms as P is adsorbed or coprecipitated with iron (oxyhydr)oxide minerals, and authigenic P ($P_{auth}$) is primarily formed in the sediment by means of "sink switching" of these phases to carbonate fluorapatite and/or vivianite (*23*–*25*).

The phase partitioning of sedimentary P is largely controlled by redox conditions in the water column and sediments. P may be preferentially released from organic matter during remineralization under anoxic conditions, leading to elevated $C_{org}$:$P_{org}$ ratios in the sediment (*26*), increased recycling of P back to the water column, and reduced formation of authigenic carbonate fluorapatite (*27*, *28*). The availability of iron (oxyhydr)oxides is also typically considered to diminish in an anoxic system, which limits the burial of $P_{Fe}$ (*23*), although this dynamic becomes more complex when considering the prevalence of low-sulfate, ferruginous (anoxic Fe-rich) oceanic conditions throughout large parts of Earth's history (*4*), whereby Fe minerals may trap a proportion of the P delivered to the sediment (*29*).

Assuming that bottom-water anoxia leads to an overall enhancement of sedimentary P regeneration, two important feedback mechanisms arise that affect global biogeochemistry, with each operating over a different time scale. First, a short-term positive-feedback mechanism (self-promoting) operates, whereby the spread of ocean anoxia results in increased P availability in the water column. This stimulates primary productivity and fuels respiration, thus further increasing P availability (*26*). These eutrophic conditions rapidly deplete water-column $O_2$ and in turn increase the rate of spread of anoxia. Second, a geologically paced negative feedback mechanism (self-inhibiting) operates on the combined C-O-P cycles, whereby the burial of $C_{org}$ in marine sediments leads to oxygenation of the atmosphere. This increase in partial pressure of oxygen ($Po_2$) drives higher rates of oxidative weathering of ancient sedimentary $C_{org}$ (*30*) and ventilates the ocean, acting to stabilize $O_2$ by both reducing the rate of $C_{org}$ burial and increasing the consumption of $O_2$ on land.

Theoretical models have previously linked the above feedbacks to the geologically rapid onset of Cretaceous ocean anoxic events and to their delayed termination through increasing atmospheric $O_2$ (*28*). It has also been shown that under an increased continental weathering input of P, self-sustaining oscillations between oxic and anoxic oceanic states might occur (*31*). We hypothesize here that the above feedbacks are in fact sufficient to explain the stepwise oxygenation of Earth's atmosphere and oceans, including apparent cyclic ocean oxygenation and deoxygenation events during the Neoproterozoic and early Phanerozoic (*6*), which are followed by the transition to a sustained oxic deep ocean.

First, the GOE can occur when the weathering of $C_{org}$ becomes the principal long-term $O_2$ sink. This can be achieved once the rate of photosynthetic $O_2$ production outpaces the consumption of $O_2$ by way of reaction with reduced gases and reduced seawater species. Second, cyclic oxygenation events in the NOE would be a likely consequence of the combined positive and negative feedbacks between ocean oxygenation and sedimentary P recycling, whereby a small shift toward a more oxygenated planetary surface results in gradual oxygenation of oceanic bottom waters. This would limit P regeneration from sediments, reducing short-term productivity and $O_2$ demand and thus further increasing dissolved $O_2$ concentrations. This positive feedback could oxygenate ocean basins, but only

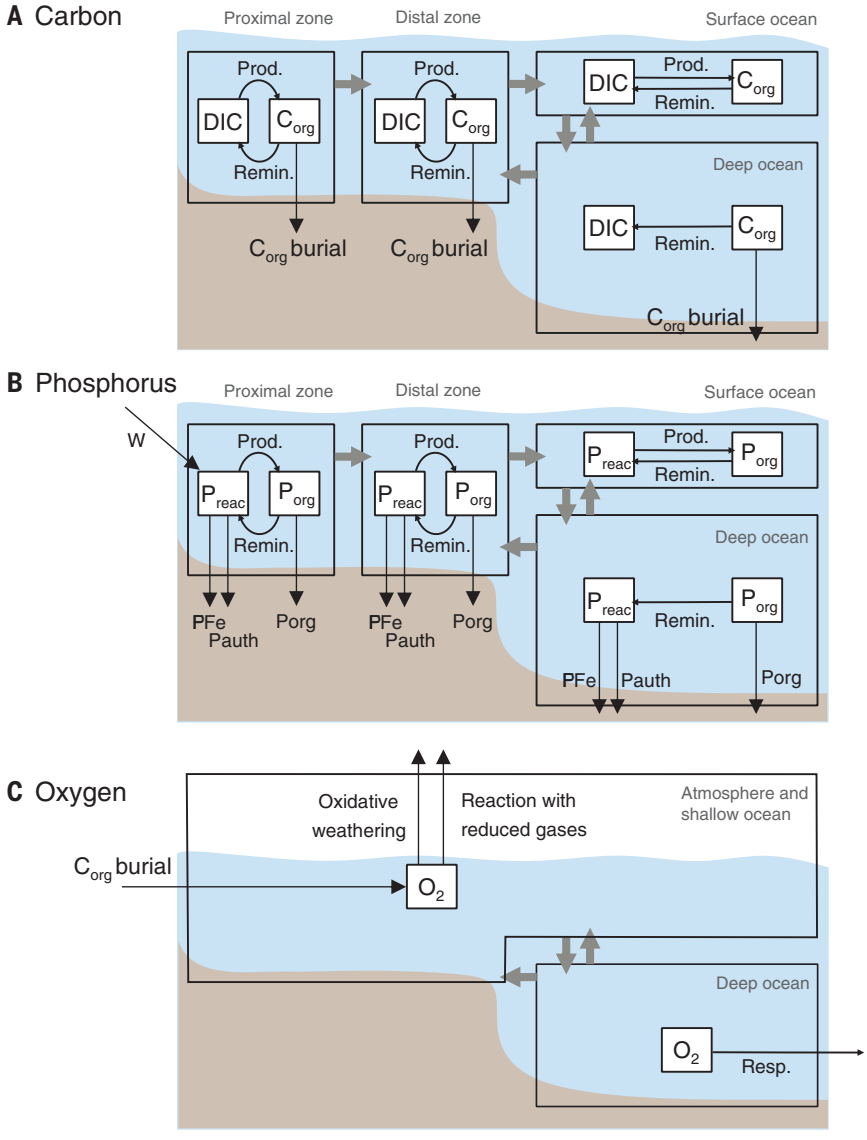



**Fig. 2. Ocean and atmosphere box model.** Boxes with background color show hydrospheric reservoirs, and gray arrows denote mixing between them. White boxes show chemical reservoirs, and black arrows denote biogeochemical fluxes. (**A**) Carbon cycle: C exists as dissolved inorganic carbon (DIC) or $C_{org}$. Prod., primary productivity; Remin., remineralization. (**B**) Phosphorus cycle: P exists as soluble reactive phosphorus ($P_{reac}$) and $P_{org}$. W, weathering. (**C**) Oxygen cycle. Single $O_2$ reservoir encompasses all ocean boxes that exchange with the atmosphere. Resp., respiration. See text for full description; see methods and other supplementary materials for equations.

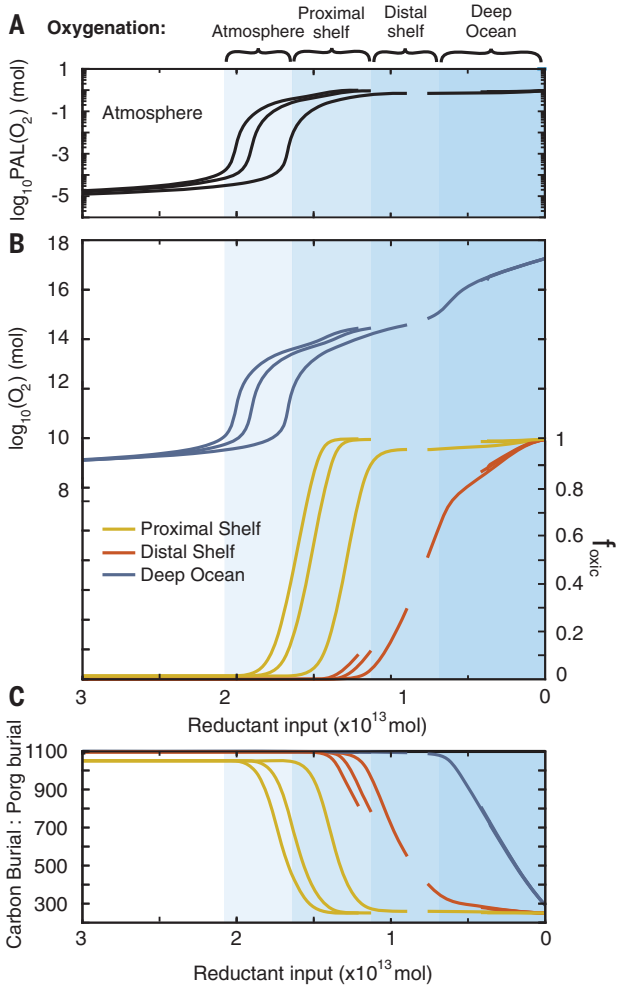



**Fig. 3. Model stable solutions with respect to overall surface redox state.** The model is run to steady state for changes to the reduced gas input rate. (**A**) Atmospheric $O_2$ reservoir. (**B**) Deep-ocean $O_2$ reservoir (mol) and shelf oxic fraction of seafloor ($f_{oxic}$). (**C**) Molar C and P burial ratio in sediments. Three lines for each zone represent different choices of redox dependence for P burial fluxes (see text). Breaks in solid lines indicate periods when no stable solution exists; in these parameter spaces, the model produces an oscillating solution.

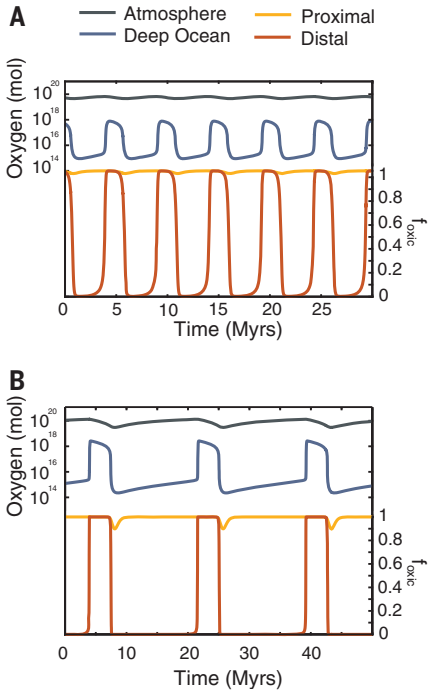



**Fig. 4. Oscillating redox solutions.** Transient model responses demonstrating limit cycles of frequency of ~5 to 20 Myr. (**A**) Conservative redox dependency for deep-ocean P burial terms (50% $P_{auth}$, 25% $P_{org}$) with $1 \times 10^{13}$ moles of $O_2$ consumption. (**B**) Stronger redox dependencies (90% $P_{auth}$, 50% $P_{org}$) with $2.5 \times 10^{13}$ moles of $O_2$ consumption (*28*).

temporarily, because reduced P availability for primary productivity then leads to less $O_2$ production over geologic time scales, thus reducing atmospheric $O_2$, which eventually leads to a return to marine anoxia. Finally, a combination of the two mechanisms outlined above can result in sustained oceanic oxygenation. In this case, a sufficiently large increase in surface redox potential, coupled with a greater contribution of oxidative weathering to overall $O_2$ regulation, allows deep-ocean oxygenation to be maintained.

## Modeling results

We test our hypothesis by building on a well-established conceptual model of marine biogeochemistry (*28*, *32*, *33*). The model tracks the global cycles of P, C, and $O_2$ in a four-box ocean system representing shelf, open-ocean, and deep-water environments (Fig. 2). We add to the model an atmospheric $O_2$ reservoir, a global geological $O_2$ cycle, oxidative weathering of $C_{org}$, and an open-ocean scavenging flux of P by upwelled Fe, basing these on other previous models (*29*, *30*).

Phosphorus-dependent primary productivity occurs in all surface ocean boxes, and redox-dependent burial of P is included in all boxes that are in contact with the sediments. As in previous versions of the model, sedimentary inventories are not calculated explicitly, and regeneration of P from sediments is addressed through the net P burial terms, which follow previous model derivations (*28*, *32*). Deep-ocean dissolved $O_2$ concentration is calculated explicitly, as is the $O_2$ content of the atmosphere. Following previous model versions, the $O_2$ content of the ocean boxes in contact with the atmosphere is represented by a "degree of anoxia" parameter, termed $f_{anoxic}$, which represents the balance between $O_2$ diffusion from the atmosphere and $O_2$ use (*32*). An optional scavenging flux of P sorption to upwelling iron particles (*29*) is implemented to test the effects of additional P drawdown in a ferruginous (iron-rich) ocean, a state that may have persisted throughout much of Earth's history (*4*). The scavenging flux directly follows previous models (*29*), operating at $O_2$ concentrations below 1 μM and removing 25% of the P that is upwelled into the open ocean. Full model equations are shown in the supplementary materials. As well as testing different limits on the scavenging flux, we also ran sensitivity tests to vary the degree of redox dependency of the $P_{org}$ and $P_{auth}$ burial terms (following previous versions of this model).

Figure 3 shows steady-state responses of the baseline model to variable fluxes of reduced gas to the surface system, representing overall changes in net surface redox over Earth's history. All other parameters remain at their present-day values, and processes vary only by means of internal feedbacks. The redox dependence of the $P_{org}$ and $P_{auth}$ burial fluxes were varied and are shown as different lines. The model runs with stronger redox dependencies (40% for $P_{auth}$, 25% for $P_{org}$) respond first to oxygenation (leftmost lines in Fig. 3), with weaker dependencies (35% for $P_{auth}$, 25% for $P_{org}$, and 10% for both $P_{auth}$ and $P_{org}$) plotted to the right of these. Under a very high reductant flux (left side of $x$ axis in Fig. 3), as proposed for the Archean ($\gg 1 \times 10^{13}$ mol $O_2$ equivalent reductant input) (*34*, *35*), $O_2$ production is overwhelmed and atmospheric $O_2$ is stable at ~$10^{-5}$ PAL, consistent with pre-GOE conditions (*36*). The $O_2$ balance is primarily maintained by reaction of $O_2$ with reduced gases (*30*). All surface and deep-ocean boxes

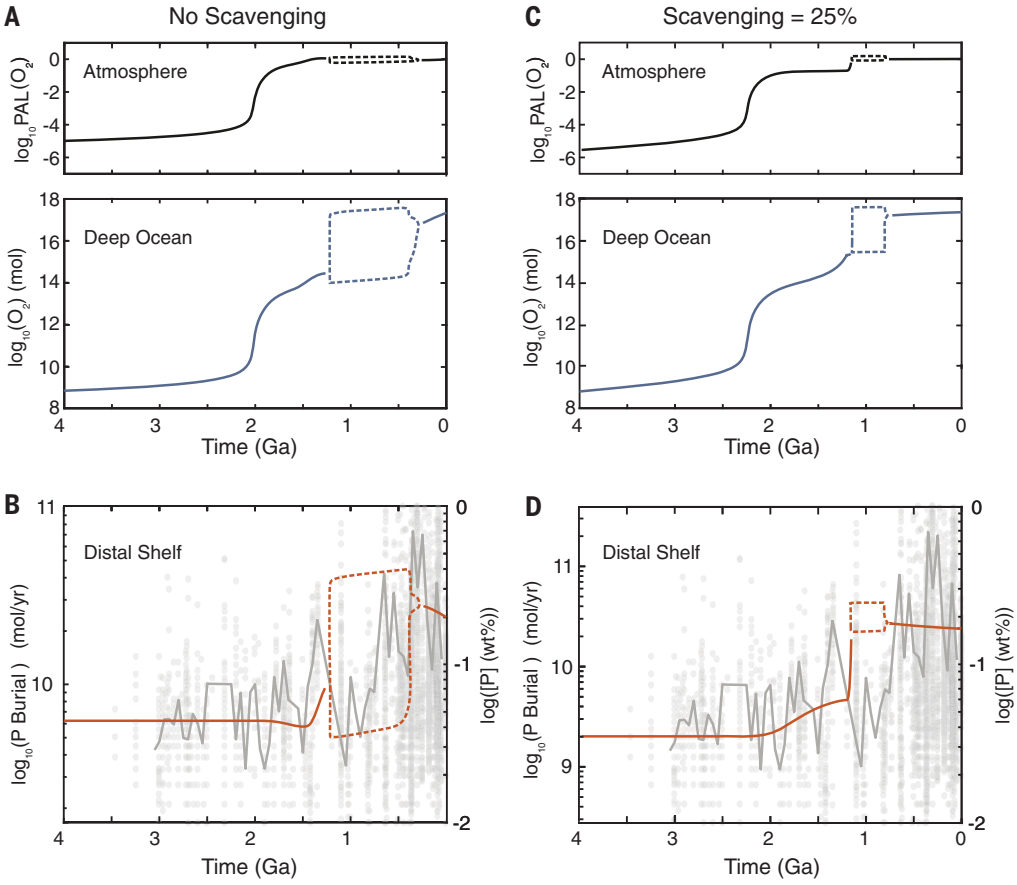



**Fig. 5. Possible $O_2$ evolution over Earth's history.** The model is run from 4 Ga to present, subject to a decrease in reductant input, illustrative of a gradual shift in net redox and demonstrating possible evolution of surface $O_2$ levels. Dotted boxes represent the boundaries of limit cycles in the solution. (**A** and **C**) Atmosphere and deep-ocean $O_2$ abundances. (**B** and **D**) Total shelf P burial rates (red; left axis) compared with P abundance in marine shales (right axis) (*29*). The gray line shows the 50-Myr binned moving average through compilation data. wt %, weight %. (A and B) No iron-bound scavenging and redox dependencies of 50% ($P_{auth}$) and 25% ($P_{org}$) (*32*), with a starting $O_2$ consumption via reductants of 45 × $10^{12}$ mol/year and a linear decrease. (C and D) Iron-bound scavenging included with a maximum rate of 5 × $10^{10}$ mol P/year. Plots show stronger redox dependencies (90% $P_{auth}$, 50% $P_{org}$) (*28*). Starting $O_2$ consumption via reductants was 45 × $10^{12}$ mol/year, with an exponential decrease.

are anoxic because of the low $O_2$ supply. A GOE occurs in the model when consumption of atmospheric $O_2$ by means of reduced gases is reduced to a value lower than the total $O_2$ source from $C_{org}$ burial (at ~1.5 × $10^{13}$ mol $O_2$ equivalent). Atmospheric $O_2$ rises by several orders of magnitude to ~0.06 to 0.25 PAL, which is broadly consistent with several estimates for the mid-Proterozoic (0.01 to 0.1 PAL) (*37*), and the $O_2$ balance is primarily controlled by oxidative weathering (*30*). Deep-ocean $O_2$ concentration also rises to ~0.1% of the modern value, whereas the proximal and distal shelf environments remain anoxic because $O_2$ diffusion from the atmosphere is not sufficient to outweigh $O_2$ demand.

Under a further decrease in reductant input, shelf environments undergo rapid oxygenation events. Gradually rising atmospheric $O_2$ concentrations cause a step change in water-column redox due to the positive feedback between bottom-water $O_2$ concentration and net P burial. Gradual ventilation of shelf bottom waters results in greater net removal of P, reducing overall $O_2$ demand and promoting further oxygenation. As the input of reduced gas declines, this transition happens first in the proximal shelf and then in the distal shelf environment because the latter is more strongly buffered against oxygenation, owing to the upwelling of P from anoxic deeper waters. Finally, as reductant input declines further, the deep ocean becomes fully oxygenated (reductant input of <0.5 × $10^{13}$ mol $O_2$ equivalent). The positive feedback between dissolved $O_2$ concentration and net P burial again causes a rapid transition. In our model, the deep ocean is oxygenated when atmospheric $O_2$ reaches 0.7 to 0.8 PAL, which is consistent with values reported in other studies (*38*).

Figure 3C demonstrates the degree of sedimentary P recycling in the model (shown as $C_{org}$:$P_{org}$). These changing ratios reflect the positive feedbacks between bottom-water redox and P recycling. This P recycling is dependent only on the extent of bottom-water anoxia, and thus the model predicts a substantial degree of recycling before the GOE. However, under global ferruginous conditions in the early Archean, P may have been more effectively trapped in the sediment (*29*) (a test of this can be found in the supplementary materials). Nevertheless, this P trap would ultimately help to stabilize $O_2$ at low levels by reducing surface ocean productivity.

The oxygenation of the distal shelf environment has the potential for oscillatory behavior (limit cycles), denoted by the break in steady-state lines in Fig. 3. The oxygenation of the entire shelf environment results in a large reduction in overall $C_{org}$ burial rates (because the shelves are the major locus of $C_{org}$ burial). Over geological time scales, this reduction in C burial sufficiently reduces the $O_2$ content of the atmosphere to return the shelves to anoxia. Figure 4 shows this cyclic response of the model under reductant inputs of 1 and 2.5 × $10^{13}$ mol $O_2$ equivalent per year. The cyclic regime includes temporary oxygenation of both the distal shelf and the deep ocean, as the rapid oxygenation event results in increased supply of oxic water by means of down-welling. These ocean oxygenation events (OOEs) last between 2 and 5 Myr for the range of redox dependencies tested in the model and occur on approximate 5- to 20-Myr time scales.

Our model demonstrates that gradual oxygenation of Earth's surface over time results in distinct oxygenation events. This occurs because, in our model, the atmosphere, continental shelves, and deep ocean act as distinct compartments of the Earth system (*39*, *40*), which are controlled by local rather than global feedbacks (*41*). First, a "great oxidation" of the atmosphere occurs, followed by oxygenation of near-shore shelf environments, and then distal shelf environments. This oxygenation of the whole shelf is manifest as an oscillating solution, which would likely lead to a series of OOEs. Finally, the deep ocean becomes resiliently oxygenated. This sequence of events tracks the apparent oxygenation history of Earth as recorded by multiple redox proxies,

including cycling between oxic and anoxic deep-ocean states during the Neoproterozoic and early Paleozoic (*6*). Our model neglects some potentially stabilizing negative feedbacks, such as the climate-silicate weathering link, although models that do include these processes still exhibit rapid shifts in marine anoxia (*42*).

In Fig. 5, we examine the potential for our model to recreate Earth's oxygenation history by performing transient model simulations under a continuous decline in reductant input constrained by mantle thermal evolution (*43*). In light of the potential for high Archean reductant availability and outgassing, and considering the relatively high rate of $O_2$ production in our model, we test starting fluxes of $4.5 \times 10^{13}$ mol $O_2$ equivalent per year at 4 Ga (*35*, *43*). With or without the inclusion of open-water P scavenging, the model is able to recreate the broad observed pattern of atmospheric and oceanic oxygenation over Earth's history. This includes a "great oxidation" of the atmosphere at around 2.5 to 2 Ga and unstable oxygenation of the deeper ocean starting at around 1 Ga, which continues until permanent oxygenation of the deep ocean is established at around 0.7 to 0.4 Ga.

We compare our model results with a compilation of P concentrations from marine shales (*29*) (Fig. 5). The data show an approximate fourfold increase in P weight percent between the Precambrian and Phanerozoic baselines, and this corresponds to an increase in P burial rate in the model shelf environment between the Proterozoic and oxygenated ocean states. Our model does not calculate sediment P weight percentages but does produce an upward baseline shift in the shelf sedimentary phosphorus burial rate, which is qualitatively consistent with the data. This increase in P burial is much greater when open-ocean scavenging is included, as the shutdown of scavenging on deep-ocean oxygenation results in substantially more P remaining in the system, in exactly the manner described by Reinhard *et al.* (*29*). We also note a two- to fourfold increase in shelf P burial rates when the model does not include open-ocean scavenging. This occurs simply because more P is trapped in the sediments when the shelves become oxygenated.

The reductant-driven oxygenation of the surface system that we analyze here, although plausible, is only one way in which to drive gradual net surface redox changes over time. Others include the long-term build-up of $C_{org}$ in the crust (*13*, *43*), the escape of hydrogen to space (*44*), or a gradually increasing supply of P to the ocean. Net redox changes driven by continual removal of hydrogen or $C_{org}$ from Earth's surface should operate in a similar way to the addition of reduced gases that we explore here. Notably, in these model runs, under a fixed present-day rate of riverine phosphate delivery and no scavenging flux, the model predicts a declining inventory of ocean P and declining rates of productivity and C burial over Earth's history as sedimentary P recycling is curtailed (full details can be found in fig. S4). When scavenging is considered, the P inventory and overall $C_{org}$ burial rates are broadly static through early Earth history and increase slightly when the deep ocean is oxygenated. Both model outputs reproduce the increasing P concentrations in shales through time (*29*), and although we cannot produce a $\delta^{13}C$ record with this model (because it lacks an inorganic C cycle), it would likely be consistent with the geological record, first because the model can produce either an increase or decrease in $C_{org}$ burial and second because ocean $\delta^{13}C$ is buffered by the adjustment of oxidative weathering rates at low $O_2$ (*30*) and by higher rates of inorganic C degassing and deposition on the early Earth (*45*). Nevertheless, although P input over Earth's history is highly uncertain (*46*), it is commonly assumed that the P inventory, along with productivity and C burial rates, has increased substantially over time. Thus, in the supplementary materials, we rerun our model under a varying riverine P input and show that the stepwise oxygenation events are indeed reproducible when Earth's gradual redox shift is driven by a steady increase in P supply to the oceans.

We demonstrate here that relationships between the global P, C, and O cycles are fundamental to understanding the oxygenation history of Earth. Our model confirms that observed oxygenation events throughout Earth's history may be driven by well-defined internal system feedbacks between these cycles, without the requirement for extensive external forcing. The results of this analysis are far-reaching. It appears that oxygenation of Earth's surface did not require any biological advances beyond simple photosynthetic cyanobacteria and was simply a matter of time, which substantially increases the possibility of high-$O_2$ worlds existing elsewhere.

## REFERENCES AND NOTES


1. J. Farquhar, H. Bao, M. Thiemens, *Science* **289**, 756–759 (2000).
2. N. J. Planavsky *et al.*, *Science* **346**, 635–638 (2014).
3. D. S. Hardisty *et al.*, *Earth Planet. Sci. Lett.* **463**, 159–170 (2017).
4. S. W. Poulton, D. E. Canfield, *Elements* **7**, 107–112 (2011).
5. L. Och, G. Shields-Zhou, *Earth Sci. Rev.* **110**, 26–57 (2012).
6. S. K. Sahoo *et al.*, *Geobiology* **14**, 457–468 (2016).
7. A. J. Krause *et al.*, *Nat. Commun.* **9**, 4081 (2018).
8. N. J. Butterfield, *Palaeontology* **58**, 5–17 (2015).
9. J. J. Brocks *et al.*, *Nature* **548**, 578–581 (2017).
10. K. M. Meyer, A. Ridgwell, J. L. Payne, *Geobiology* **14**, 207–219 (2016).
11. T. M. Lenton, R. A. Boyle, S. W. Poulton, G. A. Shields-Zhou, N. J. Butterfield, *Nat. Geosci.* **7**, 257–265 (2014).
12. L. R. Kump, M. E. Barley, *Nature* **448**, 1033–1036 (2007).
13. C.-T. A. Lee *et al.*, *Nat. Geosci.* **9**, 417–424 (2016).
14. I. H. Campbell, C. M. Allen, *Nat. Geosci.* **1**, 554–558 (2008).
15. F. Horton, *Geochem. Geophys. Geosyst.* **16**, 1723–1738 (2015).
16. A. H. Knoll, M. A. Nowak, *Sci. Adv.* **3**, e1603076 (2017).
17. T. M. Lenton, S. J. Daines, *Emerg. Top. Life Sci.* **2**, 267–278 (2018).
18. C. Goldblatt, T. M. Lenton, A. J. Watson, *Nature* **443**, 683–686 (2006).
19. M. W. Claire, D. C. Catling, K. J. Zahnle, *Geobiology* **4**, 239–269 (2006).
20. T. A. Laakso, D. P. Schrag, *Geobiology* **15**, 366–384 (2017).
21. T. A. Laakso, D. P. Schrag, *Earth Planet. Sci. Lett.* **388**, 81–91 (2014).
22. T. Tyrrell, *Nature* **400**, 525–531 (1999).
23. M. D. Krom, R. A. Berner, *Geochim. Cosmochim. Acta* **45**, 207–216 (1981).
24. K. C. Ruttenberg, R. A. Berner, *Geochim. Cosmochim. Acta* **57**, 991–1007 (1993).
25. T. Jilbert, C. P. Slomp, *Geochim. Cosmochim. Acta* **107**, 155–169 (2013).
26. P. Van Cappellen, E. D. Ingall, *Paleoceanography* **9**, 677–692 (1994).
27. C. P. Slomp, J. Thomson, G. J. De Lange, *Geochim. Cosmochim. Acta* **66**, 1171–1184 (2002).
28. I. Tsandev, C. P. Slomp, *Earth Planet. Sci. Lett.* **286**, 71–79 (2009).
29. C. T. Reinhard *et al.*, *Nature* **541**, 386–389 (2017).
30. S. J. Daines, B. J. W. Mills, T. M. Lenton, *Nat. Commun.* **8**, 14379 (2017).
31. I. C. Handoh, T. M. Lenton, *Global Biogeochem. Cycles* **17**, 1092 (2003).
32. C. P. Slomp, P. Van Cappellen, *Biogeosciences* **4**, 155–171 (2007).
33. I. Tsandev, C. P. Slomp, P. Van Cappellen, *Global Biogeochem. Cycles* **22**, GB4004 (2008).
34. P. Kharecha, J. F. Kasting, J. Siefert, *Geobiology* **3**, 53–76 (2005).
35. J. Krissansen-Totton, G. N. Arney, D. C. Catling, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4105–4110 (2018).
36. A. A. Pavlov, J. F. Kasting, *Astrobiology* **2**, 27–41 (2002).
37. D. C. Catling, M. W. Claire, *Earth Planet. Sci. Lett.* **237**, 1–20 (2005).
38. E. A. Sperling *et al.*, *Nature* **523**, 451–454 (2015).
39. K. Wallmann, *Global Biogeochem. Cycles* **17**, 1084 (2003).
40. W. S. Broecker, T. H. Peng, *Tracers in the Sea* (Eldigio Press, 1982).
41. C. J. Bjerrum, D. E. Canfield, T. W. Dahl, "Chasing Neoptoerozoic Atmospheric Oxygen Ghosts," abstract PP21D-06, presented at the American Geophysical Union Fall Meeting 2016, San Francisco, CA, 12 to 16 December 2016.
42. T. M. Lenton, S. J. Daines, B. J. W. Mills, *Earth Sci. Rev.* **178**, 1–28 (2018).
43. J. M. Hayes, J. R. Waldbauer, *Philos. Trans. R. Soc. B* **361**, 931–950 (2006).
44. D. C. Catling, K. J. Zahnle, C. McKay, *Science* **293**, 839–843 (2001).
45. G. A. Shields, B. J. W. Mills, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4318–4323 (2017).
46. B. Mills, T. M. Lenton, A. J. Watson, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9073–9078 (2014).
47. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, *Nature* **506**, 307–315 (2014).


## ACKNOWLEDGMENTS

We thank C. Slomp and K. Wallmann for sending computer code; we also thank the reviewers of this work for constructive and useful comments. **Funding:** L.J.A. is funded by a Leeds Anniversary Research Scholarship. B.J.W.M. acknowledges support from a University of Leeds Academic Fellowship. S.W.P. acknowledges support from a Leverhulme Research Fellowship and a Royal Society Wolfson Research Merit Award. B.J.W.M. and S.W.P. are funded by the UK Natural Environment Research Council (NE/R010129/1 and NE/S009663/1). **Author contributions:** L.J.A. and B.J.W.M. designed the research and developed the model. L.J.A. performed model runs. All authors 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. Model code and output data are available from the corresponding author on request.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1333/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
Tables S1 to S3
References (*48*–*52*)

11 April 2019; accepted 17 October 2019
Published online 10 December 2019
10.1126/science.aax6459

MOLECULAR BIOLOGY

# DNA loop extrusion by human cohesin

Iain F. Davidson, Benedikt Bauer, Daniela Goetz*, Wen Tang, Gordana Wutz, Jan-Michael Peters†

**Eukaryotic genomes are folded into loops and topologically associating domains, which contribute to chromatin structure, gene regulation, and gene recombination. These structures depend on cohesin, a ring-shaped DNA-entrapping adenosine triphosphatase (ATPase) complex that has been proposed to form loops by extrusion. Such an activity has been observed for condensin, which forms loops in mitosis, but not for cohesin. Using biochemical reconstitution, we found that single human cohesin complexes form DNA loops symmetrically at rates up to 2.1 kilo–base pairs per second. Loop formation and maintenance depend on cohesin's ATPase activity and on NIPBL-MAU2, but not on topological entrapment of DNA by cohesin. During loop formation, cohesin and NIPBL-MAU2 reside at the base of loops, which indicates that they generate loops by extrusion. Our results show that cohesin and NIPBL-MAU2 form an active holoenzyme that interacts with DNA either pseudo-topologically or non-topologically to extrude genomic interphase DNA into loops.**

In eukaryotic interphase cells, DNA is folded over long distances into loops and topologically associating domains (TADs) (*1*–*3*), which contribute to gene regulation and recombination (*4*, *5*). These structures depend on cohesin (*6*–*9*), a ring-shaped ATPase (Fig. 1A) that is also essential for sister chromatid cohesion and belongs to the SMC (structural maintenance of chromosomes) family of protein complexes found in all kingdoms of life [(*10*); reviewed in (*11*)].

How cohesin forms loops and TADs is unknown, but one hypothesis posits that cohesin extrudes DNA loops until it encounters convergently oriented "loop anchor" DNA sequences that are bound by the zinc finger protein CTCF (*12*, *13*). This hypothesis could explain why cohesin and CTCF colocalize (*14*, *15*) at TAD boundaries (*1*, *2*) and loop anchors (*3*), why these contain CTCF sites in convergent orientations (*3*), and why increasing cohesin's residence time on chromatin causes the formation of longer loops (*6*, *9*, *16*) and relocalization of cohesin into axial chromosomal domains called "vermicelli" (*13*, *17*). The loop extrusion hypothesis is also consistent with cohesin being highly mobile in the genome (*18*, *19*) and is supported by the observation that the related SMC complex condensin can extrude loops in vitro (*20*), a property that can explain DNA folding in mitotic chromosomes (*21*–*23*). However, there is no evidence that cohesin can form DNA loops by extrusion or that cohesin possesses a motor activity that would enable such a process, and it has been pointed out that loops could also be formed by sporadic encounters of distant DNA sequences (*11*).

Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), 1030 Vienna, Austria.
*Present address: AMS Advanced Medical Services GmbH, 68167 Mannheim, Germany.
†Corresponding author. Email: peters@imp.ac.at

## ATP hydrolysis–dependent DNA looping by cohesin and NIPBL-MAU2

We therefore analyzed whether recombinant human cohesin$^{STAG1}$ complexes (fig. S1J) (*24*) can form DNA loops by using an assay (*25*) in which yeast condensin performs loop extrusion (*20*). In these experiments, linear 48.5-kbp λ-phage DNA molecules were tethered at both ends to a glass surface in a flow cell, stained with Sytox Orange, stretched by perpendicular buffer flow, and imaged by total internal reflection (TIRF) microscopy in the presence of cohesin$^{STAG1}$ and ATP to visualize DNA looping (Fig. 1B). However, under these conditions, no loop formation was observed (0/372 DNAs analyzed; fig. S1A). Because recombinant cohesin could differ from endogenous cohesin, we generated HeLa cells in which all SCC1 alleles were modified to encode SCC1-Halo-FLAG fusion proteins and affinity-purified a mixture of cohesin$^{STAG1}$ and cohesin$^{STAG2}$ (Fig. 1C), in which cohesin$^{STAG2}$ was more abundant than cohesin$^{STAG1}$ by a factor of 3 (*26*). However, this preparation also did not cause loop formation (0/503 DNAs analyzed; Fig. 1E, fig. S1B, and movie S1).

Because condensin formed loops (*20*) under conditions in which cohesin failed to do so, we next considered differences between cohesin and condensin. Although both complexes share a ring-like architecture in which two elongated SMC proteins are connected via their hinge domains and a kleisin subunit [reviewed in (*11*)], condensins are pentamers that contain two HAWKs (HEAT repeat proteins associated with kleisins) (*27*), whereas cohesin complexes are tetramers containing only one HAWK. In somatic human cells, this can either be STAG1 or STAG2 (Fig. 1A) (*28*). However, cohesin is regulated by three additional HAWKs (Fig. 1A). NIPBL is thought to load cohesin onto DNA (*29*, *30*), whereas PDS5A and PDS5B have multiple functions, including release of cohesin from DNA (*31*), enabling cohesin acetylation (*32*), and contributing to TAD boundaries (*9*). We therefore tested whether cohesin forms loops in the presence of one of these HAWKs.

When we added recombinant PDS5A or PDS5B (fig. S1I) to HeLa cohesin, no DNA looping was observed (0/493 and 0/476 DNAs analyzed, respectively; Fig. 1E and fig. S1, C and D). However, when recombinant NIPBL bound to MAU2 (Fig. 1D) was combined with either HeLa cohesin or recombinant cohesin$^{STAG1}$, we observed looping in 238/511 (47%) and 142/344 (42%) of DNA molecules, respectively (Fig. 1, E, F, G, and M, fig. S2A, and movies S2 and S3). Looping was also observed in the absence of buffer flow [227/372 (61%) of DNAs], indicating that loops were formed actively by cohesin rather than passively by stretching DNA (Fig. 1H and fig. S2, B and C). Loops formed at a single position and incorporated DNA symmetrically from both ends (30/30 looping events; Fig. 1F and fig. S3, A and B), unlike loops formed by condensin, which extrudes DNA asymmetrically (*20*). Occasionally the base of cohesin loops moved along the DNA (Fig. 1, G and H). Similar observations were previously made using condensin mutated in a kleisin "safety belt" that normally encircles DNA (*20*), raising the possibility that cohesin lacks such a safety belt. Cohesin loops reached maximal size within 45 s from the onset of looping, and 424/447 (95%) of them were maintained during subsequent imaging for ~8 min. The remaining loops were released, returning DNA to its original stretched conformation (fig. S1E). Infrequent spontaneous cleavage of DNA likewise released looped DNA into its original shape (fig. S1F), ruling out the possibility that the observed structures were caused by recombination. The two "arms" of DNA loops were usually closely aligned (Fig. 1F) but occasionally moved apart from each other, revealing that they were only connected at the base of the loop (see figures below). Open DNA loops formed by cohesin could also be observed after paraformaldehyde fixation (Fig. 1I).

HeLa and recombinant cohesin formed loops at mean rates of 1 kbp $s^{-1}$ and 0.8 kbp $s^{-1}$, respectively (maximal rate 2.1 kbp $s^{-1}$; Fig. 1J and fig. S3, C and D), similar to the rate reported for yeast condensin [maximal rate 1.5 kbp $s^{-1}$ (*20*)]. The rate correlated inversely with the distance between the tethered DNA ends (Pearson correlation coefficient $r = -0.35$, $P = 0.001$; fig. S3E). Looping also occurred more frequently on DNA molecules in which the tethered ends were closely spaced (fig. S3F). These findings suggest that looping by cohesin, like looping by condensin (*20*), is prevented when the force required to bend DNA exceeds a threshold.

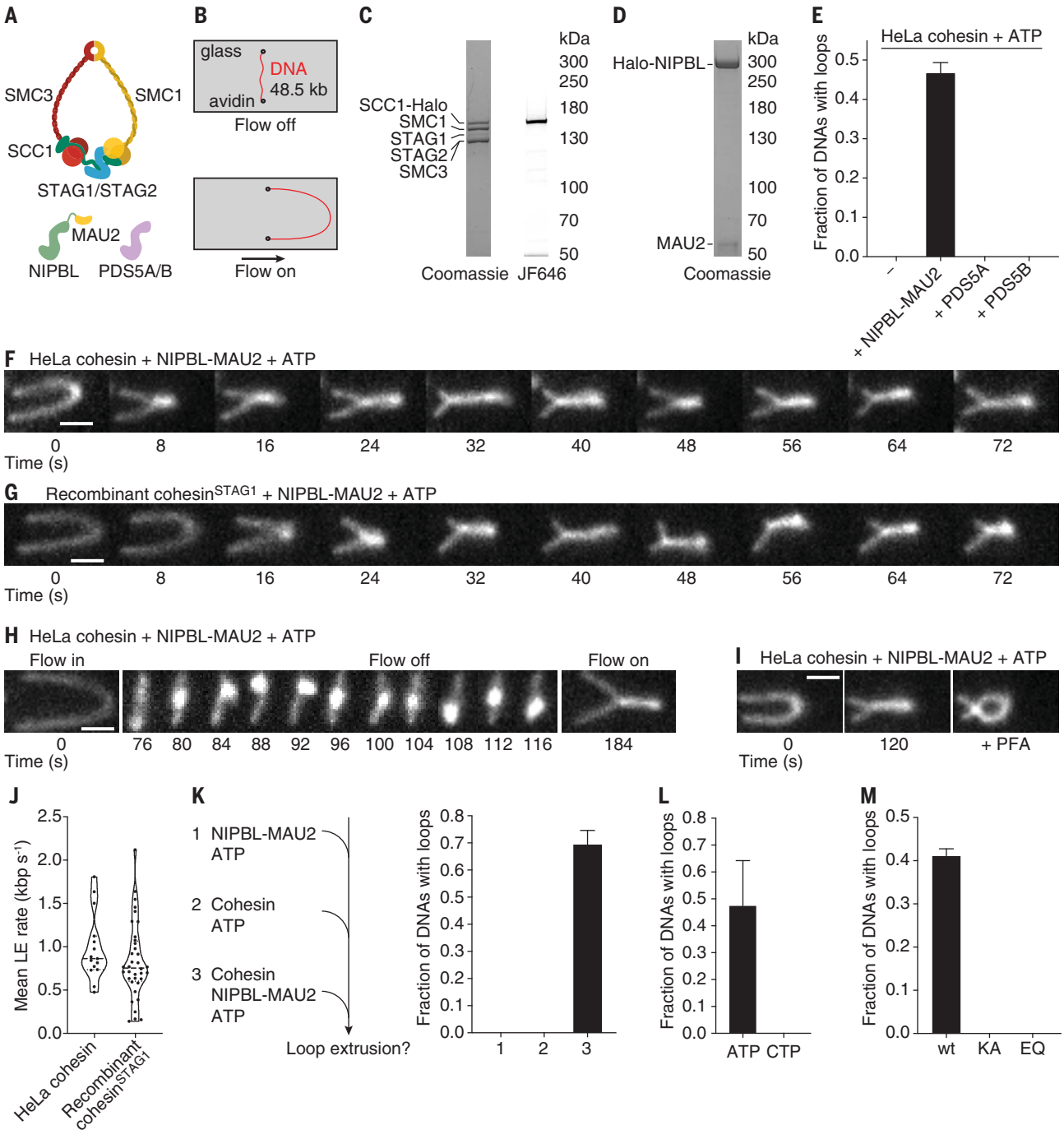



**Fig. 1. Human cohesin extrudes DNA loops in an NIPBL-MAU2– and ATP hydrolysis–dependent manner.** (**A**) Illustration of cohesin, PDS5A/B, and NIPBL-MAU2. (**B**) Illustration of a flow cell containing tethered λ-DNA molecules stretched by perpendicular flow. (**C**) Coomassie staining of HeLa cohesin after SDS-PAGE. JF646-HaloTag ligand was visualized by epi-red excitation. (**D**) Coomassie staining of recombinant NIPBL-MAU2 after SDS-PAGE. (**E**) Fraction of DNAs that formed loops (mean ± SD from three independent experiments) in the presence of the indicated components; *n* = 503, 511, 493, 476 DNAs analyzed (from left to right). (**F** and **G**) DNA looping in the presence of the indicated components. DNA was visualized by Sytox Orange staining. Scale bar, 2 μm. (**H**) As (F), except buffer flow was paused prior to loop extrusion (left). A DNA loop, visible as a bright dot, formed in the absence of buffer flow (center) and was extended by resumption of perpendicular flow (right). (**I**) As (F), except paraformaldehyde (PFA) was subsequently introduced into the flow cell, which allowed visualization of the "arms" of the DNA loop. (**J**) Mean loop extrusion rate in the presence of NIPBL-MAU2, ATP, and HeLa or recombinant cohesin$^{STAG1}$. Median and quartiles are shown; *n* = 15, 36 DNAs analyzed (from left to right). (**K**) Fraction of DNAs (mean ± SD from three independent experiments) that formed loops after sequential introduction of the indicated components; *n* = 327 DNAs analyzed. (**L**) Fraction of DNAs (mean ± SD from three independent experiments) that formed loops in the presence of HeLa cohesin, NIPBL-MAU2, and ATP or CTP; *n* = 426, 366 DNAs analyzed (from left to right). (**M**) Fraction of DNAs (mean ± SD from three independent experiments) that formed loops in the presence of NIPBL-MAU2, ATP, and wild-type (wt), SMC1 K38A/SMC3 K38A ATP binding–deficient (KA), or SMC1 E1157Q/SMC3 E1144Q ATP hydrolysis–deficient (EQ) recombinant cohesin$^{STAG1}$; *n* = 344, 377, 406 DNAs analyzed (from left to right).

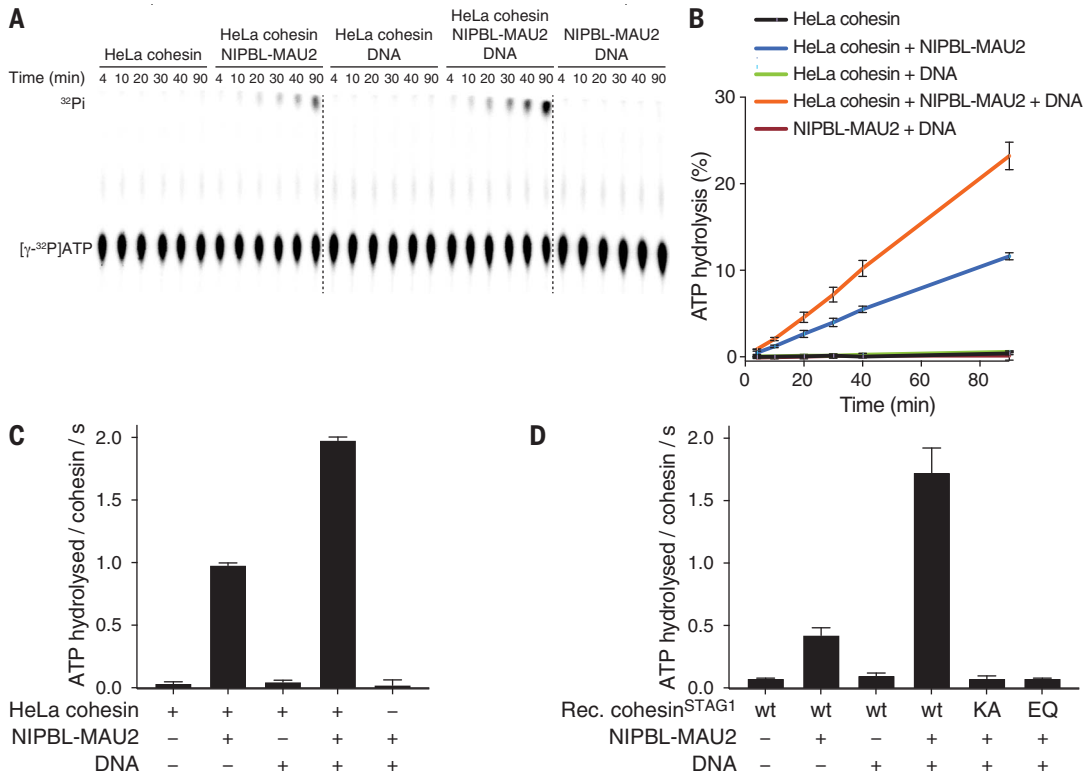



**Fig. 2. Cohesin's ATPase activity is enhanced by NIPBL-MAU2 and DNA.** (**A**) Example of thin-layer chromatography/autoradiography assay used to measure [γ-$^{32}$P]ATP hydrolysis in the presence of the indicated components. (**B**) ATP hydrolysis (mean ± SD from three independent experiments) determined from (A). (**C** and **D**) ATPase rates (mean ± SD from three independent experiments) in the presence of the indicated components determined using thin-layer chromatography/autoradiography.

To investigate whether NIPBL-MAU2 functions in looping by first binding DNA and then recruiting cohesin, we introduced NIPBL-MAU2 and cohesin sequentially into a flow cell (Fig. 1K). No looping was observed when NIPBL-MAU2 was introduced (0/327 DNAs; Fig. 1K and fig. S1G), nor when cohesin was added subsequently (0/327 DNAs; Fig. 1K). In contrast, looping occurred on 225/327 (60%) of DNA molecules when cohesin and NIPBL-MAU2 were introduced simultaneously (Fig. 1K), indicating that cohesin can only loop DNA in the presence of NIPBL-MAU2.

Like yeast cohesin (*33*, *34*), HeLa and recombinant cohesin$^{STAG1}$ hydrolyzed ATP slowly but were activated by as much as a factor of 30 by NIPBL-MAU2 and further by DNA (Fig. 2). Under these conditions, 2 (HeLa cohesin) or 1.7 (cohesin$^{STAG1}$) ATP molecules were hydrolyzed per cohesin complex per second (Fig. 2, C and D). DNA looping was supported by ATP [199/426 (47%) of DNAs analyzed] but not cytidine triphosphate (CTP) (0/366 of DNAs; Fig. 1L and fig. S1H) and did not occur in the presence of ATP binding–deficient or ATP hydrolysis–deficient mutants of cohesin$^{STAG1}$ [0/377 and 0/406 of DNAs, respectively, compared to 142/344 (41%) of DNAs in the presence of wild-type cohesin$^{STAG1}$; Fig. 1M and fig. S1, J and K]. Cohesin can therefore form DNA loops in a manner that depends on its ATPase activity and on NIPBL-MAU2.

### Cohesin and NIPBL-MAU2 are located at the base of DNA loops during extrusion

If cohesin loops DNA by extrusion, it must be present at the base of loops. Indeed, Janelia Fluor 646 (JF646)–labeled HeLa and recombinant cohesin$^{STAG1}$ (Fig. 1C and fig. S1J), as well as NIPBL-MAU2 (Fig. 3G), were present at the base of loops and moved against buffer flow during their formation (Fig. 3, A, B, and H, fig. S4, and movies S4 and S5). To determine whether looping is mediated by single or oligomeric cohesin complexes, we measured the fluorescence intensity of HeLa cohesin at the base of loops ($n$ = 62) and compared it to the intensity of cohesin bound to glass ($n$ = 815), where the oligomeric state of cohesin could be analyzed by photobleaching (for technical reasons, we were unable to detect bleaching of cohesin on DNA). For both cohesin populations, fluorescence measurements yielded single peaks with similar mean intensities (217 and 242 a.u., respectively; Fig. 3, C and D), and in the latter case most signals bleached in a single step, indicative of monomeric cohesin [37/50 (74%) of cohesin signals analyzed; Fig. 3E], or did not bleach [8/50 (16%)]. In rare cases, cohesin on glass exhibited a fluorescence intensity of ~500 a.u., indicative of a dimeric state, and bleached in two steps by going through a 250-a.u. intermediate [5/50 (10%) of cohesin signals analyzed; Fig. 3F]. These data support the hypothesis that cohesin loops DNA by extrusion and does so as a monomer.

### Cohesin translocates rapidly along DNA against the direction of buffer flow

Yeast condensin can unidirectionally and ATP-dependently translocate along extended DNA molecules at 0.06 kbp s$^{-1}$ (*35*). Our flow cells also contained some singly and doubly tethered DNA molecules that were stretched in the direction of buffer flow and not oriented perpendicular to it (fig. S2B, arrowheads). Along these, cohesin frequently translocated against the direction of buffer flow at a mean rate of 0.4 kbp s$^{-1}$ [164 translocating cohesin signals along 345 DNAs analyzed (0.48 translocating cohesin signals per DNA); Fig. 4, A, B, C, E, and F]. The fluorescence intensity distribution of translocating cohesin (Fig. 4D) closely resembled that of cohesin at the base of loops and on glass (Fig. 3, C and D), indicating that cohesin also translocated as a monomer. We did not observe cohesin translocating in the direction of buffer flow, possibly because such cohesin molecules had reached the ends of DNA molecules before they could be imaged. Like loop extrusion, cohesin translocation was strictly dependent on ATP and NIPBL-MAU2 (Fig. 4, E and F). The relationship between cohesin translocation and loop extrusion is unknown, but we speculate that translocation, which occurs at approximately half the rate of loop formation, represents a partial extrusion reaction in which cohesin translocates on a single DNA rather than along the two arms of a loop.

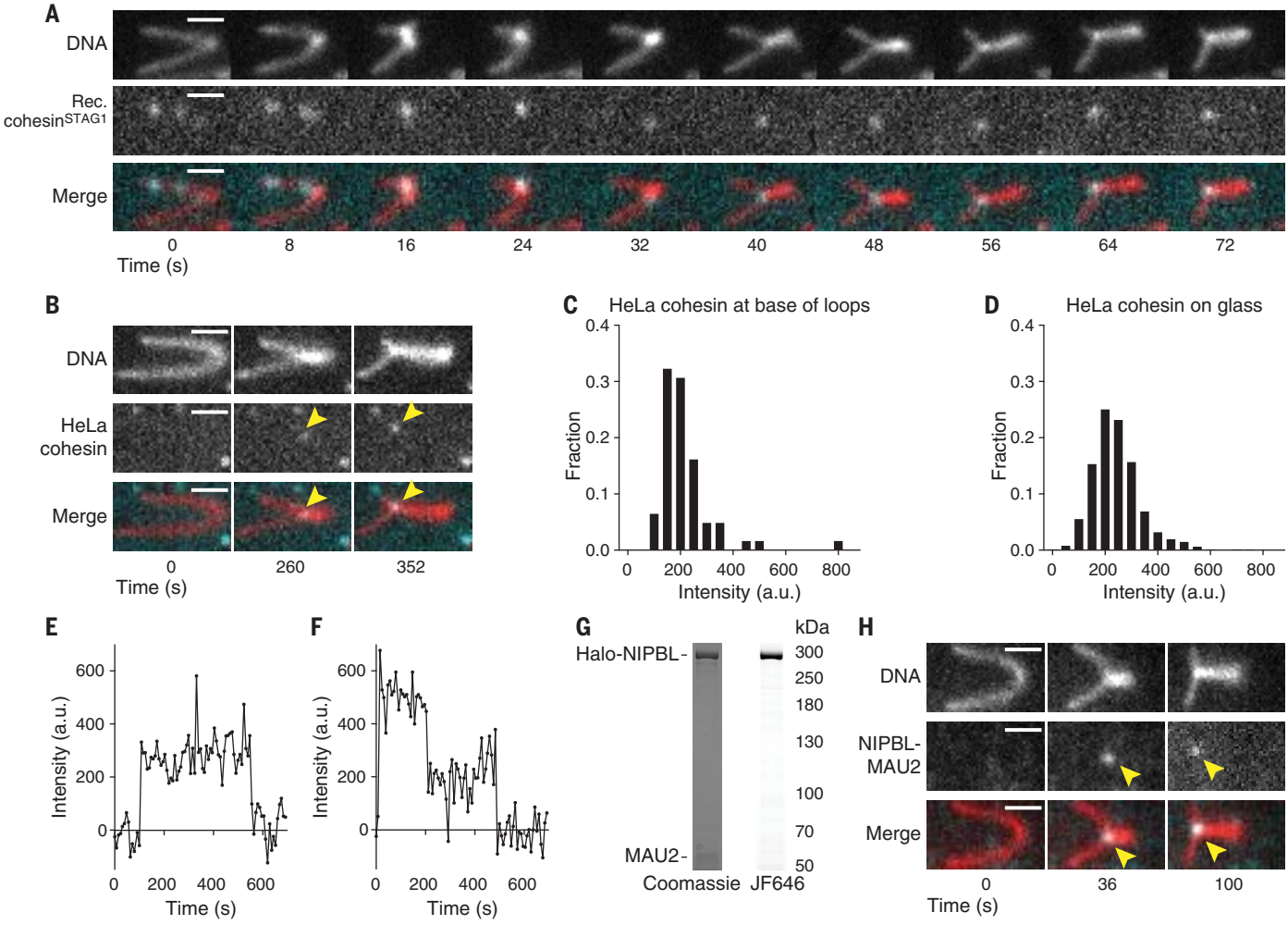



**Fig. 3. Cohesin and NIPBL-MAU2 are located at the base of DNA loops during extrusion.** (**A**) DNA looping and cohesin localization during flow-in of JF646-labeled recombinant cohesin$^{STAG1}$, NIPBL-MAU2, and ATP. DNA was visualized by Sytox Orange staining. Lower panel shows merged images. Scale bar, 2 μm. (**B**) As (A), except DNA was incubated with JF646-labeled HeLa cohesin. Yellow arrowheads indicate position of cohesin during loop extrusion. (**C** and **D**) Fluorescence intensity distribution of HeLa cohesin (C) at the base of DNA loops ($n$ = 62 cohesin signals) and (D) bound nonspecifically to the glass surface ($n$ = 815 cohesin signals). (**E**) Time trace of cohesin signal binding nonspecifically to the glass surface and bleaching in one step. (**F**) As (E), except cohesin bleached in two steps. (**G**) Coomassie staining of recombinant NIPBL-MAU2 after SDS-PAGE. JF646-HaloTag ligand was visualized by epi-red excitation. (**H**) As (A), except DNA was incubated with unlabeled HeLa cohesin, JF646-labeled NIPBL-MAU2, and ATP.

### ATP and NIPBL-MAU2 are required for ongoing loop extrusion by cohesin

The role of cohesin's ATPase activity and NIPBL-MAU2 in loop extrusion could be indirect or direct (or both) by promoting loading of cohesin onto DNA or by enabling cohesin to extrude DNA, respectively. To distinguish between these possibilities, we first initiated loop extrusion by HeLa cohesin, NIPBL-MAU2, and ATP for 8 min (Fig. 5A, step 1), then replaced the content of the flow cell with different combinations of NIPBL-MAU2, ATP, and CTP and monitored the loops for another 8 min (Fig. 5A, step 2). Throughout step 1, ~95% of loops remained stable (Fig. 5, F to I, step 1). When the content of the flow cell was replaced with NIPBL-MAU2 and ATP, 98/104 (94%) of the remaining loops persisted (Fig. 5, B and F, fig. S5A, and movie S6), indicating that loops were maintained by DNA-bound cohesin that did not exchange with freely diffusing cohesin. When we substituted ATP for CTP (Fig. 5, C and G, fig. S5B, and movie S7) or introduced ATP in the absence of NIPBL-MAU2 (Fig. 5, D and H, fig. S5C, and movie S8), only 2/95 (2%) and 3/92 (3%) of loops were maintained, respectively. ATP and NIPBL-MAU2 are therefore required for the maintenance of loops. These results, and our observations that cohesin can only loop DNA in the presence of NIPBL-MAU2 (Fig. 1K) and that NIPBL-MAU2 is located at the base of loops during their extrusion (Fig. 3H), suggest that NIPBL-MAU2 and cohesin's ATPase activity are both required for cohesin-mediated loop extrusion per se. Because 97% of DNA loops fell apart within 8 min after removal of unbound NIPBL-MAU2, most NIPBL-MAU2 must interact with cohesin transiently and dissociate from cohesin within this time period.

### The topology of cohesin-DNA interactions during loop extrusion

Cohesin is thought to mediate cohesion by topologically entrapping sister DNA molecules inside its ring structure (*36*), but cohesin can also entrap single DNAs (*37*). Thus, topological DNA entrapment could be a prerequisite for loop extrusion. Alternatively, cohesin could extrude loops "pseudo-topologically" by

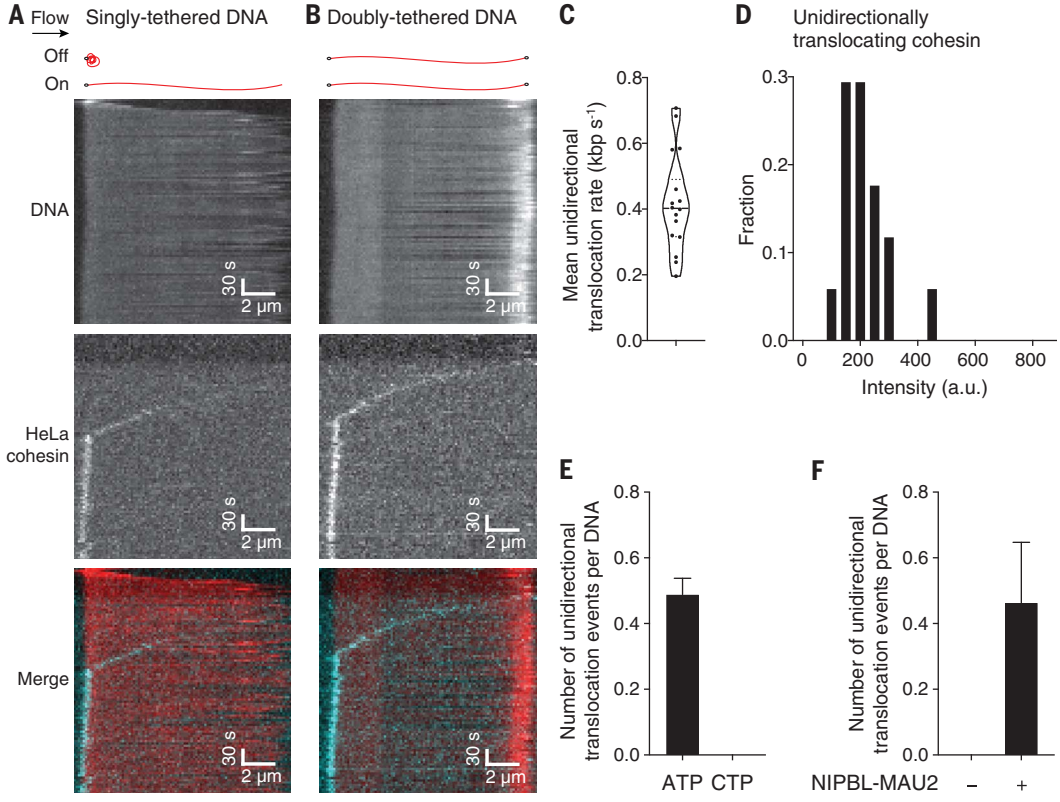



**Fig. 4. Cohesin translocates rapidly along DNA against the direction of buffer flow.** (**A** and **B**) Top: Illustrations of DNAs tethered at one or both ends and stretched in the direction of buffer flow. Lower panels: Kymographs of JF646-labeled HeLa cohesin translocating along singly tethered (A) and doubly tethered (B) DNA molecules against the direction of buffer flow in the presence of NIPBL-MAU2 and ATP. DNA was visualized by Sytox Orange staining. (**C**) Mean rate of JF646-labeled HeLa cohesin translocation against the direction of buffer flow in the presence of NIPBL-MAU2 and ATP. Median and quartiles are shown; $n$ = 16 translocation events analyzed. (**D**) Fluorescence intensity distribution of cohesin translocating unidirectionally against the direction of buffer flow; $n$ = 17 cohesin signals. (**E**) Number of JF646-labeled HeLa cohesin translocation events against the direction of buffer flow per DNA molecule (mean ± SD from three independent experiments) in the presence of NIPBL-MAU2 and either ATP or CTP; $n$ = 179, 178 DNAs analyzed (from left to right). (**F**) Number of JF646-labeled HeLa cohesin translocation events against the direction of buffer flow per DNA molecule (mean ± SD from three independent experiments) in the presence of ATP ± NIPBL-MAU2; $n$ = 177, 183 DNAs analyzed (from left to right).

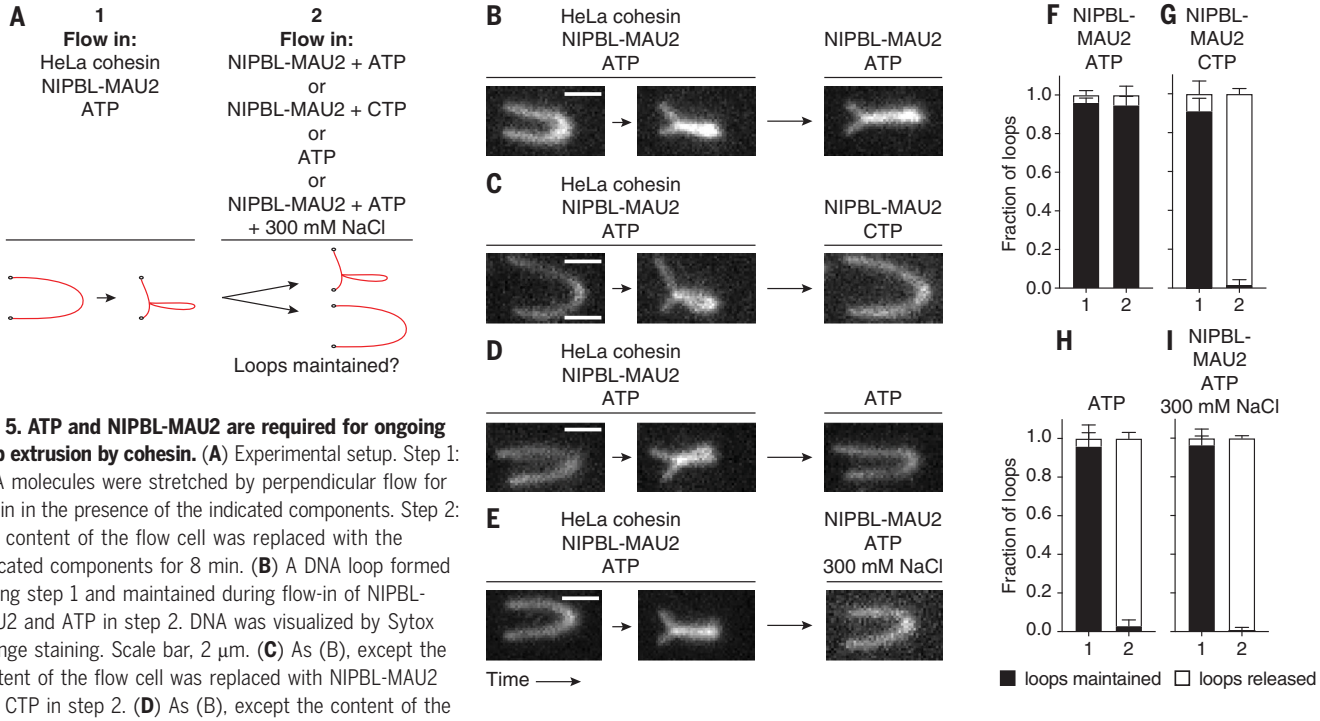



**Fig. 5. ATP and NIPBL-MAU2 are required for ongoing loop extrusion by cohesin.** (**A**) Experimental setup. Step 1: DNA molecules were stretched by perpendicular flow for 8 min in the presence of the indicated components. Step 2: The content of the flow cell was replaced with the indicated components for 8 min. (**B**) A DNA loop formed during step 1 and maintained during flow-in of NIPBL-MAU2 and ATP in step 2. DNA was visualized by Sytox Orange staining. Scale bar, 2 μm. (**C**) As (B), except the content of the flow cell was replaced with NIPBL-MAU2 and CTP in step 2. (**D**) As (B), except the content of the flow cell was replaced with ATP in step 2. (**E**) As (B), except the content of the flow cell was replaced with NIPBL-MAU2 and ATP in buffer containing 300 mM NaCl in step 2. The brightness of the step 2 image was enhanced to compensate for the reduced Sytox Orange signal in 300 mM NaCl. (**F** to **I**) Fraction of loops (mean ± SD from three independent experiments) maintained in the indicated conditions; $n$ = 108, 104, 97, 138 loops analyzed (from left to right).

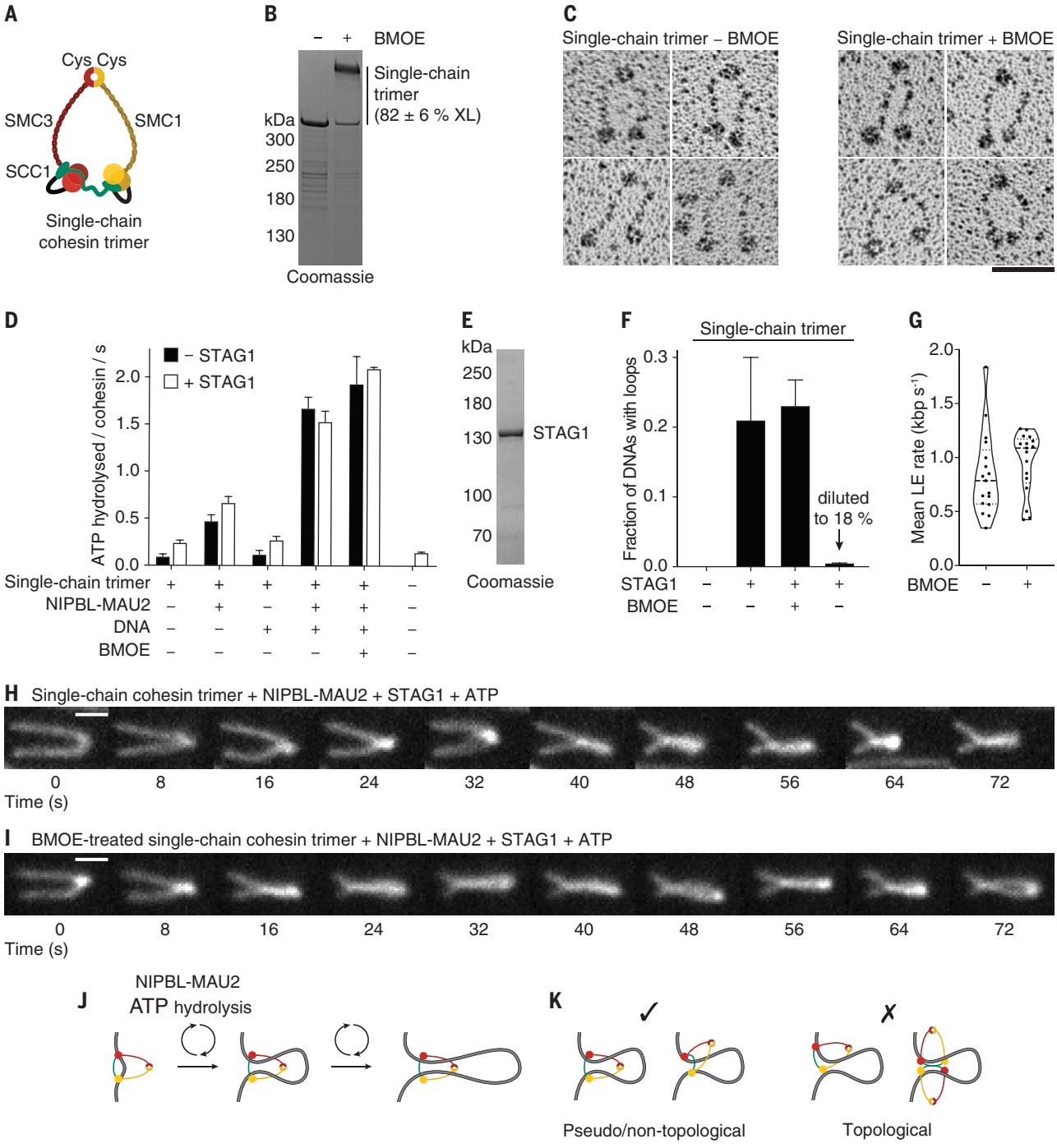



**Fig. 6. Loop extrusion by cohesin does not require opening of the tripartite ring.** (**A**) Illustration of single-chain trimer cohesin. (**B**) Coomassie staining of single-chain trimer cohesin incubated either with BMOE (+) or with BMOE that had been inactivated by preincubation with dithiothreitol (DTT) (–). Mobility shift indicates hinge–cross-linked molecules; percentage of cross-link efficiency is indicated (mean ± SD). (**C**) Rotary shadowing electron micrographs of single-chain cohesin incubated with BMOE ± DTT as described in (B). Scale bar, 50 nm. (**D**) ATPase rate (mean ± SD from three independent experiments) in the presence and absence of the indicated components, determined as in Fig. 2. (**E**) Coomassie staining of recombinant STAG1 after SDS-PAGE. (**F**) Fraction of DNAs that formed loops (mean ± SD from three independent experiments) in the presence of NIPBL-MAU2, ATP, and single-chain trimer cohesin incubated with BMOE ± DTT as described in (B) and then preincubated with or without STAG1; $n$ = 429, 481, 432, 387 DNAs analyzed (from left to right). (**G**) Mean loop extrusion rate in the presence of STAG1, NIPBL-MAU2, ATP, and single-chain trimer cohesin incubated with BMOE ± DTT as described in (B). Median and quartiles are shown; $n$ = 17, 17 DNAs analyzed (from left to right). (**H** and **I**) DNA looping in the presence of the indicated components. Single-chain trimer was incubated with BMOE ± DTT as described in (B). DNA was visualized by Sytox Orange staining. Scale bar, 2 μm. (**J** and **K**) Models of cohesin-dependent loop extrusion. (J) Ongoing loop extrusion is dependent on ATP hydrolysis by cohesin and on NIPBL-MAU2, which interacts with cohesin transiently. (K) Covalent fusion of all ring interfaces does not abolish loop extrusion by cohesin, consistent with models in which cohesin extrudes loops either non-topologically or pseudo-topologically.

threading a loop of DNA through its ring structure (Fig. 6K), or non-topologically by extruding in a manner that does not involve encircling DNA at all (*37*). To distinguish between these possibilities, we tested whether loop extrusion is resistant to high-salt treatment [a property reported for topological cohesin-DNA interactions (*29*)] and whether loop extrusion depends on the integrity of the cohesin ring and on the ability of cohesin's ring-forming subunits to transiently dissociate from each other, which would be required for topologically entrapping DNA.

To address the first question, we initiated loop formation in the presence of 50 mM NaCl (Fig. 5A, step 1) and then increased the salt concentration to 300 mM (Fig. 5A, step 2). Under these conditions, only 1/133 (<1%) of loops persisted (Fig. 5, E and I, and fig. S5D); such a result is consistent with non-topological DNA entrapment by cohesin, but also with the possibility that interactions between cohesin and NIPBL-MAU2 are salt-sensitive.

We next generated a version of cohesin$^{STAG1}$ in which the kleisin SCC1 can be cleaved by tobacco etch virus (TEV) protease (fig. S6, A and B) (*24*), mimicking cleavage of cohesin by separase in metaphase. TEV cleavage reduced but did not abolish cohesin recruitment to DNA (fig. S6C); however, it prevented loop extrusion [1/473 (<0.3%) DNAs; fig. S6D]. This result is consistent with the possibility that loop extrusion requires topological entrapment of DNA, but likewise with the possibility that SCC1's integrity is required for cohesin's extrusion function.

To distinguish between these possibilities, we engineered a form of cohesin in which all interfaces could be covalently linked to prevent topological DNA entrapment. To do this, we generated a 369-kDa in-frame fusion of SMC3, SCC1, and SMC1, which were connected by linker sequences (Fig. 6A). In this "single-chain trimer," we replaced all cysteine residues with serines and introduced a pair of cysteines in the hinge domains of SMC1 and SMC3 to allow cross-linking by bis-maleimidoethane (BMOE). Cross-linking efficiency was 82 ± 6%, as measured by SDS–polyacrylamide gel electrophoresis (PAGE) mobility shift (Fig. 6B). Rotary shadowing electron microscopy revealed that single-chain trimers formed ring-like structures indistinguishable from the ones formed by HeLa (*38*) and recombinant cohesin$^{STAG1}$ (*24*), except for the absence of a density corresponding to STAG1 or STAG2 (Fig. 6C; note that these rings appear "open" because SCC1 is difficult to visualize). Like recombinant "wild-type" cohesin$^{STAG1}$, single-chain cohesin hydrolyzed ~1.7 molecules of ATP per second in the presence of NIPBL-MAU2 and DNA (Fig. 6D). ATP hydrolysis rates were not reduced by BMOE treatment and, as observed for yeast cohesin (*34*), were unchanged in the presence of STAG1 (Fig. 6, D and E).

Although non–cross-linked single-chain trimer is an active ATPase, it did not form DNA loops (0/429 DNAs; Fig. 6F). However, when this protein was preincubated with recombinant STAG1, loop extrusion occurred frequently [100/481 (21%) of DNAs; Fig. 6, F and G, and movie S9]. Remarkably, treatment of single-chain trimer with BMOE neither prevented nor even reduced loop extrusion in the presence of STAG1 [98/432 (23%) of DNAs; Fig. 6, F and G, and movie S10]. To test whether residual non–cross-linked cohesin could account for this activity, we diluted non–cross-linked single-chain trimer bound to STAG1 to a concentration corresponding to the 18% that failed to be cross-linked in the previous experiments. This reduced the fraction of DNAs that were extruded from 21% to 0.5% (2/387 DNAs; Fig. 6F), which suggests that the residual amounts of non–cross-linked cohesin are not sufficient to explain loop extrusion after BMOE treatment. Instead, these results indicate that cohesin can extrude loops even when all its interfaces have been covalently linked and that cohesin can therefore perform this process without topologically entrapping DNA. This implies that cohesin performs loop extrusion by interacting with DNA either pseudo-topologically or non-topologically (Fig. 6K). Furthermore, these experiments revealed that STAG1 has an important role in loop extrusion, even though it is dispensable for cohesin's DNA-dependent ATPase activity.

## Discussion and conclusions

Our results show that in the presence of NIPBL-MAU2, single cohesin complexes can form DNA loops by extrusion in a manner that depends on cohesin's ATPase activity. How cohesin forms loops in the context of chromatin remains unknown. However, the high frequency and speed of loop extrusion in vitro, combined with the fact that the loop extrusion hypothesis can explain cellular phenomena such as the CTCF convergence rule (*3*) and vermicelli formation (*13*, *17*), strongly support the hypothesis that cohesin also forms loops and TADs by extrusion in vivo, thereby facilitating key processes such as gene regulation (*4*) and immunoglobulin gene recombination (*5*). Given the high speed of loop extrusion observed in our experiments (mean rate 1 kbp $s^{-1}$), current estimates of loop formation rates in HeLa cells [0.375 kbp $s^{-1}$ (*7*)], and the number of cohesin complexes that have been identified on chromatin [~160,000 during $G_1$ phase in HeLa cells (*26*)], an entire human genome could be extruded rapidly, possibly in a few minutes. This implies that loop extrusion is a highly dynamic and therefore regulatable process.

Our observation that NIPBL-MAU2 is not only required to initiate loop extrusion but is essential throughout this process has several important implications. First, it indicates that even though cohesin-NIPBL-MAU2 interactions are short-lived in vitro, the active loop-extruding holoenzyme is cohesin bound to NIPBL-MAU2. Like condensin, cohesin therefore needs two HAWKs for this process: STAG1 or STAG2 and NIPBL. In contrast, and as previously suggested (*9*, *34*), PDS5 proteins may be HAWKs that inhibit loop extrusion, as they can compete with NIPBL for binding to cohesin (*34*, *39*), do not stimulate cohesin's ATPase activity (*34*), and are required for TAD boundaries (*9*).

Second, and consistent with cellular observations (*16*, *34*, *40*), cohesin's ATPase and NIPBL-MAU2 cannot be required for cohesin loading alone but instead must have direct functions in loop extrusion (Fig. 6J). In the case of cohesin's ATPase activity, one of these functions could be the generation of force to move DNA, and in the case of NIPBL-MAU2 the stimulation of cohesin's ATPase activity (*33*, *34*) (Fig. 2). Given that cohesin's ATPase activity and NIPBL-MAU2 function in loop extrusion, it will be worthwhile to reinvestigate whether these factors do in fact load cohesin onto DNA or whether they are merely required to keep cohesin there—for example, by enabling processive loop extrusion. Consistent with this possibility, yeast cohesin's constitutive HAWK Scc3 (orthologous to human STAG1 and STAG2) is also needed for maintaining cohesin on DNA (*41*) and its NIPBL ortholog Scc2 is required for retaining cohesin on chromatin in $G_1$ phase (*42*). Such a scenario could also explain why a loading factor has never been identified for condensin. This could either be because condensin's HAWKs are performing this function or simply because there is also no bona fide loading complex for cohesin. Our identification of cohesin-NIPBL-MAU2 as a potent loop extrusion factor will enable addressing these questions and the mechanistic basis of cohesin-dependent loop extrusion.

## REFERENCES AND NOTES


1. E. P. Nora *et al.*, *Nature* **485**, 381–385 (2012).
2. J. R. Dixon *et al.*, *Nature* **485**, 376–380 (2012).
3. S. S. P. Rao *et al.*, *Cell* **159**, 1665–1680 (2014).
4. S. Cuartero *et al.*, *Nat. Immunol.* **19**, 932–941 (2018).
5. Y. Zhang *et al.*, *Nature* **573**, 600–604 (2019).
6. J. Gassler *et al.*, *EMBO J.* **36**, 3600–3618 (2017).
7. S. S. P. Rao *et al.*, *Cell* **171**, 305–320.e24 (2017).
8. W. Schwarzer *et al.*, *Nature* **551**, 51–56 (2017).
9. G. Wutz *et al.*, *EMBO J.* **36**, 3573–3599 (2017).
10. A. V. Strunnikov, V. L. Larionov, D. Koshland, *J. Cell Biol.* **123**, 1635–1648 (1993).
11. F. Uhlmann, *Nat. Rev. Mol. Cell Biol.* **17**, 399–412 (2016).
12. A. L. Sanborn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6456–E6465 (2015).
13. G. Fudenberg *et al.*, *Cell Rep.* **15**, 2038–2049 (2016).
14. V. Parelho *et al.*, *Cell* **132**, 422–433 (2008).
15. K. S. Wendt *et al.*, *Nature* **451**, 796–801 (2008).

16. J. H. I. Haarhuis *et al*., *Cell* **169**, 693–707.e14 (2017).
17. A. Tedeschi *et al*., *Nature* **501**, 564–568 (2013).
18. A. Lengronne *et al*., *Nature* **430**, 573–578 (2004).
19. G. A. Busslinger *et al*., *Nature* **544**, 503–507 (2017).
20. M. Ganji *et al*., *Science* **360**, 102–105 (2018).
21. K. Nasmyth, *Annu. Rev. Genet.* **35**, 673–745 (2001).
22. E. Alipour, J. F. Marko, *Nucleic Acids Res.* **40**, 11202–11212 (2012).
23. A. Goloborodko, M. V. Imakaev, J. F. Marko, L. Mirny, *eLife* **5**, e14864 (2016).
24. P. J. Huis in 't Veld *et al*., *Science* **346**, 968–972 (2014).
25. I. F. Davidson *et al*., *EMBO J.* **35**, 2671–2685 (2016).
26. J. Holzmann *et al*., *eLife* **8**, e46269 (2019).
27. J. N. Wells, T. G. Gligoris, K. A. Nasmyth, J. A. Marsh, *Curr. Biol.* **27**, R17–R18 (2017).
28. I. Sumara, E. Vorlaufer, C. Gieffers, B. H. Peters, J. M. Peters, *J. Cell Biol.* **151**, 749–762 (2000).
29. R. Ciosk *et al*., *Mol. Cell* **5**, 243–254 (2000).
30. T. S. Takahashi, P. Yiu, M. F. Chou, S. Gygi, J. C. Walter, *Nat. Cell Biol.* **6**, 991–996 (2004).
31. K. Shintomi, T. Hirano, *Genes Dev.* **23**, 2224–2236 (2009).
32. S. Vaur, A. Feytout, S. Vazquez, J.-P. Javerzat, *EMBO Rep.* **13**, 645–652 (2012).
33. Y. Murayama, F. Uhlmann, *Nature* **505**, 367–371 (2014).
34. N. J. Petela *et al*., *Mol. Cell* **70**, 1134–1148.e7 (2018).
35. T. Terakawa *et al*., *Science* **358**, 672–676 (2017).
36. C. H. Haering, A.-M. Farcas, P. Arumugam, J. Metson, K. Nasmyth, *Nature* **454**, 297–301 (2008).
37. M. Srinivasan *et al*., *Cell* **173**, 1508–1519.e18 (2018).
38. D. E. Anderson, A. Losada, H. P. Erickson, T. Hirano, *J. Cell Biol.* **156**, 419–424 (2002).
39. S. Kikuchi, D. M. Borek, Z. Otwinowski, D. R. Tomchick, H. Yu, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12444–12449 (2016).
40. J. Rhodes, D. Mazza, K. Nasmyth, S. Uphoff, *eLife* **6**, e30000 (2017).
41. M. B. Roig *et al*., *FEBS Lett.* **588**, 3692–3702 (2014).
42. M. Srinivasan *et al*., *eLife* **8**, e44736 (2019).

## ACKNOWLEDGMENTS

We thank C.H. Haering for sharing reagents, M. Madalinski for purifying JF646-HaloTag Ligand, G. Litos for purifying proteins, D. Drechsel for characterizing single-chain trimer cohesin, T. Lendl for image analysis support, and P. Pasierbek and other members of IMP/IMBA Biooptics facility for assistance. Electron microscopy was performed by the EM Facility of the Vienna BioCenter Core Facilities GmbH, member of Vienna BioCenter, Austria. **Funding:** Research in the laboratory of J.-M.P. is supported by Boehringer Ingelheim, the Austrian Research Promotion Agency (Headquarter grant FFG-852936), the European Research Council under the European Union's Horizon 2020 research and innovation program GA No 693949, and the Human Frontier Science Program (grant RGP0057/2018). B.B. was supported by long-term fellowships from the European Molecular Biology Organization and the Human Frontier Science Program. **Author contributions:** I.F.D., B.B., G.W., and J.-M.P. designed experiments; I.F.D. performed experiments, purified proteins, and analyzed data; B.B. generated and purified STAG1 and single-chain cohesin and optimized cohesin purification; D.G. generated NIPBL-MAU2 and performed initial ATPase assays; B.B. and W.T. generated the SCC1-Halo-Flag HeLa cell line; G.W. purified HeLa cohesin; I.F.D. and J.-M.P. wrote the manuscript with input from all authors; J.-M.P. supervised the study. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Materials will be provided upon request and completion of a material transfer agreement. Data described in this manuscript are archived at the Research Institute of Molecular Pathology and will be provided upon request.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1338/suppl/DC1
Materials and Methods
Figs. S1 to S6
Movies S1 to S10
References (*43*–*46*)

View/request a protocol for this paper from *Bio-protocol*.

2 September 2019; accepted 6 November 2019
Published online 21 November 2019
10.1126/science.aaz3418

REPORT

MOLECULAR BIOLOGY

# Human cohesin compacts DNA by loop extrusion

Yoori Kim[1]*, Zhubing Shi[1]*, Hongshan Zhang[2], Ilya J. Finkelstein[2]†, Hongtao Yu[1]†

Cohesin is a chromosome-bound, multisubunit adenosine triphosphatase complex. After loading onto chromosomes, it generates loops to regulate chromosome functions. It has been suggested that cohesin organizes the genome through loop extrusion, but direct evidence is lacking. Here, we used single-molecule imaging to show that the recombinant human cohesin-NIPBL complex compacts both naked and nucleosome-bound DNA by extruding DNA loops. DNA compaction by cohesin requires adenosine triphosphate (ATP) hydrolysis and is force sensitive. This compaction is processive over tens of kilobases at an average rate of 0.5 kilobases per second. Compaction of double-tethered DNA suggests that a cohesin dimer extrudes DNA loops bidirectionally. Our results establish cohesin-NIPBL as an ATP-driven molecular machine capable of loop extrusion.

The ring-shaped cohesin complex binds chromosomes both topologically and nontopologically and regulates diverse chromosome-based processes, including chromosome segregation, DNA repair, and transcription (*1*–*7*). Human cohesin consists of the SMC1-SMC3 heterodimeric adenosine triphosphatase (ATPase), the kleisin subunit RAD21 that links the ATPase heads, and either of the helical repeat proteins STAG1 or STAG2 (Fig. 1A). Cohesin is loaded on chromosomes by the NIPBL-MAU2 complex (*8*–*10*). Mutations of cohesin subunits and NIPBL result in human developmental diseases with multisystem dysfunctions, collectively referred to as cohesinopathy (*9*, *11*), likely because of transcriptional defects caused by cohesin deficiency. High-throughput chromosome conformation capture (Hi-C) experiments suggest that cohesin mediates the formation of chromosome loops and topologically associated domains (TADs) through a process called loop extrusion (*5*, *12*–*20*). Single-molecule studies have demonstrated that the related SMC complex condensin can extrude DNA loops (*21*, *22*). By contrast, cohesin has been reported to slide on DNA through adenosine triphosphate (ATP)–independent passive diffusion in vitro (*23*–*25*). Whether cohesin also has intrinsic loop extrusion activity remains an open question. Here, we addressed this question with single-molecule studies using recombinant human cohesin.

The cohesin loader NIPBL remains bound to cohesin on chromosomes (*26*, *27*) and is required for chromosome looping in cells (*15*).

[1]Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA. [2]Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: ilya@finkelsteinlab.org (I.J.F.); hongtao.yu@utsouthwestern.edu (H.Y.).

We thus expressed and purified the recombinant human cohesin complex alone or bound to the C-terminal region of NIPBL (NIPBL$^{C}$, residues 1163 to 2804) in insect cells (Fig. 1B and fig. S1A). Cohesin alone had low basal ATPase activity, and this activity was greatly stimulated by NIPBL$^{C}$ and DNA (Fig. 1C) (*28*, *29*). The ATPase-deficient cohesin SMC1A-E1157Q/SMC3-E1144Q (EQ) mutant exhibited minimal ATPase activity even in the presence of NIPBL$^{C}$ and DNA. Negative-stain electron microscopy showed that 51% ($n$ = 352) of the chemically cross-linked cohesin-NIPBL$^{C}$ complex particles displayed a bent-rod-like conformation with an overall length of ~50 nm (fig. S1, B and C), whereas the rest of the particles had a shorter, thicker rod shape with a length of ~33 nm. These conformations likely represent different forms of cohesin, with SMC1-SMC3 hinge domains partially or fully folded back toward their head domains, as had been previously observed for both human and yeast cohesin (*30*, *31*).

If cohesin can extrude DNA loops, then it would be expected to compact DNA. To test this possibility, we used total internal reflection fluorescence microscopy to observe aligned arrays of DNA molecules on a lipid bilayer surface in real time (Fig. 1D and movie S1) (*32*, *33*). DNA was stained with the green fluorescent dye YOYO-1. After incubating the DNA curtains with unlabeled cohesin-NIPBL$^{C}$ in the absence of buffer flow, we observed that all DNA molecules were completely compacted to the barrier in the presence of ATP (fig. S2), and this compaction could not be reversed by resuming flow.

We then fluorescently labeled NIPBL$^{C}$ with quantum dots (QDs) to visualize the cohesin-NIPBL$^{C}$ complex. The QD-labeled cohesin-NIPBL$^{C}$ complexes rapidly bound to the DNA array as soon as it entered the flow cell (Fig. 1E). We observed single QD-tagged complexes, as indicated by intermittent QD photoblinking (*34*), but could not completely rule out QD-tagged cohesin oligomers (see below). Analysis of the initial binding distribution of the cohesin-NIPBL$^{C}$ complexes on the DNA substrate showed preferential loading on AT-rich regions (fig. S3, A and B), which was also observed for cohesin alone and for condensin (*21*, *24*). We also measured the one-dimensional diffusion activity of the cohesin-NIPBL$^{C}$ complex on double-tethered DNA curtains (fig. S3, C and D). The complex was stably bound on DNA for >5 min and diffused slowly in an ATP-independent manner, with an average diffusion coefficient of 0.02 ± 0.003 $\mu m^2\ s^{-1}$ (mean ± SEM; fig. S3, E and F), which is eightfold lower than that reported for human cohesin topologically bound to DNA (*25*). Thus, the cohesin-NIPBL$^{C}$ complex displays very limited diffusion activity on DNA.

After the initial binding of cohesin-NIPBL$^{C}$ complexes to single-tethered DNA, we observed time-dependent, gradual DNA compaction (Fig. 1F and movie S2). We then measured the extent and rate of cohesin-mediated DNA compaction at different applied forces (i.e., flow rates; Fig. 1, F to H) (*35*). At an initial applied force of 0.3 pN, the DNA was almost completely compacted, with an average rate of 0.5 kb $s^{-1}$. When the initial applied force increased to 0.6 pN, we detected incomplete DNA condensation with a correspondingly slower compaction rate. There was minimal DNA compaction by cohesin-NIPBL$^{C}$ at 0.8 pN of applied force. Although the applied force changes for DNA molecules as they begin compacting, the force-sensitive cohesin translocation is qualitatively similar to condensin-mediated DNA looping (*22*). Reducing the salt concentration from 50 to 25 mM increased the extent and rate of DNA compaction by cohesin at high flow rates (fig. S4), suggesting that stronger cohesin-DNA or cohesin-cohesin interactions (as can occur in a crowded nuclear milieu) can raise the force threshold of compaction.

Cohesin alone (without NIPBL$^{C}$) with its RAD21 subunit labeled with QDs did not bind or compact DNA even in the presence of ATP (fig. S5A), indicating a requirement for NIPBL in loading cohesin onto DNA. The slowly hydrolyzable ATP analog adenylyl-imidodiphosphate (AMP-PNP) could not support efficient DNA compaction (fig. S5, A and B). The ATPase-deficient cohesin EQ mutant, which was expected to retain nucleotide-binding activity, was also deficient in DNA compaction even in the presence of NIPBL$^{C}$ and ATP. Thus, cohesin-NIPBL$^{C}$-dependent DNA compaction requires ATP hydrolysis.

We never detected stepwise DNA condensation events, suggesting that DNA compaction by cohesin-NIPBL$^{C}$ is not mediated through search and capture of distant DNA segments. Instead, we observed frequent cohesin-NIPBL$^{C}$ slippage events at 0.3 pN of applied force

**Fig. 1. Human cohesin-NIPBL$^C$ compacts linear DNA.** (**A**) Schematic representation of human cohesin-NIPBL. CC, coiled coil. (**B**) Coomassie staining of purified recombinant human cohesin and cohesin-NIPBL$^C$. WT, wild type. (**C**) ATPase activities (mean ± SEM) of human cohesin and cohesin-NIPBL$^C$ (50 nM) in the absence or presence of 500 nM 40-bp double-stranded DNA (dsDNA). (**D**) Illustration of DNA curtains bound by human cohesin-NIPBL$^C$. One end of DNA is tethered to the surface. (**E**) Image of fluorescently labeled cohesin-NIPBL$^C$ (magenta) on single-tethered DNA molecules stained with YOYO-1 (green). (**F**) Representative kymographs showing DNA condensation mediated by cohesin-NIPBL$^C$ at two applied forces. Red gradient triangle indicates the protein injection time window. The concentration of protein traversing the flow cell was diluted for a few minutes by constant buffer flow. (**G** and **H**) Quantification of the percentage of DNA length condensed (G) and the DNA condensation rate (H). Boxplots indicate the median, 10th, and 90th percentiles of the distribution. *p*-values are obtained from two-tailed *t* test: ****$p < 0.0001$. At least 25 DNA molecules were measured for each condition.

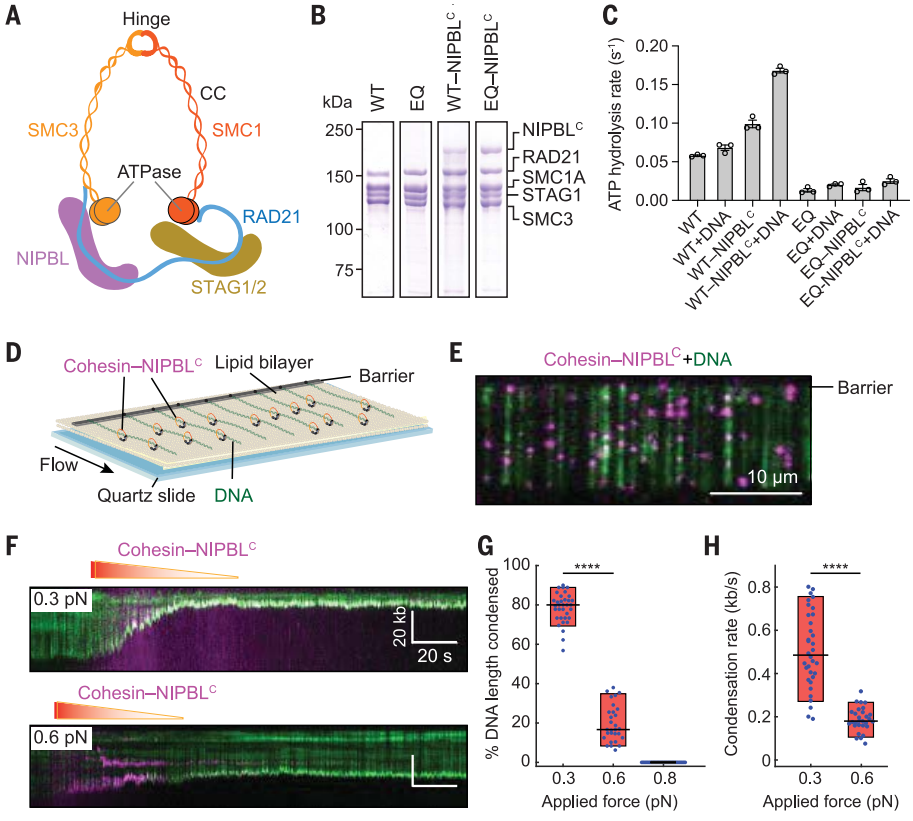
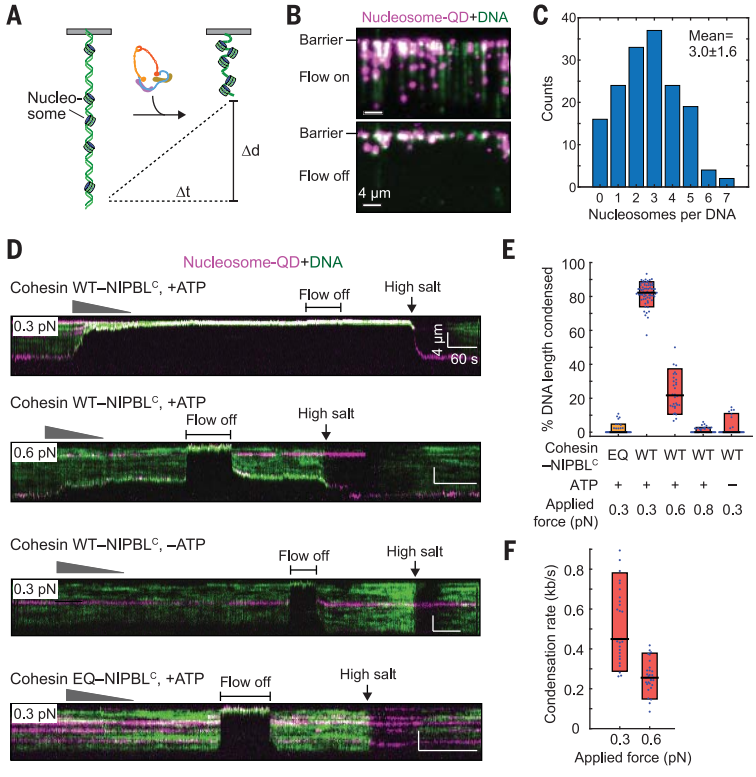



**Fig. 2. Cohesin-NIPBL$^C$ compacts nucleosomal DNA.** (**A**) Illustration of nucleosomal DNA compaction by cohesin-NIPBL$^C$. (**B**) Image of QD-labeled nucleosomes deposited on single-tethered DNA curtain with or without flow. (**C**) Distribution of the number of nucleosomes per DNA (mean ± SD). (**D**) Representative kymographs of the compaction of nucleosome-bound DNA by WT or EQ cohesin-NIPBL$^C$ at different applied forces with or without ATP. (**E**) Percentage of the length of nucleosome-bound DNA condensed in (D). (**F**) Compaction rate of nucleosome-bound DNA by cohesin-NIPBL$^C$. At least 25 DNA molecules were measured for each condition in (E) and (F).



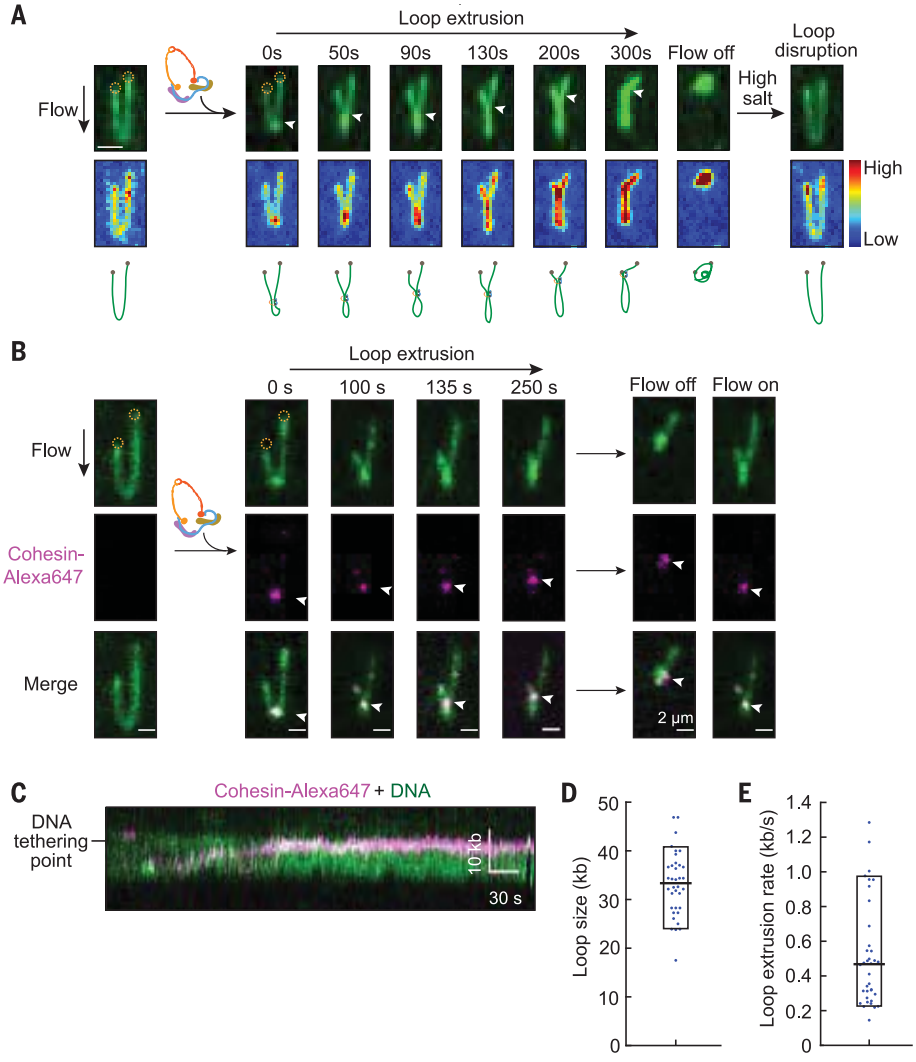



**Fig. 3. Real-time visualization of loop extrusion by cohesin-NIPBL$^{C}$.** (**A**) Time course showing DNA loop extrusion by cohesin-NIPBL$^{C}$ on YOYO-1–stained U-shaped DNA (top). Scale bar, 2 μm. Both DNA ends (dashed circle) are tethered to the surface and the extruding loop is extended at 0.1 ml/min buffer flow. Upon cohesin-NIPBL$^{C}$ injection, a small loop appeared at the tip of the DNA and elongated until the base of the loop (arrow) reached one tethering point. A brief shutoff of flow retracted DNA completely, indicating that DNA was not stuck to the surface. Injection of a high-salt buffer (500 mM NaCl) disrupted the loop. The scaled color map (middle panel) shows that the DNA intensity matches the growing loop. The schematic drawing (bottom panel) depicts a model of loop extrusion. (**B**) Time-course montage of loop extrusion showing the localization of labeled cohesin-NIPBL$^{C}$ (indicated by white arrowheads) at the base of the DNA loop. Turning the flow on and off showed that the cohesin-NIPBL$^{C}$ complex moved with the DNA, confirming that it was indeed bound to the DNA loop. (**C**) Representative kymograph showing the movement of a labeled cohesin-NIPBL$^{C}$ complex toward the DNA-tethering points. (**D** and **E**) Quantification of the loop size (D) and the rate of loop extrusion (E) by Alexa647-labeled cohesin-NIPBL$^{C}$.

(fig. S6). These observations suggested that cohesin can slide backward on DNA. The rate of the backtracking was similar to that of compaction (fig. S6C). The gradual, processive, ATP-dependent DNA compaction by cohesin-NIPBL$^{C}$ with occasional slippage indicated that cohesin is a bona fide molecular motor.

After DNA was fully compacted by cohesin-NIPBL$^{C}$ at 0.3 pN, increasing the applied force to 0.8 pN did not fully extend the DNA, and the bound cohesin only backtracked slightly (fig. S6A). This suggested that the completion of DNA compaction might lead to the formation of more stable cohesin-DNA assemblies, providing a plausible explanation for the observation that ATP is not required to maintain cohesin-dependent TADs after their formation in cells (*18*).

Cohesin can bind DNA by entrapment of DNA inside the lumen of its ring (topological binding) or by physical interaction with DNA that does not involve the opening of its ring (nontopological binding). Topological DNA binding is salt resistant (*29*). Injection of high-salt buffer dislodged the bound cohesin and fully reversed DNA compaction (fig. S6A and movie S3). Thus, cohesin-NIPBL$^{C}$ that is capable of loop extrusion might not be topologically bound to DNA. Analysis of the fluorescent intensities from DNA and QD-protein complexes indicated that cohesin-mediated DNA compaction preferably occurred at the untethered DNA ends (figs. S6A and S7). The underlying reason for this preference is unclear but could be due to the ease of formation of seed DNA loops near the ends or flow-induced higher occupancy of cohesin at DNA ends. Regardless, our data indicated that cohesin-NIPBL$^{C}$ is a processive DNA motor that compacts DNA.

DNA is packaged into chromatin in the nucleus. We next tested whether cohesin-NIPBL$^{C}$ could compact nucleosome-bound DNA (Fig. 2A). We incorporated one to six QD-labeled human nucleosomes on each DNA substrate using salt dialysis (Fig. 2, B and C) and visualized the nucleosome-bound DNA substrate upon injection of unlabeled cohesin-NIPBL$^{C}$ at multiple applied forces (Fig. 2D). The extent of nucleosome-DNA compaction was ~80% (Fig. 2E) and the average rate of the compaction was 0.5 kb $s^{-1}$ (Fig. 2F). Compaction of nucleosome-bound DNA was also force sensitive and dependent on ATP hydrolysis (Fig. 2, D to F). Thus, cohesin-NIPBL$^{C}$ compacts nucleosome-bound DNA with properties similar to those of naked DNA. Nucleosomes impede the movement of topologically loaded cohesin (*23*, *24*). Our finding that cohesin-NIPBL$^{C}$ compacts naked and nucleosome-bound DNA with similar rates again suggested that the loop-extruding cohesin might not be topologically loaded. Furthermore, high-salt washout of cohesin revealed that nucleosomes themselves were not repositioned during cohesin-dependent compaction (Fig. 2D). These data suggested that cohesin can act on chromatin without having to displace or slide nucleosomes.

To directly visualize loop extrusion by cohesin, we prepared U-shaped DNA by tethering both DNA ends to the surface and monitored the looping events in real time (movie S4). A small loop formed immediately after cohesin-NIPBL$^{C}$

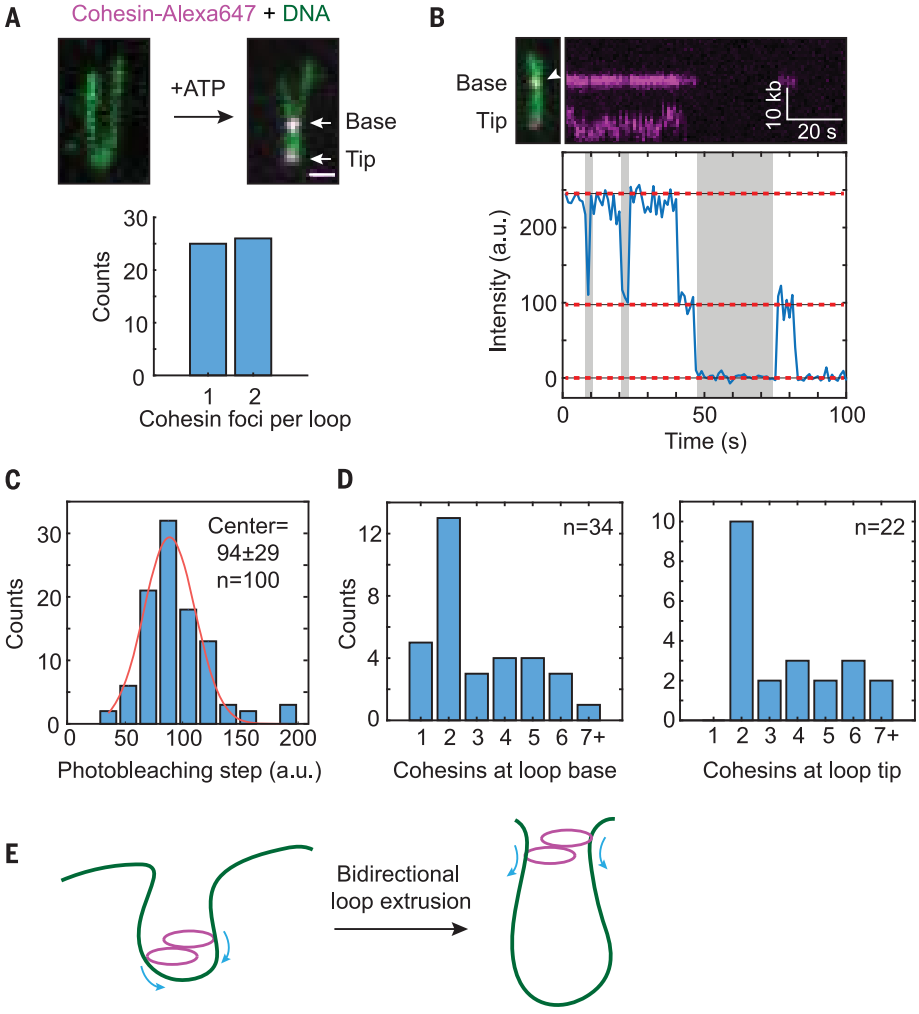



**Fig. 4. Cohesin-NIPBL$^{C}$ dimers promote loop extrusion.** (**A**) Top panel: Representative image of Alexa647-labeled cohesin-NIPBL$^{C}$ complexes bound to both the base and the tip of a DNA loop. Bottom panel: Number of Alexa647-cohesin foci on each DNA loop ($n$ = 51 DNA molecules). Scale bar, 2 μm. (**B**) Representative two-step photobleaching trace of cohesin at the loop base plotted with the corresponding DNA loop image and its kymograph shown above the trace. Dashed red lines are the photobleaching steps; gray shading is intermittent blinking indicating single fluors. a.u., arbitrary units. (**C**) Intensity distribution of single photobleaching steps of Alexa647-cohesin at the DNA loop fit to a Gaussian distribution (red line). (**D**) Distribution of the number of cohesin-NIPBL$^{C}$ molecules at the DNA loop base (left panel) and at the loop tip (right panel). (**E**) Model for bidirectional loop extrusion by a cohesin-NIPBL$^{C}$ dimer.

injection and gradually elongated until the motor stalled or one side of the loop reached either DNA-tethering point (Fig. 3A, fig. S8A, and movies S5 to S7). Both arms of the U-shaped DNA were shortened during the process. These results are consistent with cohesin-NIPBL$^{C}$ extruding DNA loops bidirectionally, as unidirectional asymmetric loop extrusion is expected to shorten only one arm of the U-shaped DNA. The loop was stably maintained for a few minutes, and injection of a high-salt buffer quickly restored the DNA to its original U shape (movie S8). As expected, we did not detect any looping activity in the absence of ATP or with the cohesin EQ mutant.

To visualize cohesin-NIPBL$^{C}$ at the base of the loop, we directly labeled cohesin-NIPBL$^{C}$ containing SNAP$_{f}$-tagged STAG1 with Alexa Fluor 647 dye (Alexa647). Alexa647-labeled cohesin completely compacted DNA with a similar rate as the unlabeled complex (fig. S8B). Cohesin-NIPBL$^{C}$ bound at the base of the extruding loop (Fig. 3B, fig. S8C, and movie S9). Consistent with symmetric, bidirectional loop extrusion, cohesin-NIPBL$^{C}$ moved toward DNA-tethering points during the process (Fig. 3C). The average size of extruded loops was 33 kb and the mean rate of loop extrusion was 0.5 kb s$^{-1}$ (Fig. 3, D and E).

A single condensin complex can extrude DNA loops asymmetrically and in one direction (*22*). A single cohesin complex might be able to perform symmetric loop extrusion. Alternatively, two cohesin complexes might act in concert to extrude loops in both directions. To determine how many cohesin molecules perform loop extrusion, we analyzed photobleaching steps of Alexa647-labeled cohesin-NIPBL$^{C}$ on DNA loops. On ~50% of DNA loops, we observed two fluorescent foci, with one each at the loop base and tip (Fig. 4A). The number of photobleaching steps of both Alexa647-cohesin foci peaked at two (Fig. 4, B to D). Cohesin foci on DNA that did not form loops had no discrete peaks for the number of photobleaching steps. Thus, the loop-extruding complexes most frequently contained two cohesin molecules. These data suggested that a cohesin-NIPBL$^{C}$ dimer might be the minimal functional unit for loop extrusion.

Collectively, our results support a model in which a cohesin-NIPBL$^{C}$ dimer extrudes DNA loops symmetrically in both directions (Fig. 4E). Bidirectional loop extrusion by a cohesin dimer also explains the observed and simulated Hi-C maps of chromosome loops (*5*, *19*). We observed loop extrusion by recombinant human cohesin-NIPBL$^{C}$ at relatively low applied forces and low salt concentrations. We anticipate that cellular crowding, MAU2, the N-terminal region of NIPBL, and other cohesin interactors further stabilize cohesin on DNA and enhance its intrinsic loop extrusion activity in cells.

High-salt buffer dislodged loop-extruding cohesin from DNA. Cohesin-NIPBL$^{C}$ bound to DNA at low salt exhibited diffusion kinetics much slower than those of topologically loaded cohesin. Finally, nucleosomes do not hinder cohesin-mediated DNA compaction. These findings suggest that cohesin mediates loop extrusion through nontopological or pseudo-topological interactions with DNA.

## REFERENCES AND NOTES

1. J. H. Haarhuis, A. M. Elbatsh, B. D. Rowland, *Dev. Cell* **31**, 7–18 (2014).
2. C. Morales, A. Losada, *Curr. Opin. Cell Biol.* **52**, 51–57 (2018).

3. F. Uhlmann, *Nat. Rev. Mol. Cell Biol.* **17**, 399–412 (2016).
4. G. Zheng, H. Yu, *Sci. China Life Sci.* **58**, 1089–1098 (2015).
5. Á. Sedeño Cacciatore, B. D. Rowland, *Curr. Opin. Genet. Dev.* **55**, 11–18 (2019).
6. J. M. Peters, A. Tedeschi, J. Schmitz, *Genes Dev.* **22**, 3089–3114 (2008).
7. K. Nasmyth, C. H. Haering, *Annu. Rev. Genet.* **43**, 525–558 (2009).
8. R. Ciosk *et al.*, *Mol. Cell* **5**, 243–254 (2000).
9. E. T. Tonkin, T. J. Wang, S. Lisgo, M. J. Bamshad, T. Strachan, *Nat. Genet.* **36**, 636–641 (2004).
10. E. Watrin *et al.*, *Curr. Biol.* **16**, 863–874 (2006).
11. T. Bose, J. L. Gerton, *J. Cell Biol.* **189**, 201–210 (2010).
12. K. Nasmyth, *Annu. Rev. Genet.* **35**, 673–745 (2001).
13. E. Alipour, J. F. Marko, *Nucleic Acids Res.* **40**, 11202–11212 (2012).
14. J. Gassler *et al.*, *EMBO J.* **36**, 3600–3618 (2017).
15. J. H. I. Haarhuis *et al.*, *Cell* **169**, 693–707.e14 (2017).
16. S. S. P. Rao *et al.*, *Cell* **171**, 305–320.e24 (2017).
17. G. Wutz *et al.*, *EMBO J.* **36**, 3573–3599 (2017).
18. L. Vian *et al.*, *Cell* **173**, 1165–1178.e20 (2018).
19. G. Fudenberg *et al.*, *Cell Rep.* **15**, 2038–2049 (2016).
20. A. Tedeschi *et al.*, *Nature* **501**, 564–568 (2013).
21. T. Terakawa *et al.*, *Science* **358**, 672–676 (2017).
22. M. Ganji *et al.*, *Science* **360**, 102–105 (2018).
23. I. F. Davidson *et al.*, *EMBO J.* **35**, 2671–2685 (2016).
24. J. Stigler, G. O. Çamdere, D. E. Koshland, E. C. Greene, *Cell Rep.* **15**, 988–998 (2016).
25. M. Kanke, E. Tahara, P. J. Huis In't Veld, T. Nishiyama, *EMBO J.* **35**, 2686–2698 (2016).
26. J. Rhodes, D. Mazza, K. Nasmyth, S. Uphoff, *eLife* **6**, e30000 (2017).
27. G. Zheng, M. Kanchwala, C. Xing, H. Yu, *eLife* **7**, e33920 (2018).
28. N. J. Petela *et al.*, *Mol. Cell* **70**, 1134–1148.e7 (2018).
29. Y. Murayama, F. Uhlmann, *Nature* **505**, 367–371 (2014).
30. M. T. Hons *et al.*, *Nat. Commun.* **7**, 12523 (2016).
31. F. Bürmann *et al.*, *Nat. Struct. Mol. Biol.* **26**, 227–236 (2019).
32. I. F. Gallardo *et al.*, *Langmuir* **31**, 10310–10317 (2015).
33. M. M. Soniat *et al.*, *Methods Enzymol.* **592**, 259–281 (2017).
34. M. Kuno, D. P. Fromm, H. F. Hamann, A. Gallagher, D. J. Nesbitt, *J. Chem. Phys.* **112**, 3117–3120 (2000).
35. J. M. Schaub, H. Zhang, M. M. Soniat, I. J. Finkelstein, *Langmuir* **34**, 14882–14890 (2018).


## ACKNOWLEDGMENTS

We thank Z. Ouyang and S. Kikuchi for their initial biochemical characterization of cohesin and its loader complex and the staff of the Electron Microscopy Core Facility at University of Texas Southwestern Medical Center for technical support. **Funding:** This study was supported by the Howard Hughes Medical Institute, the National Institutes of Health (GM120554 to I.J.F. and GM124096 to H.Y.), the Cancer Prevention and Research Institute of Texas (RP160667-P2 to H.Y.), and the Welch Foundation (F-1808 to I.J.F. and I-1441 to H.Y.). I.J.F. is a CPRIT Scholar in Cancer Research. **Competing interests:** The authors declare no competing interests. **Author contributions:** Y.K. designed and performed most single-molecule assays and analyzed the data. Z.S. performed protein purifications, ATPase assays, and negative-stain electron microscopy. H.Z. performed nucleosome experiments, the diffusion assay, and analyzed the data. H.Y. initiated and cosupervised the project. I.J.F. cosupervised the project and provided all instruments and materials needed for the single-molecule experiments. All authors contributed to the writing of the manuscript. **Data and materials availability:** All data are available in the manuscript or the supplementary material.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1345/suppl/DC1
Materials and Methods
Figs. S1 to S8
References (*36*–*43*)
Movies S1 to S9

View/request a protocol for this paper from *Bio-protocol*.

10 September 2019; resubmitted 11 November 2019
Accepted 16 November 2019
10.1126/science.aaz4475

HIGH-PRESSURE PHYSICS

# Imaging stress and magnetism at high pressures using a nanoscale quantum sensor

S. Hsieh[1,2]*, P. Bhattacharyya[1,2]*, C. Zu[1]*, T. Mittiga[1], T. J. Smart[3], F. Machado[1], B. Kobrin[1,2], T. O. Höhn[1,4], N. Z. Rui[1], M. Kamrani[5], S. Chatterjee[1], S. Choi[1], M. Zaletel[1], V. V. Struzhkin[6], J. E. Moore[1,2], V. I. Levitas[5,7,8], R. Jeanloz[3], N. Y. Yao[1,2]†

Pressure alters the physical, chemical, and electronic properties of matter. The diamond anvil cell enables tabletop experiments to investigate a diverse landscape of high-pressure phenomena. Here, we introduce and use a nanoscale sensing platform that integrates nitrogen-vacancy (NV) color centers directly into the culet of diamond anvils. We demonstrate the versatility of this platform by performing diffraction-limited imaging of both stress fields and magnetism as a function of pressure and temperature. We quantify all normal and shear stress components and demonstrate vector magnetic field imaging, enabling measurement of the pressure-driven $\alpha \leftrightarrow \epsilon$ phase transition in iron and the complex pressure-temperature phase diagram of gadolinium. A complementary NV-sensing modality using noise spectroscopy enables the characterization of phase transitions even in the absence of static magnetic signatures.

In hybrid quantum-sensing devices, sensors are directly integrated into existing toolsets ranging from biological imaging to materials spectroscopy (*1*–*4*). Here, we demonstrate the versatility of a platform based on quantum spin defects combined with static high-pressure technologies (*5*, *6*). In particular, we instrument diamond anvil cells (DACs) with a layer of nitrogen-vacancy (NV) centers directly at the culet, enabling the pursuit of two complementary objectives in high-pressure science: understanding the strength and failure of materials under pressure (e.g., the brittle-ductile transition) and discovering and characterizing exotic phases of matter (e.g., pressure-stabilized high-temperature superconductors) (*7*–*11*). Achieving these goals hinges upon the sensitive in situ imaging of signals within the high-pressure chamber. For the first goal, measuring the local stress environment permits the direct observation of inhomogeneities in plastic flow and the formation of line defects. For the second goal, the ability to spatially resolve field distributions can provide a direct image of complex order parameters and textured phenomena such as magnetic domains. However, the enormous stress gradients generated near the sample limit the utility of most conventional tabletop spectroscopy techniques; as a result, one is often restricted to measuring bulk properties averaged over the entire DAC geometry.

Our approach to these challenges is to use an ensemble of NV centers [~1 part per million (ppm) density] implanted ~50 nm from the surface of the diamond anvil culet (Fig. 1, A and B). Each NV center represents an atomic-scale defect (i.e., a substitutional nitrogen impurity adjacent to a vacancy) inside the diamond lattice and exhibits an $S = 1$ electronic spin ground state (*12*). In the absence of external fields, the $|m_s = \pm 1\rangle$ spin sublevels are degenerate and separated by $D_{gs} = (2\pi) \times 2.87$ GHz from the $|m_s = 0\rangle$ state. Crucially, both the nature and energy of these spin states are sensitive to local changes in stress, temperature, and magnetic and electric fields (Fig. 1C) (*13*–*19*). These spin states can be optically initialized and read out, as well as coherently manipulated through microwave fields. Their energy levels can be probed by performing optically detected magnetic resonance (ODMR) spectroscopy, which measures a change in the NV's fluorescence intensity when an applied microwave field is on resonance between two NV spin sublevels (Fig. 1D), thus enabling a variety of external signals to be sensed over a wide range of environmental conditions (*1*, *20*, *21*).

Here, we focus on the sensing of stress and magnetic fields, wherein the NV is governed by the Hamiltonian (*18*, *22*), $H = H_0 + H_B + H_S$, with $H_0 = D_{gs} S_z^2$ (zero-field splitting), $H_B = \gamma_B \vec{B} \cdot \vec{S}$ (Zeeman splitting), and $H_S = [\alpha_1(\sigma_{xx} + \sigma_{yy}) + \beta_1 \sigma_{zz}] S_z^2 + [\alpha_2(\sigma_{yy} - \sigma_{xx}) + \beta_2(2\sigma_{xz})](S_y^2 - S_x^2) + [\alpha_2(2\sigma_{xy}) + \beta_2(2\sigma_{yz})](S_x S_y + S_y S_x)$ capturing the NV's response to the local diamond stress tensor, $\overleftrightarrow{\sigma}$ (Fig. 1C). In the above, $\gamma_B \approx (2\pi) \times 2.8$ MHz/G is the gyromagnetic ratio, $\{\alpha_{1,2}, \beta_{1,2}\}$ are the stress susceptibility coefficients (*17*–*19*, *23*), $\hat{z}$ is the NV orientation axis, and $\hat{x}$ is defined such that the $xz$ plane contains one of the carbon-vacancy bonds (Fig. 1E). In general, the resulting ODMR spectra exhibit eight resonances arising from the four possible crystallographic orientations of the NV (Fig. 1D). By extracting the energy shifting and splitting of the spin sublevels for each NV orientation group, one obtains an overconstrained set of equations enabling the reconstruction of either the (six component) local stress tensor or the (three component) vector magnetic field (*23*).

In our experiments, we use a miniature DAC (Fig. 1, A and B) consisting of two opposing anvils compressing either a beryllium copper or rhenium gasket (*24*). The sample chamber defined by the gasket and diamond-anvil culets is filled with a pressure-transmitting medium (either a 16:3:1 methanol/ethanol/water solution or cesium iodide) to provide a quasi-hydrostatic environment. Microwave excitation is applied with a 4-μm-thick platinum foil compressed between the gasket and anvil pavilion facets (fig. S1); scanning confocal microscopy (with a transverse diffraction-limited spot size of ~600 nm, containing ~$10^3$ NVs) allows us to obtain two-dimensional ODMR maps across the culet.

We begin by probing the stress tensor across the culet surface (up to $P = 48$ GPa as shown in fig. S7) using two different cuts of diamond [i.e., (111)-cut and (110)-cut culet]. For a generic stress environment, the intrinsic degeneracy associated with the four NV orientations is not sufficiently lifted, implying that individual resonances cannot be resolved. To resolve these resonances while preserving the stress contribution, we sequentially tune a precisely controlled external magnetic field to be perpendicular to each of the different NV orientations (*23*). For each perpendicular field choice, three of the four NV orientations exhibit a strong Zeeman splitting proportional to the projection of the external magnetic field along their symmetry axes. Notably, this enables one to resolve the stress information encoded in the remaining NV orientation, whereas the other three groups of NVs are spectroscopically split away. Using this method, we obtain sufficient information to extract the full stress tensor, as depicted in Fig. 2. A number of intriguing features are observed at the interface between the culet and the sample chamber, which provide insight into both elastic (reversible) and plastic (irreversible) deformations.

At low pressures ($P = 4.9$ GPa), the normal stress along the loading axis, $\sigma_{ZZ}$, is spatially uniform (Fig. 2A), whereas all shear stresses, $\{\sigma_{XY}, \sigma_{XZ}, \sigma_{YZ}\}$, are minimal (Fig. 2B). The axes $\{\hat{X}, \hat{Y}, \hat{Z}\}$ correspond to the lab frame, whereas $\{\hat{x}, \hat{y}, \hat{z}\}$ correspond to the NV frame (Fig. 1, A and E). These observations are in agreement with conventional stress continuity predictions for the interface between a solid and an ideal fluid (*25*). Moreover, $\sigma_{ZZ}$ is consistent with the independently measured pressure inside the sample chamber (by ruby fluorescence),

[1]Department of Physics, University of California, Berkeley, CA 94720, USA. [2]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Department of Earth and Planetary Science, University of California, Berkeley, CA 94720, USA. [4]Fakultät für Physik, Ludwig-Maximilians-Universität München, 80799 Munich, Germany. [5]Department of Aerospace Engineering, Iowa State University, Ames, IA 50011, USA. [6]Geophysical Laboratory, Carnegie Institution of Washington, Washington, DC 20015, USA. [7]Department of Mechanical Engineering, Iowa State University, Ames, IA 50011, USA. [8]Ames Laboratory, Division of Materials Science and Engineering, Ames, IA 50011, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: norman.yao@berkeley.edu

demonstrating the NV's potential as a built-in pressure scale (*26*). In contrast to the uniformity of $\sigma_{ZZ}$, the field profile for the mean lateral stress, $\sigma_{\perp} \equiv \frac{1}{2}(\sigma_{XX} + \sigma_{YY})$, exhibits a concentration of forces toward the center of the culet (Fig. 2A). Using the measured $\sigma_{ZZ}$ as a boundary condition, we perform finite-element simulations to reproduce this spatial pattern (*23*).

Upon increasing pressure ($P = 13.6$ GPa), a spatial gradient in $\sigma_{ZZ}$ emerges (Fig. 2B, inset). This qualitatively distinct feature is consistent with the solidification of the pressure-transmitting medium into its glassy phase above $P_g \approx 10.5$ GPa (*27*). Crucially, this demonstrates our ability to characterize the effective viscosity of solids and liquids under pressure. To characterize the sensitivity of our system, we perform ODMR spectroscopy with a static applied magnetic field and pressure under varying integration times and extract the frequency uncertainty from a Gaussian fit. We observe a stress sensitivity of $\{0.023, 0.030, 0.027\}$ GPa/$\sqrt{\text{Hz}}$ for hydrostatic, average normal, and average shear stresses, respectively. This is consistent with the theoretically derived stress sensitivity, $\eta_S \sim \frac{\Delta\nu}{\xi C\sqrt{Nt}} = \{0.017, 0.022, 0.020\}$ GPa/$\sqrt{\text{Hz}}$, where $N$ is the number of NV centers, $\Delta\nu$ is the linewidth, $\xi$ is the relevant stress susceptibility, $t$ is the integration time, and $C$ is an overall factor accounting for measurement infidelity (*23*). In combination with diffraction-limited imaging resolution, this sensitivity makes it possible to measure and ultimately control the full stress tensor distribution across a sample.

Having characterized the stress environment, we use the NV centers as an in situ magnetometer to detect phase transitions inside the high-pressure chamber. Analogous to the case of stress, we observe a magnetic sensitivity of 12 μT/$\sqrt{\text{Hz}}$, in agreement with the theoretically estimated value, $\eta_B \sim \frac{\delta\nu}{C\gamma_B B\sqrt{Nt}} = 8.8$ μT/$\sqrt{\text{Hz}}$. Assuming a point dipole located a distance $d \sim 5$ μm from the NV layer, this corresponds to an experimentally measured magnetic moment sensitivity: $7.5 \times 10^{-12}$ emu/$\sqrt{\text{Hz}}$ (Fig. 1F).

After determining the sensitivity, we begin by directly measuring the magnetization of iron as it undergoes the pressure-driven $\alpha \leftrightarrow \epsilon$ phase transition from body-centered cubic (bcc) to hexagonal close-packed (hcp) crystal structures (*28*); crucially, this structural phase transition is accompanied by the depletion of the magnetic moment, and it is this change in the iron's magnetic behavior that we image. The sample chamber is loaded with a ~10-μm polycrystalline iron pellet as well as a ruby microsphere (pressure scale), and we apply an external magnetic field $B_{ext}$~180 G. As before, by performing a confocal scan across the culet, we acquire a two-dimensional magnetic resonance map (Fig. 3). At low pressures (Fig. 3A), near the iron pellet, we observe substantial shifts in the eight NV resonances, owing to the presence of a ferromagnetic field from the iron pellet. As one increases pressure (Fig. 3B), these shifts begin to diminish, signaling a reduction in the magnetic susceptibility. Finally, at the highest pressures ($P \sim$ 22 GPa, Fig. 3C), the magnetic field from the

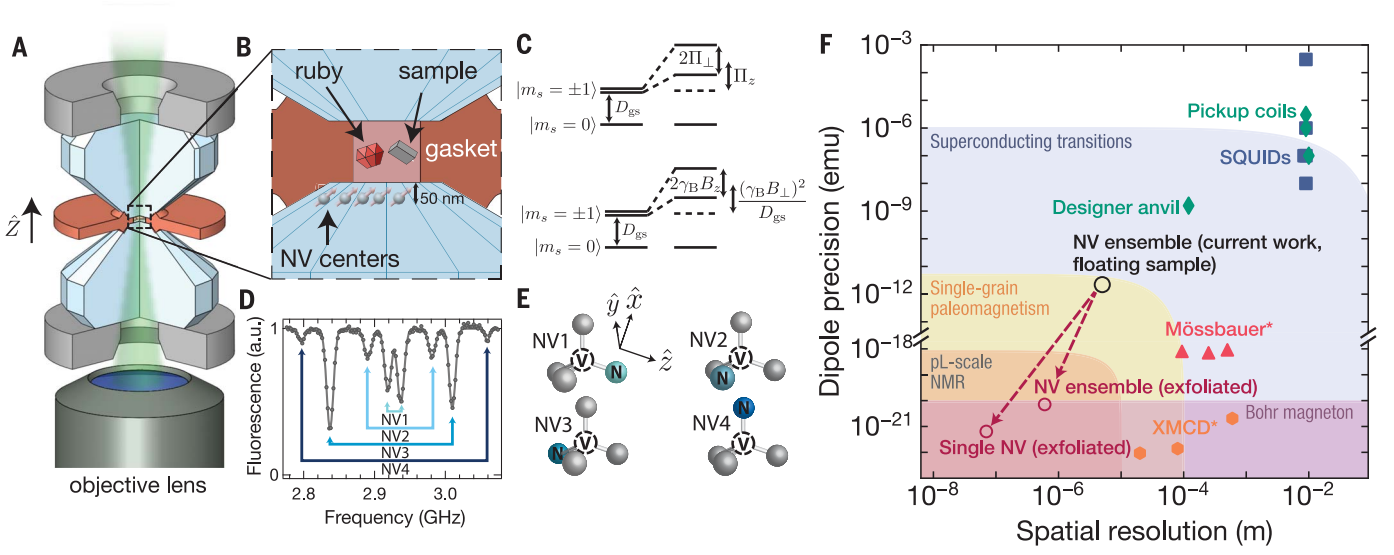



**Fig. 1. NV centers integrated into a diamond anvil cell.** (**A**) Schematic of the DAC geometry. Two opposing anvils are compressed by a nonmagnetic steel cell and cubic boron nitride backing plates (gray). NV centers are initialized and read out using a 532-nm laser focused to a diffraction-limited spot (~600 nm), which is scanned across the culet surface. (**B**) The DAC sample chamber is defined by the gasket-anvil assembly (diagram not to scale); it is loaded with the sample of interest, a pressure-transmitting medium, and a single ruby microsphere (pressure calibration). A ~50-nm layer of NV centers is embedded into the diamond anvil directly below the sample chamber. (**C**) Top: Stress both shifts and splits the $m_s = \pm 1$ sublevels at first order; in particular, the shifting is characterized by $\Pi_z = \alpha_1(\sigma_{xx} + \sigma_{yy}) + \beta_1\sigma_{zz}$, and the splitting is characterized by $\Pi_\perp^2 = [\alpha_2(\sigma_{yy} - \sigma_{xx}) + \beta_2(2\sigma_{xz})]^2 + [\alpha_2(2\sigma_{xy}) + \beta_2(2\sigma_{yz})]^2$. Bottom: An axial magnetic field splits the $m_s = \pm 1$ sublevels at first order, but a transverse magnetic field leads to shifts only at second order. (**D**) A representative ODMR spectrum from an NV center ensemble under an applied magnetic field. (**E**) Each pair of resonances in (D) corresponds to one of the four NV crystallographic orientations. (**F**) Comparison of high-pressure magnetometry techniques. We define the spatial resolution as a characteristic sensor length scale over which the sample magnetism is integrated. Estimates for our current work are shown assuming a sample suspended in a pressure medium 5 μm away from the culet (black open circle). We project that by exfoliating a sample directly onto the culet surface and using 5-nm implanted NV centers, the distance from the sample can be substantially reduced, thus improving both dipole precision and spatial resolution (open red circles). Inductive methods [pickup coils (green diamonds) and superconducting quantum interference devices (SQUIDs) (blue squares)] integrate the magnetization of a sample over the coil's area (*23*); to this end, the diameter associated with the coil is taken as the "spatial resolution" although in principle, the sample inside the chamber can be substantially smaller. By contrast, high-energy photon scattering techniques [x-ray magnetic circular dichroism (orange hexagons), and Mössbauer spectroscopy (pink triangles)] probe atomic-scale magnetism (*23*); the length scale for these methods is shown here as the spot size of the excitation beam.

pellet has decreased by more than two orders of magnitude.

To quantify this phase transition, we reconstruct the full vector magnetic field produced by the iron sample from the aforementioned two-dimensional NV magnetic resonance maps (Fig. 3, D to F). We then compare this information with the expected field distribution at the NV layer inside the culet, assuming the iron pellet generates a dipole field (*23*). This enables us to extract an effective dipole moment as a function of applied pressure (Fig. 3G). To identify the critical pressure, we fit the transition using a logistic function (*23*). This procedure yields the transition at $P = 16.7 \pm 0.7$ GPa (Fig. 3J).

In addition to changes in the magnetic behavior, another key signature of this first-order transition is the presence of hysteresis. We investigate this by slowly decompressing the diamond anvil cell and monitoring the dipole moment; the decompression transition occurs at $P = 10.5 \pm 0.7$ GPa (Fig. 3J), suggesting a hysteresis width of ~6 GPa, consistent with a combination of intrinsic hysteresis and finite shear stresses in the methanol/ethanol/water pressure-transmitting medium (*28*). Taking the average of the forward and backward hysteresis pressures, we find a critical pressure of $P_c = 13.6 \pm 3.6$ GPa, in excellent agreement with independent measurements by Mössbauer spectroscopy, where $P_c \approx 12$ GPa (Fig. 3J) (*28*).

Next, we demonstrate the integration of our platform into a cryogenic system, enabling us to make spatially resolved in situ measurements across the pressure-temperature (*P-T*) phase diagram of materials. Specifically, we investigate the magnetic *P-T* phase diagram of the rare-earth element gadolinium (Gd) up to pressures $P \approx 8$ GPa and between temperatures $T = 25$ to 340 K. Owing to an interplay between localized 4f electrons and mobile conduction electrons, Gd represents an interesting playground for studying metallic magnetism; in particular, the itinerant electrons mediate RKKY-type interactions between the local moments, which in turn induce spin-polarization of the itinerant electrons (*29*). Moreover, much like its rare-earth cousins, Gd exhibits a series of pressure-driven structural phase transitions from hcp to samarium-type (Sm-type) to double hcp (dhcp) (Fig. 4) (*30*). The interplay between these different structural phases, various types of magnetic ordering, and metastable transition dynamics leads to a complex magnetic *P-T* phase diagram that remains the object of study to this day (*29*–*31*).

In analogy to our measurements of iron, we monitor the magnetic ordering of a Gd flake by using the NV's ODMR spectra at two different locations inside the culet: close to and far away from the sample (the latter to be used as a control) (fig. S15). Because of thermal contraction of the DAC (which induces a change in pressure), each experimental run traces a distinct non-isobaric path through the *P-T* phase diagram (Fig. 4C, blue curves). In addition to these DC magnetometry measurements, we also operate the NV sensors in a complementary mode, i.e., as a noise spectrometer.

We begin by characterizing Gd's well-known ferromagnetic Curie transition at ambient pressure, which induces a sharp jump in the splitting of the NV resonances at $T_C = 292.2 \pm 0.1$ K (Fig. 4D). As depicted in Fig. 4A, upon increasing pressure, this transition shifts to lower temperatures, and consonant with its second-order nature (*32*), we observe no hysteresis (Fig. 4A, inset); this motivates us to fit the data and extract $T_C$ by solving a regularized Landau free-energy equation (*23*). Combining all of the low-pressure data (Fig. 4C, red squares), we find a linear decrease in the Curie temperature at a rate $dT_C/dP = -18.7 \pm 0.2$ K/GPa, consistent with prior studies using both DC conductivity and AC-magnetic susceptibility (*30*). Unexpectedly, this linear decrease extends well into the Sm-type phase. Upon increasing pressure above ~6 GPa (path [b] in Fig. 4C), we observe the loss of ferromagnetic (FM) signal (Fig. 4B), indicating a first-order structural transition into the paramagnetic (PM) dhcp phase (*30*). In stark contrast to the previous Curie transition, there is no revival of a ferromagnetic signal even after heating up (~315 K) and substantially reducing the pressure (to < 0.1 GPa).

A few remarks are in order. The linear decrease of $T_C$ well beyond the ~2-GPa structural transition between hcp- and Sm-type is consistent with the "sluggish" equilibration

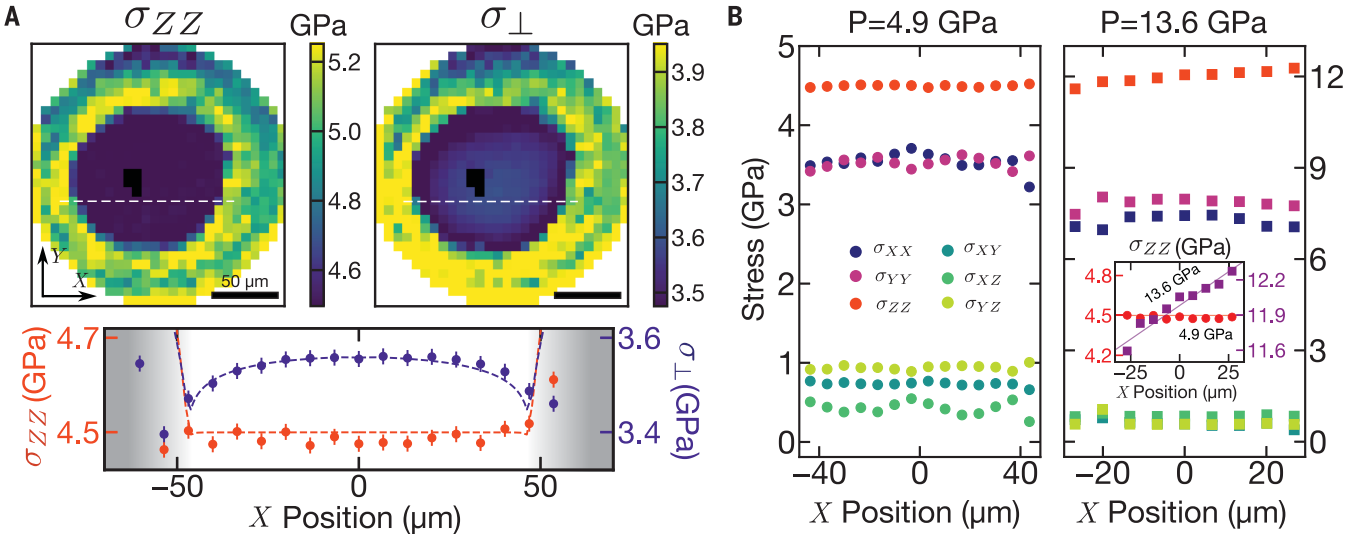



**Fig. 2. Full tensorial reconstruction of the stresses in a (111)-cut diamond anvil.** (**A**) Spatially resolved maps of the loading stress (left) and mean lateral stress (right), $\sigma_\perp = \frac{1}{2}(\sigma_{XX} + \sigma_{YY})$, across the culet surface. In the inner region, where the culet surface contacts the pressure-transmitting medium (16:3:1 methanol/ethanol/water), the loading stress is spatially uniform, whereas the lateral stress is concentrated toward the center; this qualitative difference is highlighted by a linecut (taken along the white-dashed line) of the two stresses (below), and reconstructed by finite-element analysis (orange and purple dashed lines). The black pixels indicate where the NV spectrum was obfuscated by the ruby microsphere. (**B**) Comparison of all stress tensor components in the fluid-contact region at $P = 4.9$ GPa and $P = 13.6$ GPa. At $P = 13.6$ GPa, the pressure-transmitting medium has entered its glassy phase, and we observe a spatial gradient in the loading stress $\sigma_{ZZ}$ (inset).

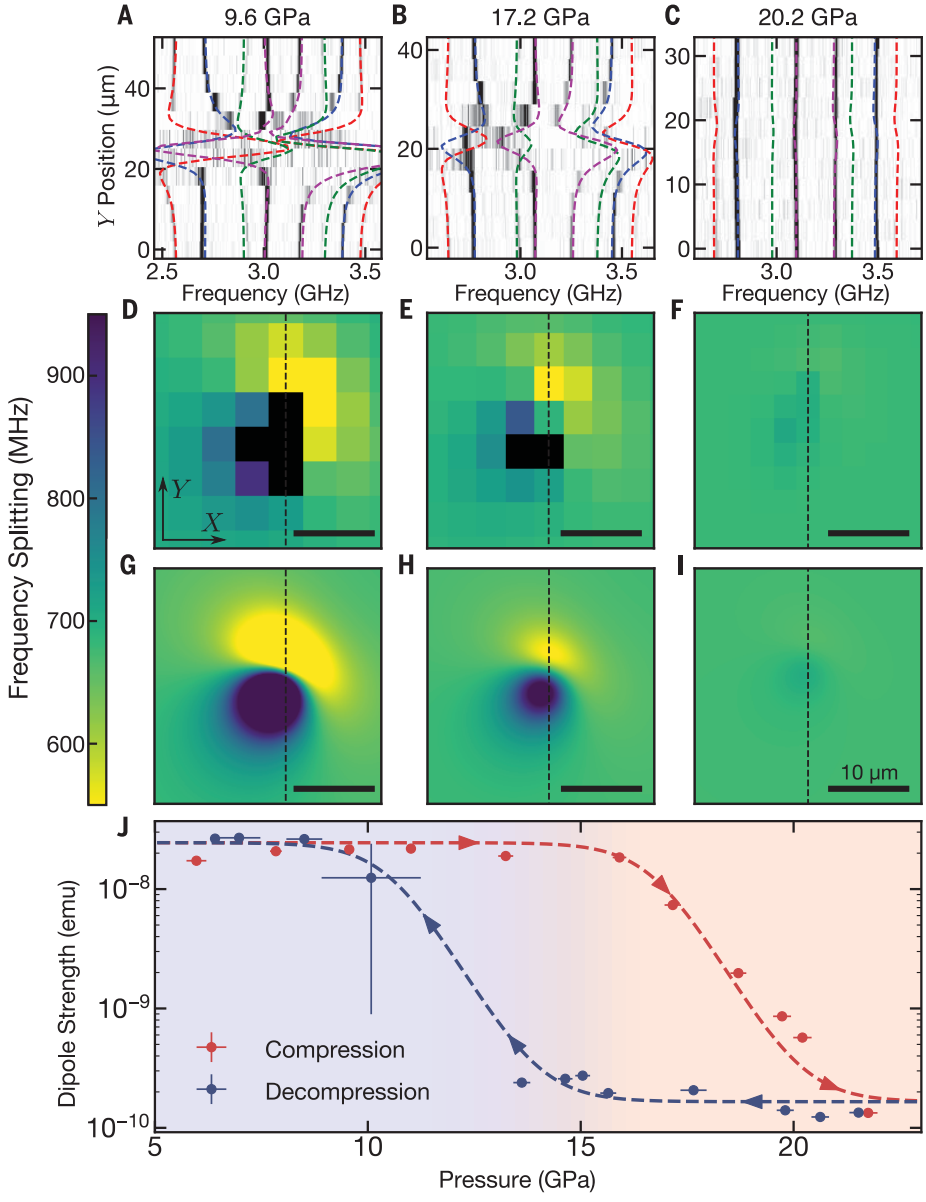



**Fig. 3. Imaging iron's $\alpha \leftrightarrow \epsilon$ phase transition.** Applying an external magnetic field ($B_{ext} \sim 180$ G) induces a dipole moment in the polycrystalline iron pellet that generates a spatially varying magnetic field across the culet of the diamond anvil. By mapping the ODMR spectra across the culet surface, we reconstruct the local magnetic field that characterizes the iron pellet's magnetization. (**A** to **C**) Comparison between the measured ODMR spectra (dark regions correspond to resonances) and the theoretical resonance positions (different colors correspond to different NV crystallographic orientations) across vertical spatial cuts (i.e., *Y* position indicates location along the black-dashed line shown in the two-dimensional scans below) at pressures of 9.6, 17.2, and 20.2 GPa, respectively (16:3:1 methanol/ethanol/water solution). (**D** to **F**) Map of the measured energy difference of a particular NV crystallographic orientation [blue lines in (A) to (C)]. Black pixels correspond to ODMR spectra where the splitting could not be accurately extracted owing to large magnetic field gradients (fig. S12). (**G** to **I**) Theoretical reconstruction of the energy differences shown in (D) to (F) (*23*). Data depicted in (A) to (C) are taken along the thin black dashed lines. (**J**) Measured dipole moment of the iron pellet as a function of applied pressure at room temperature, for both compression (red) and decompression (blue). Based on the hysteresis observed (~6 GPa), we find the critical pressure $P_c = 13.6 \pm 3.6$ GPa, in excellent agreement with previous studies (*28*).

between these two phases at low temperatures (*30*). The metastable dynamics of this transition are strongly pressure and temperature dependent, suggesting that different starting points (in the *P-T* phase diagram) can exhibit markedly different behaviors (*30*). To highlight this, we probe two different transitions out of the paramagnetic Sm-type phase by tailoring specific paths in the *P-T* phase diagram. By taking a shallow path in *P-T* space, we observe a small change in the local magnetic field across the structural transition into the PM dhcp phase at ~6 GPa (Fig. 4C, path [c], orange diamonds). By taking a steeper path in *P-T* space, one can also investigate the magnetic transition into the antiferromagnetic (AFM) Sm-type phase at ~150 K (Fig. 4C, path [d], green triangle). In general, these two transitions are extremely challenging to probe via DC magnetometry because their signals arise only from small differences in the susceptibilities between the various phases (fig. S18).

To this end, we demonstrate a complementary NV sensing modality based on noise spectroscopy, which can probe phase transitions even in the absence of a direct magnetic signal (*33*). Specifically, returning to Gd's ferromagnetic Curie transition, we monitor the NV's depolarization time, $T_1$, as the phase transition is crossed (Fig. 4D). Normally, the NV's $T_1$ time is limited by spin-phonon interactions and increases sharply as the temperature is decreased. Here, we observe a markedly disparate behavior. In particular, using nanodiamonds drop-cast on a Gd foil at ambient pressure, we find that the NV $T_1$ is nearly temperature independent in the paramagnetic phase, before exhibiting a kink and subsequent decrease upon entering the ferromagnetic phase (Fig. 4D). We note two intriguing observations: first, one possible microscopic explanation for this behavior is that $T_1$ is dominated by Johnson-Nyquist noise from the thermal fluctuations of charge carriers inside Gd (*34*, *35*). Gapless critical spin fluctuations or magnons in the ordered phase, although expected, are less likely to cause this signal (*23*). Second, we observe that the Curie temperature, as identified by $T_1$-noise spectroscopy, is ~10 K higher than that observed via DC magnetometry (Fig. 4D). Similar behavior has previously been reported for the surface of Gd (*29*, *36*), suggesting that our noise spectroscopy could be more sensitive to surface physics.

Further stress characterization of other fluids and solids may provide insights into

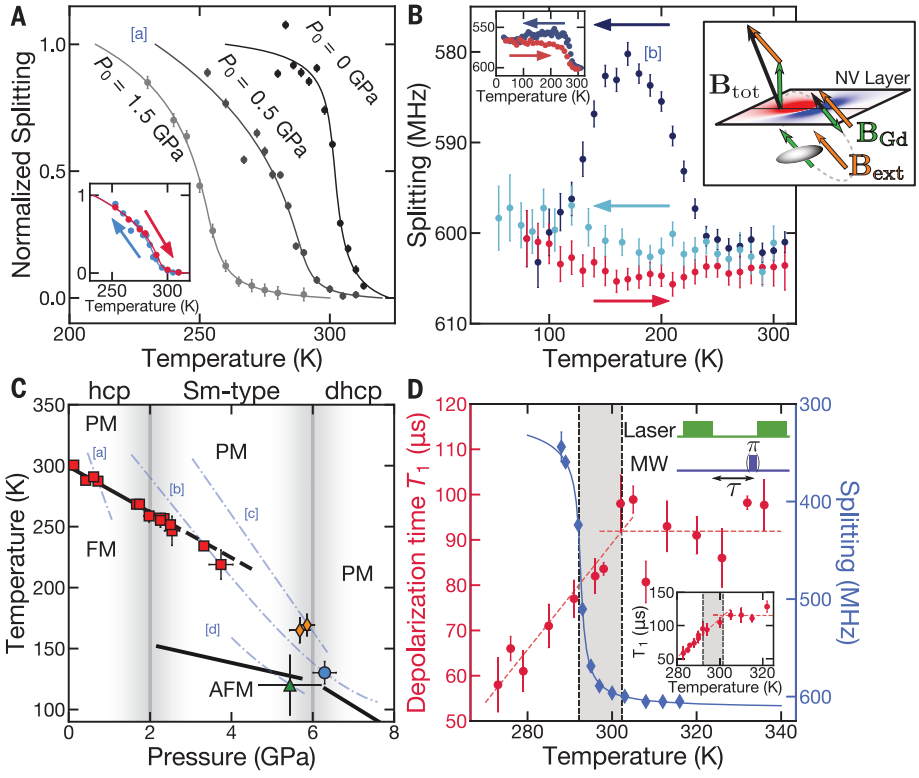



**Fig. 4. Magnetic P-T phase diagram of gadolinium.** A ~ 30 μm by 30 μm by 25 μm polycrystalline Gd foil is loaded into a beryllium copper gasket with a cesium iodide pressure medium. An external magnetic field, $B_{ext}$ ~ 120 G, induces a dipole field, $B_{Gd}$, detected by the splitting of the NVs [right inset, (B)]. (**A**) The FM Curie temperature $T_C$ decreases with increasing pressure up to ~4 GPa. NV splittings for three *P-T* paths, labeled by their initial pressure $P_0$, are shown. The *P-T* path for run [a] ($P_0$ = 0.5 GPa) is shown in (C). The cool-down (blue) and heat-up (red) of a single *P-T* cycle shows negligible hysteresis (inset). (**B**) If a *P-T* path starting in hcp is taken into the dhcp phase (at pressures ≳6 GPa) (*30*), the FM signal is lost and not reversible, as shown in (C) (path [b]). Upon cool-down (dark blue), we observe the aforementioned Curie transition, followed by the loss of FM signal at 6.3 GPa, 130 K. But upon heat-up (red) and second cool-down (light blue), the FM signal is not recovered. When the pressure does not go beyond ~6 GPa, the FM signal is recoverable (left inset) (*23*). (**C**) Magnetic *P-T* phase diagram of Gd. At low pressures, we observe the linear decrease of $T_C$ (black line) with slope −18.7 ± 0.2 K/GPa, in agreement with previous measurements (*30*). This linear regime extends into the Sm-type phase (black dashed line) owing to the slow dynamics of the hcp → Sm-type transition (*30*). When starting in the Sm-type phase, we no longer observe a FM signal, but rather a small change in the magnetic field at either the transition from Sm-type to dhcp (orange diamonds) or from PM to AFM (green triangle), depending on the *P-T* path. The bottom two phase boundaries (black lines) are taken from (*31*). (**D**) At ambient pressure, we observe a Curie temperature, $T_C$ = 292.2 ± 0.1 K, by using DC magnetometry (blue data). Using nanodiamonds drop-cast onto a Gd foil (and no applied external magnetic field), we find that the depolarization time ($T_1$) of the NVs is qualitatively different in the two phases (red data). $T_1$ is measured using the pulse sequence shown in the top right inset. The $T_1$ measurement on another nanodiamond exhibits nearly identical behavior (bottom inset).

mechanical phenomena such as viscous flow, plastic deformation, and pressure-dependent yield strength. Such information is challenging to obtain by either numerical finite-element simulations or more conventional experimental methods and may ultimately allow control of the deviatoric- as well as normal-stress conditions in high-pressure experiments (*37*).

The high sensitivity and close proximity of our sensor enables the measurement of signals in settings that are beyond the capabilities of existing techniques (Fig. 1F). Such settings include, for example, nuclear magnetic resonance (NMR) at picoliter volumes (*38*) and single-grain remnant magnetism (*39*), as well as phenomena that exhibit spatial textures such as magnetic skyrmions (*4*) and superconducting vortices (*40*).

Although our work uses NV centers, the techniques developed here can be readily extended to other atomic defects. For instance, recent developments on all-optical control of silicon-vacancy centers in diamond may allow for microwave-free stress imaging with improved sensitivities (*41*). In addition, one can consider defects in other anvil substrates beyond diamond; indeed, recent studies have shown that moissanite (6H silicon carbide) hosts optically active defects that show promise as local sensors (*41*). In contrast to millimeter-scale diamond anvils, moissanite anvils can be manufactured at centimeter or larger scales, and therefore support larger sample volumes that ameliorate the technical requirements of many experiments. Finally, the suite of sensing capabilities previously demonstrated for NV centers (i.e., electric, thermal, gyroscopic precession, etc.) can now straightforwardly be extended to high-pressure environments, opening up a large range of experiments for quantitatively characterizing materials at such extreme conditions.

## REFERENCES AND NOTES


1. G. Kucsko *et al.*, *Nature* **500**, 54–58 (2013).
2. P. Maletinsky *et al.*, *Nat. Nanotechnol.* **7**, 320–324 (2012).
3. J. Cai, F. Jelezko, M. B. Plenio, *Nat. Commun.* **5**, 4065 (2014).
4. Y. Dovzhenko *et al.*, *Nat. Commun.* **9**, 2712 (2018).
5. A. Jayaraman, *Rev. Mod. Phys.* **55**, 65–108 (1983).
6. H. Mao, X.-J. Chen, Y. Ding, B. Li, L. Wang, *Rev. Mod. Phys.* **90**, 015007 (2018).
7. E. Wigner, H. Huntington, *J. Chem. Phys.* **3**, 764–770 (1935).
8. H. Horii, S. Nemat-Nasser, *Philos. Trans. A Math. Phys. Eng. Sci.* **319**, 337–374 (1986).
9. E. Gilioli, L. Ehm, *IUCrJ* **1**, 590–603 (2014).
10. A. P. Drozdov, M. I. Eremets, I. A. Troyan, V. Ksenofontov, S. I. Shylin, *Nature* **525**, 73–76 (2015).
11. M. Somayazulu *et al.*, *Phys. Rev. Lett.* **122**, 027001 (2019).
12. M. W. Doherty *et al.*, *Phys. Rep.* **528**, 1–45 (2013).
13. V. M. Acosta *et al.*, *Phys. Rev. Lett.* **104**, 070801 (2010).
14. J. R. Maze *et al.*, *Nature* **455**, 644–647 (2008).
15. F. Dolde *et al.*, *Nat. Phys.* **7**, 459–463 (2011).
16. P. Ovartchaiyapong, K. W. Lee, B. A. Myers, A. C. B. Jayich, *Nat. Commun.* **5**, 4429 (2014).
17. M. W. Doherty *et al.*, *Phys. Rev. Lett.* **112**, 047601 (2014).
18. M. S. J. Barson *et al.*, *Nano Lett.* **17**, 1496–1503 (2017).
19. L. G. Steele *et al.*, *Appl. Phys. Lett.* **111**, 221903 (2017).
20. F. Casola, T. van der Sar, A. Yacoby, *Nat. Rev. Mater.* **3**, 17088 (2018) Review Article.
21. T. Mittiga *et al.*, *Phys. Rev. Lett.* **121**, 246402 (2018).
22. P. Udvarhelyi, V. O. Shkolnikov, A. Gali, G. Burkard, A. Pályi, *Phys. Rev. B* **98**, 075201 (2018).
23. See supplementary materials for additional details.
24. E. Sterer, M. P. Pasternak, R. D. Taylor, *Rev. Sci. Instrum.* **61**, 1117–1119 (1990).
25. G. Falkovich, *Fluid Mechanics* (Cambridge Univ. Press, ed. 2, 2018).
26. A. Dewaele, P. Loubeyre, M. Mezouar, *Phys. Rev. B Condens. Matter Mater. Phys.* **70**, 094112 (2004).
27. S. Klotz, J.-C. Chervin, P. Munsch, G. Le Marchand, *J. Phys. D Appl. Phys.* **42**, 075413 (2009).
28. R. Taylor, M. Pasternak, R. Jeanloz, *J. Appl. Phys.* **69**, 6126–6128 (1991).
29. L. Oroszlány, A. Deák, E. Simon, S. Khmelevskyi, L. Szunyogh, *Phys. Rev. Lett.* **115**, 096402 (2015).
30. A. Jayaraman, in *Metals* , vol. 1 of *Handbook on the Physics and Chemistry of Rare Earths* (Elsevier, 1978), pp. 707–747.
31. G. K. Samudrala, G. M. Tsoi, S. T. Weir, Y. K. Vohra, *High Press. Res.* **34**, 385–391 (2014).

32. P. Hargraves, R. A. Dunlap, D. J. W. Geldart, S. P. Ritcey, *Phys. Rev. B Condens. Matter* **38**, 2862–2864 (1988).
33. S. Chatterjee, J. F. Rodriguez-Nieva, E. Demler, *Phys. Rev. B* **99**, 104425 (2019).
34. S. Kolkowitz *et al*., *Science* **347**, 1129–1132 (2015).
35. As opposed to isolated NV samples, where $T_1$ is limited by spin-phonon interactions.
36. H. Tang *et al*., *Phys. Rev. Lett.* **71**, 444–447 (1993).
37. B. Feng, V. I. Levitas, R. J. Hemley, *Int. J. Plast.* **84**, 33–57 (2016).
38. P. Kehayias *et al*., *Nat. Commun.* **8**, 188 (2017).
39. D. R. Glenn *et al*., *Geochem. Geophys. Geosyst.* **18**, 3254–3267 (2017).
40. L. Thiel *et al*., *Nat. Nanotechnol.* **11**, 677–681 (2016).
41. M. Atatüre, D. Englund, N. Vamivakas, S.-Y. Lee, J. Wrachtrup, *Nat. Rev. Mater.* **3**, 38–51 (2018).
42. S. Hsieh *et al*., Imaging stress and magnetism at high pressures using a nanoscale quantum sensor, Zenodo (2019); doi.org/10.5281/zenodo.3334404.


## ACKNOWLEDGMENTS

We gratefully acknowledge fruitful discussions with J. Analytis, R. Birgeneau, J. Choi, K. de Greve, M. Eremets, Z. Geballe, F. Hellman, A. Jarmola, J. Jeffries, I. Kim, M. Kunz, D.-H. Lee, S. Lewin, P. Maurer, R. Ramesh, G. Samudrala, E. Zepeda-Alarcon, and R. Zieve. We thank D. Budker, P. Fischer, and H. Zhou for a careful reading of the manuscript and helpful comments. We are especially grateful to M. Barson and M. Doherty for sharing their raw data on stress susceptibilities. We thank C. Laumann for introducing us to the idea of integrating NV centers into diamond anvil cells. **Funding:** This work was supported as part of the Center for Novel Pathways to Quantum Coherence in Materials, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award no. DE-AC02-05CH11231. N.Y.Y. acknowledges support from the David and Lucile Packard foundation and the W. M. Keck foundation. S.H. acknowledges support from the National Science Foundation Graduate Research Fellowship under grant no. DGE-1752814. T.J.S. and R.J. acknowledge support from the U.S. Department of Energy National Nuclear Security Administration under grant no. DE-NA0003842. V.I.L. and M.K. acknowledge support from the Army Research Office (grant W911NF-17-1-0225) and the National Science Foundation (grants CMMI-1536925 and DMR-1904830). **Author contributions:** S.H., P.B., C.Z., T.M., T.J.S., and T.O.H. performed the experiments and collected data. F.M., B.K., N.Z.R., S. Chatterjee, S. Choi, M.Z., J.E.M., and N.Y.Y. developed theoretical models and methodology. S.H, P.B., C.Z., T.M., T.O.H., F.M., B.K., and S. Chatterjee performed the data analysis. M.K. and V.I.L. performed the finite-element simulations. V.V.S. provided equipment and technical expertise for the diamond anvil cells. N.Y.Y. and R.J. conceived the study and supervised the project. All authors contributed to discussions of the data and to the writing of the manuscript. **Competing interests:** University of California (coinventors S.H., P.B., C.Z., T.M., T.J.S., F.M., B.K., S.C., J.E.M., R.J., and N.Y.Y.) filed for a provisional patent (62/782,262) that relates to sensing at high pressures using an apparatus that integrates defects in diamond with diamond anvil cells. **Data and materials availability:** Published data are available on the Zenodo public database (*42*).

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/[6471/1349/suppl/DC1
Supplementary Text
Figs. S1 to S21
Tables S1 and S2
References (*43*–*70*)

21 December 2018; accepted 6 November 2019
10.1126/science.aaw4352

HIGH-PRESSURE PHYSICS

# Measuring magnetic field texture in correlated electron systems under extreme conditions

King Yau Yip[1]*, Kin On Ho[1]*, King Yiu Yu[1]*, Yang Chen[1], Wei Zhang[1], S. Kasahara[2], Y. Mizukami[3], T. Shibauchi[3], Y. Matsuda[2], Swee K. Goh[1,4]†, Sen Yang[1,4]†

Pressure is a clean, continuous, and systematic tuning parameter among the competing ground states in strongly correlated electron systems such as superconductivity and magnetism. However, owing to the restricted access to samples enclosed in high-pressure devices, compatible magnetic field sensors with sufficient sensitivity are rare. We used nitrogen vacancy centers in diamond as a spatially resolved vector field sensor for material research under pressure at cryogenic temperatures. Using a single crystal of $BaFe_2(As_{0.59}P_{0.41})_2$ as a benchmark, we extracted the superconducting transition temperature, the local magnetic field profile in the Meissner state, and the critical fields. The method developed in this work offers a distinct tool for probing and understanding a range of quantum many-body systems.

Strongly correlated electronic systems support a wide variety of phases that are sensitive to external perturbations. For example, the superconducting transition temperature $T_c$ is sensitive to both the interaction strength and the density of states at the Fermi level (*1*), which can in turn be tuned by changing external parameters. Moreover, superconductivity is frequently found to compete with other phases, including magnetically, structurally, and electronically ordered states [for example, (*2*–*5*)]. Therefore, the ability to subject the material system to suitable tuning parameters provides a major experimental tool for reaching new phases.

One of the most successful tuning parameters is hydrostatic pressure, which changes the electronic structure and the interaction strength without introducing additional chemical inhomogeneity to the sample. For many systems, pressure is the only way to reach certain quantum states. Pressure has played an influential role in stabilizing superconductivity by suppressing the competing phases. For example, in the heavy fermion intermetallic $CePd_2Si_2$, pressure suppresses the antiferromagnetic state and induces a superconducting phase with a $T_c$ that peaks at ~28 kbar (*2*). In the iron-based system $BaFe_2As_2$, superconductivity can similarly be induced by means of pressure on the suppression of the spin-density wave state (*6*). More recently, a superconducting state with a remarkably high $T_c$ of 250 to 260 K has been reported in $LaH_{10-\delta}$ at ~200 GPa (*7*, *8*). These results not only reinforce the view that pressure is a powerful tuning parameter but also call for the need to study the microscopic details of superconductivity under extremely high pressure.

To generate high pressure, the sample is enclosed in a pressure cell that is orders-of-magnitude larger than the sample itself. Moreover, to ensure a stable pressure environment, electrical accessibility to the sample volume is severely restricted. Cryogenic conditions impose further restrictions. Under these demanding experimental conditions, very few detection methods can be applied. Placing a robust dc magnetic field sensor in the immediate vicinity of the sample, in particular, is a major experimental challenge.

A negatively charged nitrogen vacancy (NV) center is a point defect in diamond with a spin-1 ground state. Owing to its spin-dependent fluorescence rate, electron spin resonance (ESR) spectrum can be measured with the optically detected magnetic resonance (ODMR) method. From these spectra, we can derive the magnetic field with sensitivity of microtesla $Hz^{-1/2}$ (*9*–*13*), as well as electric field, temperature, and mechanical strain [detailed descriptions are provided in (*20*)] (*14*–*19*). NV centers can sense both the magnitude and the direction of the field, under pressures of up to 60 GPa (*16*, *21*). Furthermore, owing to its small size, a NV center naturally provides high spatial resolution, making the microscopic study of quantum many-body features possible. With these motivations in mind, we combined the field-sensing capability of NV centers and the optical accessibility of a moissanite anvil cell to probe the local magnetic field configuration around the sample at high pressures. In this work, we demonstrate and benchmark the potential of this approach by directly probing the diamagnetism associated with superconductivity in a type II superconductor, $BaFe_2(As_{0.59}P_{0.41})_2$, under pressure.

The sample we used is a piece of single crystalline $BaFe_2(As_{1-x}P_x)_2$ with $x = 0.41$, which comes from the ultraclean family $BaFe_2(As_{1-x}P_x)_2$ (*22*). At $x = 0.33$, $T_c$ is maximized and displays clear evidence of a quantum critical point (*23*, *24*). Hence, $BaFe_2(As_{1-x}P_x)_2$ is an ideal platform with which to explore the interplay between superconductivity and quantum criticality.

An exploded view of the interior of our pressure cell is shown in Fig. 1A, and the zoom-in view illustrates the relationship between the sample and NV center reference frames. A photograph and fluorescence image in the immediate vicinity of the sample are shown in Fig. 1B. The close proximity of the microcoil to the sample ensures the efficient transmission of the microwave power to the sample space, where the NV centers are located. The bright spots in the fluorescence image are from NV centers in diamond particles, which are spread on the sample surface and mixed with pressure-transmitting fluid. The typical size of diamond particles (1 μm) was chosen to be smaller than the optical resolution for better sensitivity but larger than the vortex lattice constant $a_V$ (*20*). In this work, three diamond particles were chosen strategically: $NV_C$ is near the center of the sample, $NV_E$ is off to the side near the edge of the sample, and $NV_F$ is far away from the sample.

Under a weak external magnetic field, for $T > T_c$, the sample is in the normal state, and the magnetic field felt by the NV centers is the same as the external magnetic field (Fig. 1C). However, upon cooling below $T_c$, the expulsion of the magnetic field from the sample alters the field profile near the surface of the material, and this can be felt by the NV centers on the sample surface: For $NV_C$, the effective magnetic field is greatly reduced, whereas for $NV_E$, the effective magnetic field is greatly enhanced (Fig. 1D). When the superconductor is warmed across $T_c$, the diamagnetic response associated with superconductivity disappears at $T_c$. Additionally, for a type II superconductor, when the applied field is higher than the lower critical field ($H_{c1}$), the field begins to thread through the sample, resulting in a vortex state. The vortex state can be completely destroyed above the upper critical field ($H_{c2}$), at which the superconductor returns to the normal state. All these in-field behaviors can be profiled by the NV centers located right on the sample surface.

The pulse sequence shown in Fig. 2A is used to collect our ODMR data. Because the data were collected upon warming up in a weak magnetic field, our experiment probes the diamagnetism associated with superconductivity. To avoid heating caused by microwaves and laser irradiation, we devised a measurement protocol to mitigate measurement-induced

[1]Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China. [2]Department of Physics, Kyoto University, Kyoto 606-8502, Japan. [3]Department of Advanced Materials Science, University of Tokyo, Kashiwa 277-8561, Japan. [4]Shenzhen Research Institute, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China.
*These authors contributed equally to this work.
†Corresponding author. Email: skgoh@cuhk.edu.hk (S.K.G.); syang@cuhk.edu.hk (S.Y.)

perturbations to the superconducting state (*20*). Representative ODMR spectra at 8.3 kbar from the three diamond particles are shown in Fig. 2, B to D, when the sample temperature (~7.7 K) was much lower than $T_c$ (~20.4 K at 8.3 kbar, determined by using ac susceptibility at zero field) (*20*). The ODMR spectra show different splittings. The splittings are caused by the Zeeman effect because of the magnetic field of the surroundings. Therefore, the ODMR spectra provides a means to detect the magnetic field. Zeeman splitting of $NV_C$ (~64 MHz) is 10 times smaller than that of $NV_E$ (~658 MHz) when $T < T_c$, whereas the difference is much smaller when $T > T_c$ (Fig. 2G). The ODMR spectra of $NV_C$ at different temperatures are shown in Fig. 2E, from which the splitting was extracted and plotted in Fig. 2F. Upon warming up, the degree of splitting remains nearly constant initially but then experiences a marked increase that sets in at ~17 K. Above 21 K, the splitting levels off again. To demonstrate the relevance of this feature to superconductivity, we additionally collected the ac susceptibility data in the same experiment, which we could do because of the additional modulation coil added to our experimental configuration. Using the microcoil as the pickup coil, a sharp drop in the ac susceptibility, signifying the superconducting transition (*25*, *26*), was detected at the same temperature (Fig. 2F). The two methods agreed well on the measurement of $T_c$.

The change of the local magnetic field distribution can also be seen in the temperature evolution of the splitting for $NV_E$ and $NV_F$ (Fig. 2G). Contrary to the behavior of $NV_C$, the degree of splitting decreases for $NV_E$ upon warming up. As a reference, the splitting for $NV_F$ is nearly constant in temperature, which can be understood because $NV_F$ is so far from the superconductor that the total field does not change. These observations are well in line with the expectation from the diamagnetism associated with superconductivity, as explained earlier. There was no noticeable change of linewidth or in the overall contrast of the ESR lines. This is because the diamond particles were much smaller in size compared with the magnetic field gradient induced by the diamagnetism associated with superconductivity. Because of the finite size of the sample and the spacing between the diamond particles and the sample, there was a residual magnetic field that caused a Zeeman splitting of ~64 MHz for $NV_C$ at low temperatures. There was also a difference in the Zeeman splitting for three diamond particles when the sample was in the normal state because these diamond particles were randomly oriented relative to the applied magnetic field.

One of the major advantages of our technique is demonstrated in Fig. 2H. As discussed above, both the transverse and longitudinal components of the field relative to a given NV center can be calculated from its ODMR spectrum. This provides the means to reconstruct the field vector. When the sample is in the normal state, the orientation of the NV center can be calibrated against the applied field direction, which is directed along the *c* axis of the sample. This gives an effective magnetic field vector along the *c* axis that can be tracked as a function of temperature (*20*). With these considerations, we determined the effective magnetic field vector felt by $NV_C$, $NV_E$, and $NV_F$ at 8.3 kbar. For $NV_C$, the field vector shortens and tilts away from the vertical direction upon entering the superconducting state. This is consistent with $NV_C$ being on the top of the sample, and that the diamagnetism associated with superconductivity causes the field lines to bend around the sample. However, for $NV_E$, the field vector lengthens and only tilts slightly in the superconducting state. Again, this is consistent with $NV_E$ being located off to the side of the sample, so that the field lines there remain vertical but denser in the Meissner state. Last, the field vector sensed by $NV_F$ remains practically constant across the superconducting phase transition, in stark contrast to the behavior of $NV_E$ and $NV_C$. The ability to collect the complete vectorial information with spatial resolution under extreme conditions

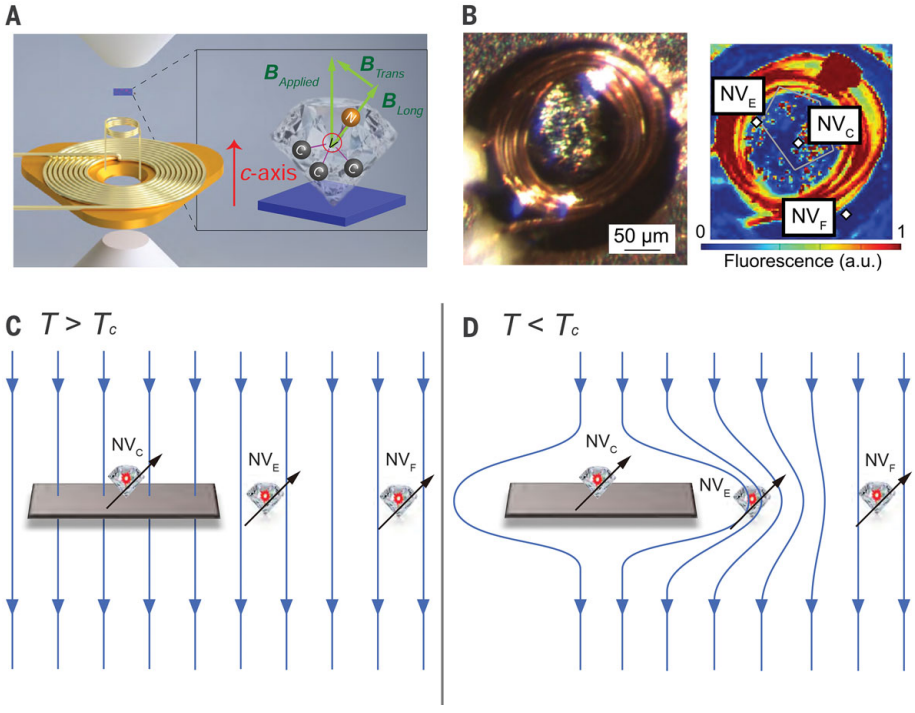



**Fig. 1. Schematic illustration of experimental configurations and detection concepts.** (**A**) An exploded view of our pressure cell design. The sample (blue) is located in the high-pressure chamber together with a collection of diamond particles. Each diamond particle is a sensitive local field sensor. The laser is directed toward the high-pressure chamber through the top moissanite anvil. The microwave is provided by a miniature microcoil in close proximity to the sample, allowing an efficient transmission of microwave power without causing the sample to heat up. The larger coil is added to serve as the modulation coil for auxiliary ac susceptibility measurements (*26*, *34*). The metal part beneath the modulation coil is the gasket. The zoomed-in picture shows two coordinate systems used in this study. One is the sample frame where the *c* axis is the stacking direction of the FeAs planes, and the other is the NV center frame. The external applied magnetic field in this study is always along the sample *c* axis. The diamond particles, each of which contain ~1 million NV centers with four possible quantization axes, are randomly oriented relative to the sample frame (*20*). (**B**) (Left) Photograph of the microcoil with sample on top of the anvil. (Right) Fluorescence image from the confocal scan showing the microcoil and NV centers. The shape of the sample is traced by the pentagon. The location of three particular diamond particles—$NV_C$, $NV_E$, and $NV_F$—are marked. $NV_C$ is near the center of the top surface, $NV_E$ is near the edge, and $NV_F$ is far away from sample and serves as a control sensor. The fluorescence is collected between 650 and 800 nm. (**C** and **D**) Magnetic field profile around the sample under a weak applied magnetic field when (C) $T > T_c$ and (D) $T < T_c$. The expulsion of the magnetic field when $T < T_c$ is a consequence of the diamagnetism associated with superconductivity. The alteration of the field profile in the presence of the superconductor provides an ideal platform with which to demonstrate the performance of our sensor to probe the complete field vectors with spatial resolution under pressure.

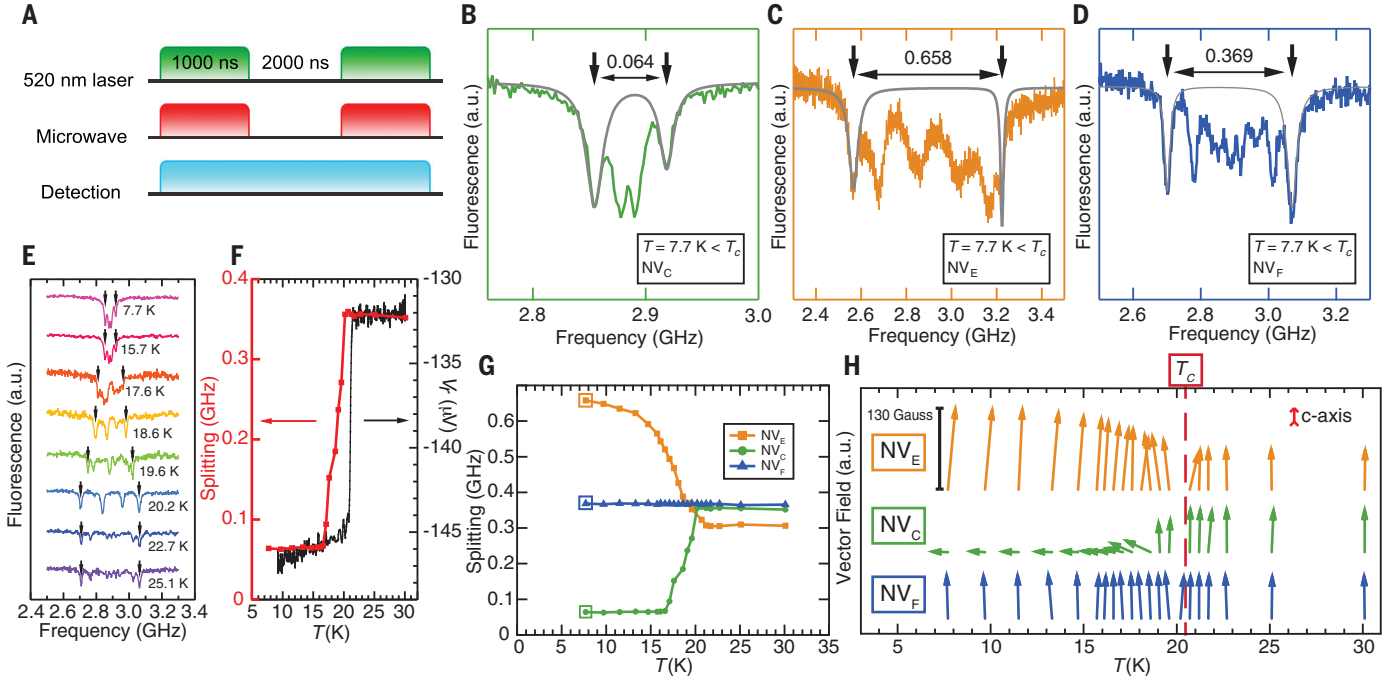



**Fig. 2. The diamagnetism associated with superconductivity sensed by NV centers for $BaFe_2(As_{0.59}P_{0.41})_2$ at 8.3 kbar.** (**A**) Pulse sequence used for ODMR measurements. (**B** to **D**) ODMR spectra of each diamond particle at 7.7 K. The Lorentzian fits for determining the Zeeman splitting are marked with gray lines. (**E**) ODMR spectra of NV centers in $NV_C$ at different temperatures. (**F**) Comparison between the ODMR method (red) and ac susceptibility method (black) in determining the transition temperature $T_c$. (**G**) The change of the Zeeman splitting for NV centers in $NV_C$, $NV_E$, and $NV_F$ as a function of temperature. (**H**) The variation of the local magnetic field vectors felt by NV centers in $NV_C$, $NV_E$, and $NV_F$ across the superconducting phase transition. The vertical direction is the *c* axis of the sample. The ODMR measurements were conducted with a laser power of 10 μW and a peak microwave power of 30 mW. An external *B* field of 68 G is applied along the *c* axis of the sample.

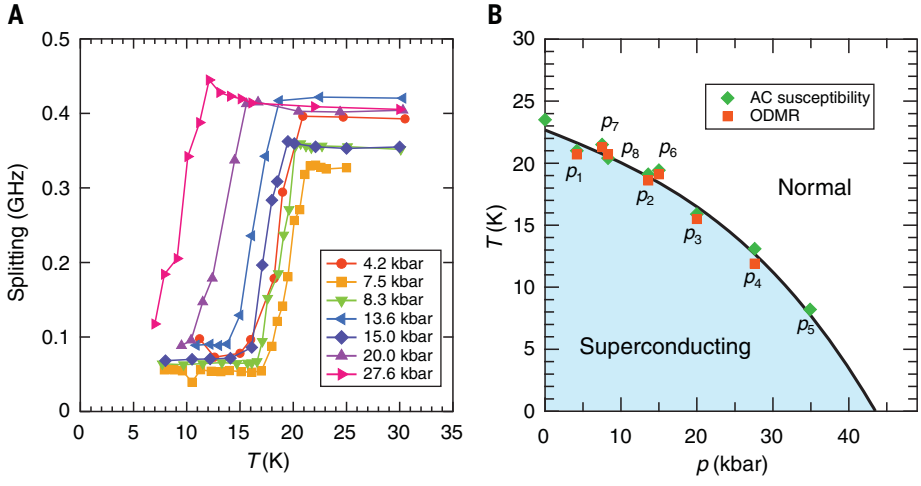



**Fig. 3. Temperature-pressure phase diagram constructed for $BaFe_2(As_{0.59}P_{0.41})_2$ using NV centers.** (**A**) The diamagnetism associated with superconductivity measured by the Zeeman splittings of NV centers under different pressures. The applied magnetic field is (70 ± 5) G. (**B**) The change of $T_c$, measured with the ODMR method (green diamond) and ac susceptibility (red square), against the applied pressure. "$p_1$... $p_8$" shows the sequence of the applied pressures. The error bars are smaller than the symbol sizes.

represents one of the key advances of our technique.

Next, we illustrate the performance of our setup under a varying pressure. In a separate run, we calibrated our ODMR shift against the shift of the ruby fluorescence spectrum up to 60 kbar, confirming our capability to sense the pressure (*20*) and to conduct ODMR experiments at high pressures. We aimed to show that our setup does not lose sensitivity to the superconducting transition when pressure is varied. The temperature dependence of the Zeeman splitting of $NV_C$ at seven different pressure points is shown in Fig. 3A, from which the pressure dependence of $T_c$ can be detected. Additional supporting ac susceptibility data can be found in (*20*). The resultant *T-p* phase diagram (Fig. 3B), where *p* is pressure, shows a suppression of the superconducting state with pressure. This is consistent with $x = 0.41$ being located at the overdoped side of the superconducting dome (*25*). To verify reproducibility, we also collected data on releasing pressure. The overall smooth evolution of $T_c$ against *p* shows that the system is in the elastic regime. This series of experiments confirms the performance of our technique.

The transition width for the two methods in Fig. 2F exhibits a noticeable difference.

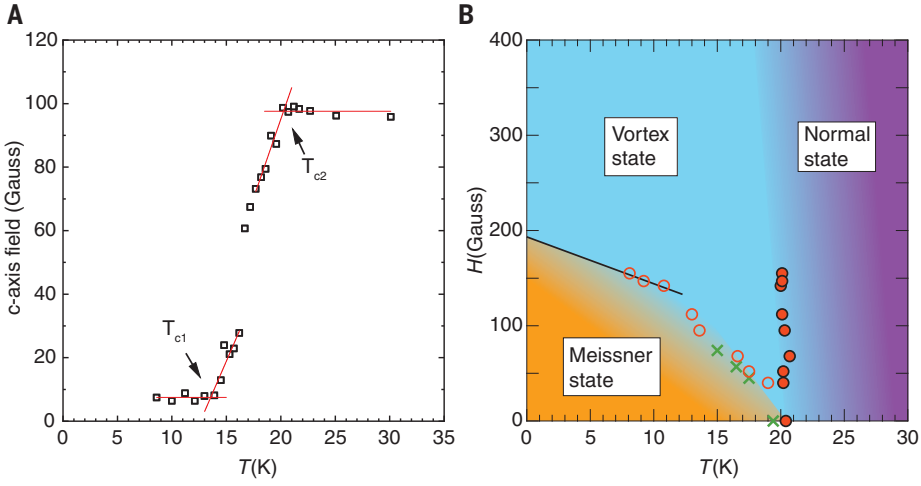



**Fig. 4. Measurement of the lower critical field $H_{c1}(T)$ and the upper critical field $H_{c2}(T)$ of $BaFe_2(As_{0.59}P_{0.41})_2$.** (**A**) The magnetic field along the *c* axis measured for $NV_C$. The applied field along the *c* axis is 95 G, which can be determined from the data at 30 K. The definitions of $T_{c1}$ and $T_{c2}$ are shown. (**B**) Phase diagram showing $\alpha H_{c1}(T)$ (red open circles) and $H_{c2}(T)$ (red solid circles) at 8.3 kbar. Here, the geometry factor $\alpha$ for a thin slab with lengths $l_c$ along the field and $l_a$ perpendicular to the field can be calculated by $\alpha = \tanh\sqrt{0.36(l_c/l_a)}$ (*35*), where $l_c/l_a$ ~ 0.8. Therefore, $\alpha$ is ~0.5. The black line acts as a guide for the eyes. Additional $\alpha H_{c1}(T)$ for 15 kbar are added to the phase diagram for comparison (green crosses). The error bars are smaller than the symbol sizes.

This is because with the application of a magnetic field, the vortex state can be stabilized in a type II superconductor. The larger width for the ODMR-based technique is caused by the NV center, located in the close proximity of the sample, beginning to sense the penetrating field in the form of vortices (ac susceptibility, which probes the average response of the whole sample, is much less sensitive to the vortex state). To probe the phase boundaries, we calculated the magnetic field along the sample *c* axis sensed by $NV_C$. The temperature dependence of the resultant field at 8.3 kbar is shown in Fig. 4A. Below $T_{c1}$ and above $T_{c2}$, the *c* axis field is temperature independent. However, between $T_{c1}$ and $T_{c2}$, a rapid rise of the *c* axis field is detected. This is a consequence of the entry of the magnetic field lines in the form of vortices at $T > T_{c1}$ and the full penetration of the applied magnetic field for $T > T_{c2}$. Using the data at 30 K, which is in the normal state, we can calibrate the value of the applied magnetic field. Thus, this magnetic field must be proportional to $H_{c1}$ at $T_{c1}$ and equal to $H_{c2}$ at $T_{c2}$. Hence, our ODMR data offer the possibility to detect the transition from the Meissner state to the vortex state under pressure.

Repeating the measurements at different applied fields, we can trace out $\alpha H_{c1}(T)$ and $H_{c2}(T)$ for $x = 0.41$ at 8.3 kbar (Fig. 4B), where $H_{c1}(T)$ is the boundary between the Meissner state and the vortex state, whereas $\alpha \sim 0.5$ is a numerical constant that depends on the geometry of the sample. From $H_{c1}(T)$, the temperature dependence of the London penetration depth can be deduced, allowing discussion of the superconducting gap function (*27*, *28*). $H_{c1}(T)$ appears linear at low temperatures and extrapolates to 384 G at 0 K. Both the linearity and the extrapolated $H_{c1}(0)$ value are in good agreement with previous $H_{c1}$ studies conducted for this family of Fe-based superconductors by means of micro-Hall probe array (*27*). On the other hand, the initial slope $|dH_{c2}/dT|_{T_c}$ is proportional to the square of the quasiparticle effective mass relative to the free-electron mass. The almost vertical $H_{c2}(T)$ is consistent with the strongly correlated nature of the material system.

We have successfully demonstrated the use of NV centers in diamond as a vector magnetic field sensor with superior spatial resolution and field sensitivity in pressure cells under cryogenic conditions. The spatial resolution of the protocol shown here can be pushed to <100 nm (*20*). This resolution offers a distinct opportunity to sense the dynamics of magnetically related features such as magnetic domains, vortices (*29*–*32*), and skyrmions in pressure cells. As a noninvasive and contactless method, it can be used to study systems that are too small or too delicate for traditional macroscopic field sensors, such as flakes of two-dimensional materials (*33*). Furthermore, this approach is not limited to magnetic-field sensing. NV center is sensitive to other physical parameters, such as local electric fields and mechanical strain. Therefore, the method demonstrated here can be used in other applications besides magnetic field related processes and becomes a powerful tool in the study of quantum physics in strongly correlated systems under pressure.

## REFERENCES AND NOTES


1. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, ed. 2, 1996).
2. N. D. Mathur *et al.*, *Nature* **394**, 39–43 (1998).
3. S. S. Saxena *et al.*, *Nature* **406**, 587–592 (2000).
4. S. K. Goh *et al.*, *Phys. Rev. Lett.* **114**, 097002 (2015).
5. S. Hosoi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8139–8143 (2016).
6. J. Paglione, R. L. Greene, *Nat. Phys.* **6**, 645–658 (2010).
7. M. Somayazulu *et al.*, *Phys. Rev. Lett.* **122**, 027001 (2019).
8. A. P. Drozdov *et al.*, *Nature* **569**, 528–531 (2019).
9. F. Jelezko, J. Wrachtrup, *Phys. Status Solidi* **203**, 3207–3225 (2006) (a).
10. M. W. Doherty *et al.*, *Phys. Rep.* **528**, 1–45 (2013).
11. G. Balasubramanian *et al.*, *Nature* **455**, 648–651 (2008).
12. J. R. Maze *et al.*, *Nature* **455**, 644–647 (2008).
13. L. Rondin *et al.*, *Rep. Prog. Phys.* **77**, 056503 (2014).
14. P. Neumann *et al.*, *Nano Lett.* **13**, 2738–2742 (2013).
15. G. Kucsko *et al.*, *Nature* **500**, 54–58 (2013).
16. M. W. Doherty *et al.*, *Phys. Rev. Lett.* **112**, 047601 (2014).
17. A. Waxman *et al.*, *Phys. Rev. B Condens. Matter Mater. Phys.* **89**, 054509 (2014).
18. N. M. Nusran *et al.*, *New J. Phys.* **20**, 043010 (2018).
19. K. Joshi *et al.*, *Phys. Rev. Appl.* **11**, 014035 (2019).
20. Materials and methods are available as supplementary materials.
21. L. Steele *et al.*, *Appl. Phys. Lett.* **111**, 221903 (2017).
22. S. Kasahara *et al.*, *Phys. Rev. B Condens. Matter Mater. Phys.* **81**, 184519 (2010).
23. K. Hashimoto *et al.*, *Science* **336**, 1554–1557 (2012).
24. T. Shibauchi, A. Carrington, Y. Matsuda, *Annu. Rev. Condens. Matter Phys.* **5**, 113–135 (2014).
25. L. E. Klintberg *et al.*, *J. Phys. Soc. Jpn.* **79**, 123706 (2010).
26. K. Y. Yip *et al.*, *Phys. Rev. B* **96**, 020502 (2017).
27. C. Putzke *et al.*, *Nat. Commun.* **5**, 5679 (2014).
28. Y. Lamhot *et al.*, *Phys. Rev. B Condens. Matter Mater. Phys.* **91**, 060504 (2015).
29. L. Rondin *et al.*, *Nat. Commun.* **4**, 2279 (2013).
30. M. Pelliccione *et al.*, *Nat. Nanotechnol.* **11**, 700–705 (2016).
31. L. Thiel *et al.*, *Nat. Nanotechnol.* **11**, 677–681 (2016).
32. Y. Schlussel *et al.*, *Phys. Rev. Appl.* **10**, 034032 (2018).
33. Y. Cao *et al.*, *Nature* **556**, 43–50 (2018).
34. P. L. Alireza, S. R. Julian, *Rev. Sci. Instrum.* **74**, 4728–4731 (2003).
35. E. H. Brandt, *Phys. Rev. B Condens. Matter Mater. Phys.* **60**, 11939–11942 (1999).
36. K. Y. Yip *et al.*, Data for Measuring magnetic field texture in correlated electron systems under extreme conditions. Zenodo (2019); doi:10.5281/zenodo.3490189


## ACKNOWLEDGMENTS

We thank D. Dasari, R. Liu, E. Shipton, J. Wrachtrup, and K. Xia for fruitful discussions. We thank S. K. Li for the technical help. **Funding:** S.K., Y.Ma., T.S., and Y.Mi. acknowledge financial supports from JST CREST (JPMJCR19T5), Grants-in-Aid for Scientific Research (KAKENHI) (15H02106, 15H03688, 15KK0160, 18H01177, 18H05227, 18K13492, 18K18727, and 19H00649) and on Innovative Areas "Topological Material Science" (15H05852) "Quantum Liquid Crystals" (19H05824) from the Japan Society for the Promotion of Science (JSPS). T.S. acknowledges the support from the Mitsubishi Foundation. S.K.G. acknowledges financial support from Hong Kong RGC (GRF/14300418, GRF/14300419, and GRF/14301316). S.Y. acknowledges financial support from Hong Kong RGC (ECS/24304617, GRF/14304618, and GRF/14304419), CUHK start-up grant, and the Direct Grants. **Author contributions:** S.K.G. and S.Y. conceived the idea, designed the experiment,

and supervised the project; S.K., Y.Ma., T.S., and Y.Mi. provided the superconductor sample; K.Y.Yip, K.O.H., K.Y.Yu, Y.C., and W.Z. prepared the pressure cell; K.Y.Yip, K.O.H., K.Y.Yu, and S.Y. performed the experiment and analyzed data; S.K.G. and S.Y. wrote the paper; and all authors commented on the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All experimental data shown in the main text and supplementary materials are available at Zenodo (*36*).

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1355/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S19
Tables S1 and S2
References (*37*–*43*)

19 December 2018; accepted 6 November 2019
10.1126/science.aaw4278

HIGH-PRESSURE PHYSICS

# Magnetic measurements on micrometer-sized samples under high pressure using designed NV centers

Margarita Lesik[1]*, Thomas Plisson[2]*†, Loïc Toraille[1]*, Justine Renaud[3], Florent Occelli[2], Martin Schmidt[1], Olivier Salord[3], Anne Delobbe[3], Thierry Debuisschert[4], Loïc Rondin[1], Paul Loubeyre[2], Jean-François Roch[1]†

Pressure can be used to tune the interplay among structural, electronic, and magnetic interactions in materials. High pressures are usually applied in the diamond anvil cell, making it difficult to study the magnetic properties of a micrometer-sized sample. We report a method for spatially resolved optical magnetometry based on imaging a layer of nitrogen-vacancy (NV) centers created at the surface of a diamond anvil. We illustrate the method using two sets of measurements realized at room temperature and low temperature, respectively: the pressure evolution of the magnetization of an iron bead up to 30 gigapascals showing the iron ferromagnetic collapse and the detection of the superconducting transition of magnesium dibromide at 7 gigapascals.

Compression of a solid directly changes its electron density, inducing a variety of magnetic phenomena such as magnetic quantum criticality and high-spin to low-spin transitions (*1*). In addition, pressure can be used to tune superconductivity in a wide range of systems, such as cuprates, transition-metal dichalcogenides, iron pnictides, heavy fermions, and topological superconductors. Recently, a new class of high-temperature superconductors has been discovered in H-rich hydrides at high pressure, with the reported critical temperature of 203 K in $H_3S$ (*2*) and 250 K in $LaH_{10}$ at pressures >150 GPa (*3*, *4*). Systematic exploration of these materials requires sensitive magnetic characterization that can be routinely operated in the 100-GPa range (*5*).

Great efforts have already been made to adapt magnetic detection methods to diamond anvil cell (DAC) specificities (*6*). The macroscopic magnetic susceptibility of the compressed sample can be measured using inductively coupled coils. However, the maximum size of the sample that can be integrated in the DAC decreases with increasing pressure. Two strategies have been pursued to circumvent this poor scalability that is detrimental to detection sensitivity. One is to insert the detection coil inside the sample chamber (*7*) and the other is to miniaturize the whole DAC to integrate it in a superconducting quantum interference device (SQUID) to benefit from the intrinsic high sensitivity of SQUID measurements (*8*). Synchrotron-based methods, such as x-ray magnetic circular dichroism, x-ray emission spectroscopy (XES), and nuclear resonant forward scattering, are also widely used, having been made possible by the development of focused and high-brightness synchrotron x-ray beams. By addressing atomic or nuclear resonance lines, these methods are element specific and resolve local magnetism stemming from a given electronic order (*9*). However, no theoretical framework has been unanimously agreed upon for the interpretation of the signals obtained with these techniques (*10*, *11*) and therefore their quantitative analysis remains challenging. Furthermore, synchrotron-based techniques, which mainly probe nuclear or electronic transitions, are not directly sensitive to magnetic phenomena such as the Meissner effect, which is an unambiguous signature of superconductivity.

We report an optical magnetometry method based on nitrogen-vacancy (NV) color centers used as in situ quantum sensors (Fig. 1A). The main advantages of this method are the easy sample preparation, the tabletop instrumentation, the mapping of the stray magnetic field with micrometer spatial resolution, and the absence of any sensitivity decrease with the sample size down to the micrometer scale. Being noninvasive, the method can be easily combined with complementary structural determination by x-ray diffraction. The method is based on optically detected magnetic resonance (ODMR), which exploits the triplet fine structure of the negatively charged $NV^-$ center ground state (Fig. 1B). ODMR relies on the dependence of the luminescence intensity of the $NV^-$ center whether it is excited from $m_S = 0$ (high-intensity) or $m_S = \pm 1$ (low-intensity) spin states (*12*). This property is used to perform a spectroscopy of the $m_S = 0 \rightarrow \pm 1$ transition excited by a microwave (MW) signal, the frequency of which is scanned across the resonance. Under a continuous MW excitation in the absence of external perturbation, the spin-dependent luminescence recorded as a function of the MW frequency exhibits a resonance peak at 2.87 GHz (Fig. 1C). An external magnetic field applied on the NV center then splits this resonance, leading to the direct measurement of the magnetic field amplitude from the ODMR spectrum (*13*). The influence of strain on this electron spin transition was investigated up to a pressure of ~60 GPa in Doherty et al. (*14*), who envisioned magnetic detection in a DAC as the main goal. Here, we fabricated the NV centers in the diamond anvil and developed a technique that enables magnetometry based on them. We illustrate the efficiency of this technique using two examples: a quantitative measurement of the magnetization of iron up to 30 GPa under ambient temperature around the α–ε phase transition (*15*) and the diamagnetic response associated with superconductivity in $MgB_2$ at 7 GPa and 30 K (*16*).

We used a focused ion beam extracted from a nitrogen plasma to create the NV centers at the culet of an ultrapure synthetic (IIas) diamond anvil (*17*), with a layer of ~$10^4$ defects/$\mu m^2$ at a depth of 20 nm below the surface of the anvil culet (*18*). As shown in Fig. 1A, the optical excitation of these shallow NV centers using a laser of 532 nm wavelength and the detection of their luminescence are performed through the anvil. We then implemented the wide-field ODMR scheme of Steinert et al. (*19*) to record on a camera the image of the spin-state-dependent luminescence emitted by the NV layer as a function of microwave frequency [see the supplementary materials (*18*) for additional information]. The implanted diamond anvil was mounted on a membrane DAC and a rhenium gasket laterally confined the sample. An external single-turn coil was positioned on the gasket as a MW antenna for the ODMR (Fig. 1A). As demonstrated in Steele et al. (*20*), the microwave excitation coil could be embedded in the anvil by metal deposition on the culet covered with a layer of diamond grown using plasma vapor deposition. A ruby crystal was used as a pressure gauge. Microscopic samples of iron or $MgB_2$ were positioned in the sample chamber directly in contact with the implanted anvil (Fig. 2) and embedded in a pressure-transmitting medium consisting of nitrogen or argon, respectively.

In the case of iron, various samples were loaded to test geometric and size effects. The detailed analysis focuses on one of the iron beads [see fig. S14 in the supplementary materials (*18*) for the signals associated with the other samples]. A magnetic field of $B_0 \approx 11$ mT, created by a permanent magnet, was applied to magnetize the iron samples and to split and resolve the resonances associated with the

[1]Laboratoire Aimé Cotton, CNRS, Univ. Paris-Sud, ENS Paris-Saclay, Université Paris-Saclay, 91405 Orsay Cedex, France. [2]CEA, DAM, DIF, 91297 Arpajon, France. [3]Orsay Physics S. A., 13710 Fuveau, France. [4]Thales Research & Technology, 91767 Palaiseau Cedex, France.
*These authors contributed equally to this work.
†Corresponding author. E-mail: thomas.plisson@cea.fr (T.P.); jean-francois.roch@ens-paris-saclay.fr (J.-F.R.)

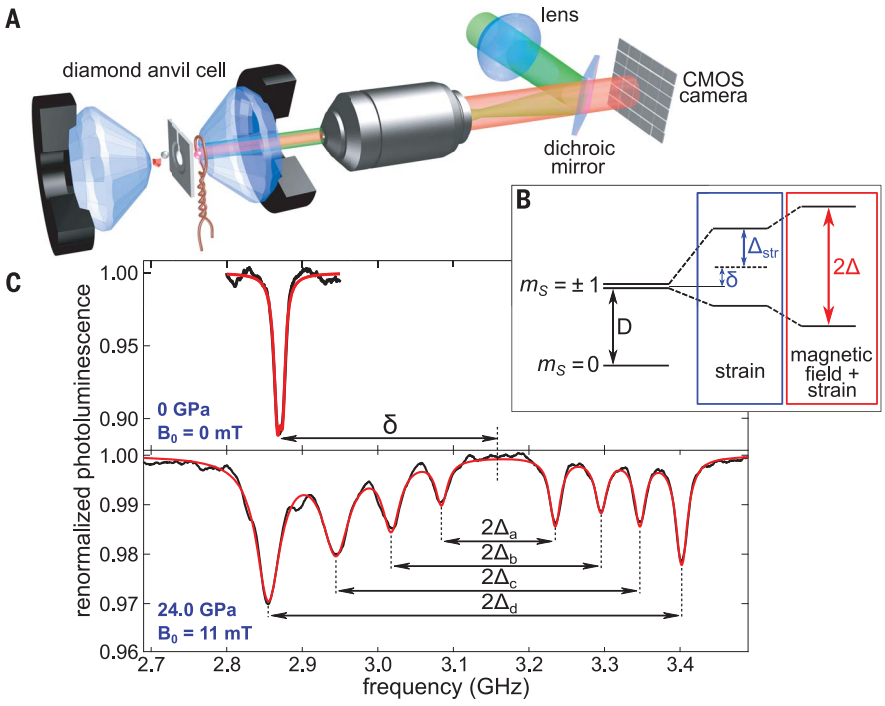



**Fig. 1. Implementation of NV magnetometry in a DAC.** (**A**) Scheme of the DAC setup. A 250-μm-wide disk of NV centers was implanted in the 300-μm-wide culet of one of the two anvils shown in blue. A rhenium gasket (in gray) enclosed the sample and a ruby pressure gauge. A single-loop wire was placed on the gasket to generate the microwave excitation. A green laser with 532-nm wavelength was used to excite the red luminescence of the NV centers. The electron spin resonance was detected through its effect on the luminescence by imaging the layer of NV centers on a camera. (**B**) Electronic structure of the NV center ground state with the modification of the energy levels induced by the strain in the anvil and the magnetic field. Hydrostatic compression shifts the resonance by δ, whereas nonhydrostatic strain splits the resonance into two components with a frequency splitting of $2\Delta_{str}$. The additional influence of the applied magnetic field leads to a total frequency splitting of 2Δ. (**C**) Typical resonance spectra recorded with an ensemble of NV centers. In the absence of any perturbation, the spectrum consists of a single resonance at D = 2.87 GHz. The combination of the strain with the projections of the magnetic field on the four NV orientations in the crystal leads to eight resonance peaks.

four orientations of the NV centers existing in a 100-oriented diamond (*21*).

The energy levels of a given NV center are modified by the magnetic field and by the strain field existing in the anvil. The combination of these two perturbations results in both a shift and a splitting of the MW resonance frequency (*22*). As illustrated in Fig. 1B, the hydrostatic component of the strain shifts the resonance frequency, whereas its nonhydrostatic component and the magnetic field both split the resonance around its center frequency. Extracting the magnetic field created by the iron bead magnetization from these resonances recorded at high pressure requires taking into account the competing effects of the magnetic field and the strain field, which add quadratically (*23*). A typical spectrum obtained in the experiment is shown in Fig. 1C. To first order, the effect of the magnetic field is proportional to its longitudinal component along the N-V axis; this leads to a spectrum consisting of four double resonances linked to the four NV orientations. A map of the measured raw frequency splittings in an area surrounding an iron bead (red square in the sample image in Fig. 2) is shown for each family of NV centers at the pressure of 24 GPa (color plots in Fig. 2). The splitting induced by the iron bead magnetization is on the order of a few megahertz, which can be resolved over the splitting induced by $B_0$. Even before any data analysis, such images can be recorded live during the experiment, providing direct evidence of pressure-induced modifications of the magnetic properties. This optical mapping has a resolution of ~1 μm,

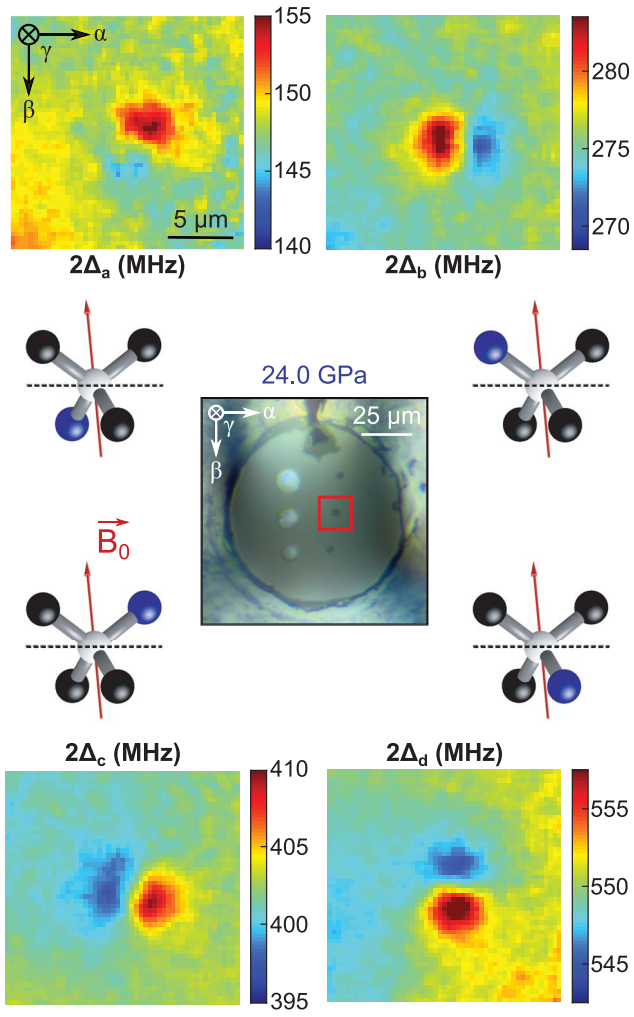



**Fig. 2. Frequency splittings associated with the magnetization of an iron bead.** The center image shows the iron samples inside the gasket. The four color plots show the maps of frequency splittings at 24 GPa for the four NV orientations across the region surrounding an iron bead (red square in the center image). For each NV orientation, the N atom is shown in blue, the vacancy V is shown in white, and the carbon atoms of the lattice are shown in black. The splittings combine the effect of nonhydrostatic strain in the anvil, the applied magnetic field, and the stray magnetic field created by the bead magnetization. α, β, and γ are reference axes linked to the anvil used to identify the four NV orientations. The dotted lines indicate the orientation of the anvil surface.

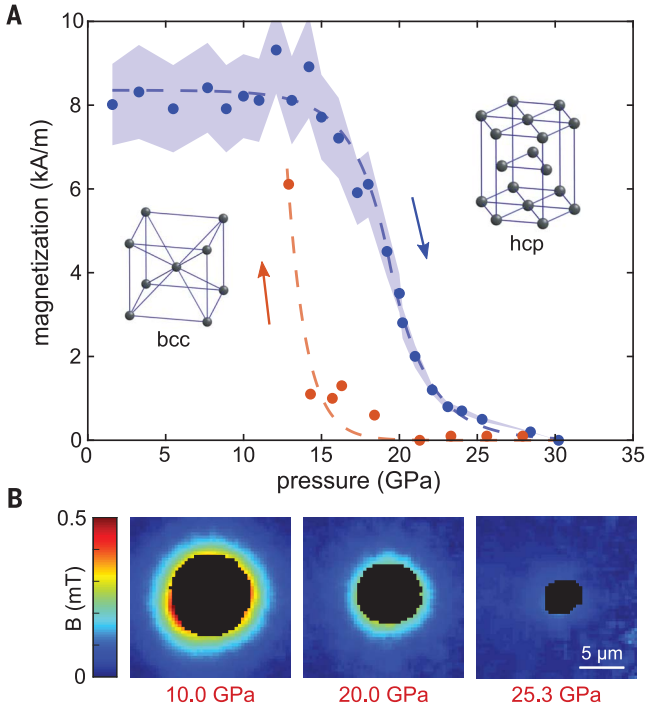



**Fig. 3. Observation of the α–ε transition of iron.** (**A**) Evolution of the bead magnetization inferred at each pressure. The data obtained during the pressure increase are shown by blue dots and those obtained during the pressure release are shown by red dots. The shaded area shows the uncertainty interval on the magnetization value during the pressure increase. The dotted lines are guides for the eye. bcc, body-centered cubic α-phase; hcp, hexagonal close-packed ε-phase. (**B**) Evolution with pressure of the amplitude of the magnetic field created by the iron bead. The mask shown in black is associated with the criteria set on the ODMR contrast.

allowing us to reveal inhomogeneities in the sample magnetization or anisotropies in the strain distribution.

After ascribing each pair of resonances to a given 111 axis of the diamond anvil, the stray magnetic field created by the iron bead can be quantitatively extracted from the correlated information that is embedded in the spectrum shown in Fig. 1C (*24*). First, the strain component was extracted from the ODMR signal using a reference area of the image selected to be far from the bead where the stray magnetic field is negligible. Right below the bead, the NV centers experience a strong transverse magnetic field, which mixes the $m_S = \pm 1$ states and induces a strong decrease of the ODMR contrast (*25*). This quenching leads to a background signal without a direct link to the iron bead magnetization. Additionally, the microwave excitation is screened by the eddy currents induced in the iron sphere; this effect also contributes to the decrease of the ODMR contrast. As shown in Fig. 3B, the area below the center of the bead was excluded by applying a mask to the data according to a threshold set on the contrast [see the supplementary materials (*18*) for the data processing]. At high pressures, the bead magnetization decreases and the masked area decreases accordingly.

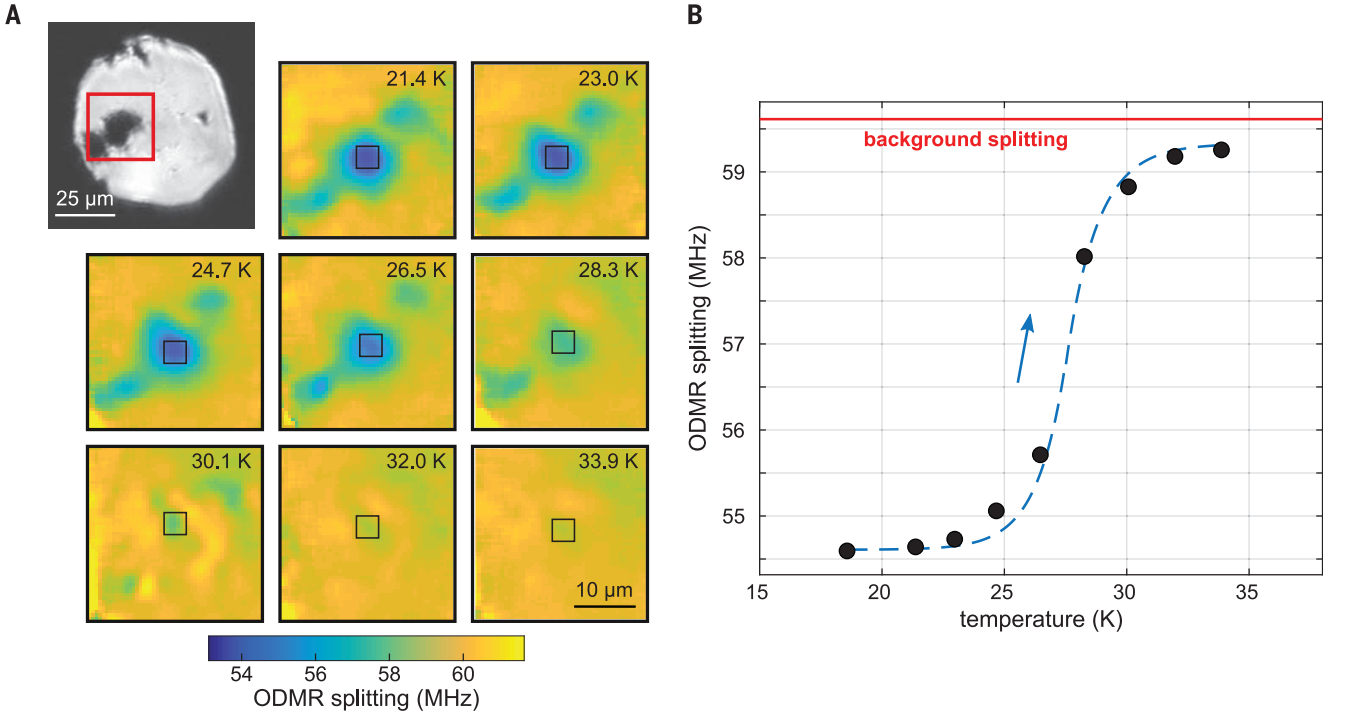



**Fig. 4. Exclusion of the magnetic field by the superconducting state of $MgB_2$.** (**A**) Maps of the ODMR frequency splitting above the $MgB_2$ sample recorded for an increasing temperature. A control magnetic field $B_0 \approx 1.8$ mT is applied and induces a background ODMR splitting. The total ODMR splitting combines the influence of $B_0$ with the strain in the anvil. Below 30 K, the exclusion of the magnetic field is observed above the $MgB_2$ sample owing to the diamagnetic response associated with superconductivity. This effect disappears above 30 K, leading to a homogeneous distribution of the ODMR splitting. Inset: Optical image of the sample. The red square indicates where the ODMR splitting is mapped. (**B**) Evolution of ODMR splitting above the sample when the temperature is increased. The data points are averaged on the black squares of (A). The dotted line is a guide for the eye.

The magnetization $M$ of the iron bead was then determined by fitting the relevant part of the magnetic field map with a simple magnetic dipole model (*18*). At low pressure, we obtained $M \approx 8 \pm 1$ kA•m$^{-1}$. The evolution of the magnetization with pressure is shown in Fig. 3A. The magnetic field of the iron bead decreases as the pressure increases from 15 GPa up to ~30 GPa (Fig. 3B), above which no value of the magnetization can be inferred. This result demonstrates the sensitivity of the detection scheme and is consistent with previous XES (*26*, *27*) and SQUID measurements (*28*) that reported magnetic signals well above the α–ε transition threshold. Whether a remnant of the α phase is responsible for this magnetic signature up to 30 GPa could be investigated by implementing NV-based magnetometry on an x-ray diffraction beamline, combining structural characterization and a direct measurement of the magnetization on the same sample at each pressure. Finally, upon the release of pressure in the DAC, we observed the reappearance of the sample magnetization (Fig. 3A) with the expected hysteresis behavior related to the hysteresis of the structural transition (*15*).

This technique can be straightforwardly implemented at low temperature to observe the diamagnetic response associated with superconductivity. As a proof-of-principle experiment using a testbed compound, we chose a sample of $MgB_2$ that was confined in the DAC at 7 GPa and first cooled in a cryostat in the absence of external magnetic field. At a temperature of ~18 K, an external magnetic field ≈1.8 mT was applied. The direction of this applied magnetic field was chosen parallel to the 100-diamond axis so that the four NV orientations in the crystal had identical responses. As shown in Fig. 4A, the comparison between the frequency shift in the ODMR spectra of the NV centers located above the sample and the homogeneous background gives a direct image of the exclusion of the magnetic field above the $MgB_2$ sample. This shielding is induced by the superconductor diamagnetic response to the applied magnetic field. Under heating of the DAC, the exclusion of the magnetic field disappeared >30 K (Fig. 4B). This behavior agrees with the reported pressure evolution of the critical temperature of $MgB_2$ superconductivity (*16*).

It would be worthwhile to investigate how this direct optical detection of superconductivity can be implemented at a pressure range >100 GPa. This may require adapting the excitation and readout wavelengths of the NV center to compensate for the pressure-induced energy shifts of the electronic levels (*14*). NV engineering based on controlled nitrogen doping during the plasma-assisted growth of a diamond layer (*29*) or using laser writing (*30*) can bury a thin sheet of NV centers at a depth where the influence of strain in the anvil could be less detrimental. We anticipate a major research direction to be the investigation of high-temperature superconductivity in the various superhydride compounds that can be directly synthesized at high pressure, such as $H_3S$ (*2*), $LaH_{10}$ (*3*, *4*), $UH_7$ (*31*), and $FeH_5$ (*32*). The technique could also enable the observation of the predicted magnetic properties of metallic hydrogen (*33*–*36*), for which various challenging experimental probes have already been proposed (*37*, *38*).

## REFERENCES AND NOTES


1. H. K. Mao, X. J. Chen, Y. Ding, B. Li, L. Wang, *Rev. Mod. Phys.* **90**, 015007 (2018).
2. A. P. Drozdov, M. I. Eremets, I. A. Troyan, V. Ksenofontov, S. I. Shylin, *Nature* **525**, 73–76 (2015).
3. M. Somayazulu *et al.*, *Phys. Rev. Lett.* **122**, 027001 (2019).
4. A. P. Drozdov *et al.*, *Nature* **569**, 528–531 (2019).
5. L. P. Gorkov, V. Z. Kresin, *Rev. Mod. Phys.* **90**, 011001 (2018).
6. M. I. Eremets, *High-Pressure Experimental Methods* (Oxford Univ. Press, 1996).
7. D. D. Jackson *et al.*, *Rev. Sci. Instrum.* **74**, 2467–2471 (2003).
8. A. Marizy, B. Guigue, F. Occelli, B. Leridon, P. Loubeyre, *High Press. Res.* **37**, 465–474 (2017).
9. R. Torchio *et al.*, *Phys. Rev. Lett.* **107**, 237202 (2011).
10. O. Mathon *et al.*, *J. Synchrotron Radiat.* **11**, 423–427 (2004).
11. J. P. Rueff, A. Shukla, *Rev. Mod. Phys.* **82**, 847–896 (2010).
12. M. W. Doherty *et al.*, *Phys. Rep.* **528**, 1–45 (2013).
13. J. M. Taylor *et al.*, *Nat. Phys.* **4**, 810–816 (2008).
14. M. W. Doherty *et al.*, *Phys. Rev. Lett.* **112**, 047601 (2014).
15. W. A. Bassett, E. Huang, *Science* **238**, 780–783 (1987).
16. C. Buzea, T. Yamashita, *Supercond. Sci. Technol.* **14**, R115–R146 (2001).
17. M. Lesik *et al.*, *Phys. Status Solidi., A Appl. Mater. Sci.* **210**, 2055–2059 (2013).
18. See the supplementary materials for additional information.
19. S. Steinert *et al.*, *Rev. Sci. Instrum.* **81**, 043705 (2010).
20. L. G. Steele *et al.*, *Appl. Phys. Lett.* **111**, 221903 (2017).
21. L. Toraille *et al.*, *Nano Lett.* **18**, 7635–7641 (2018).
22. J. Teissier, A. Barfuss, P. Appel, E. Neu, P. Maletinsky, *Phys. Rev. Lett.* **113**, 020503 (2014).
23. M. S. J. Barson *et al.*, *Nano Lett.* **17**, 1496–1503 (2017).
24. M. Chipaux *et al.*, *Eur. Phys. J. D* **69**, 166 (2015).
25. J.-P. Tetienne *et al.*, *New J. Phys.* **14**, 103033 (2012).
26. A. Monza *et al.*, *Phys. Rev. Lett.* **106**, 247201 (2011).
27. B. W. Lebert *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20280–20285 (2019).
28. Q. Wei, C. McCammon, S. A. Gilder, *Geochem. Geophys. Geosyst.* **18**, 4646–4654 (2017).
29. M. Lesik *et al.*, *Phys. Status Solidi., A Appl. Mater. Sci.* **213**, 2594–2600 (2016).
30. Y.-C. Chen *et al.*, *Nat. Photonics* **11**, 77–80 (2017).
31. I. A. Kruglov *et al.*, *Sci. Adv.* **4**, eaat9776 (2018).
32. C. M. Pépin, G. Geneste, A. Dewaele, M. Mezouar, P. Loubeyre, *Science* **357**, 382–385 (2017).
33. R. P. Dias, I. F. Silvera, *Science* **355**, 715–718 (2017).
34. X.-D. Liu, P. Dalladay-Simpson, R. T. Howie, B. Li, E. Gregoryanz, *Science* **357**, eaan2286 (2017).
35. A. F. Goncharov, V. V. Struzhkin, *Science* **357**, eaam9736 (2017).
36. P. Loubeyre, F. Occelli, P. Dumas, Observation of a first order phase transition to metal hydrogen near 425 GPa. arXiv:1906.05634 [cond-mat.mtrl-sci] (13 June 2019).
37. E. Babaev, A. Sudbø, N. W. Ashcroft, *Phys. Rev. Lett.* **95**, 105301 (2005).
38. J. P. Carbotte, E. J. Nicol, T. Timusk, *Phys. Rev. Lett.* **121**, 047002 (2018).
39. M. Lesik, T. Plisson, L. Toraille, J. Renaud, F. Occelli, M. Schmidt, O. Salord, A. Delobbe, T. Debuisschert, L. Rondin, P. Loubeyre, J.-F. Roch, Replication data for: Magnetic measurements on micron-size samples under high pressure using designed NV centers, Zenodo (2019); https://doi.org/10.5281/zenodo.3249952.


## ACKNOWLEDGMENTS

We thank V. Jacques, K. De Hantsetters, N. Vast, and F. Bouquet for fruitful discussions. **Funding:** This project received funding from the European Union's research and innovation program (H2020-FETFLAG-2018-2020) under grant agreement no. 820394 (ASTERIQS), the Agence Nationale de la Recherche (projects Envie-FIB and ASPEN), the Investissements d'Avenir program through the LabEx PALM (ANR-10-LABX-0039-PALM), the Paris-Saclay Strategic Research Initiative for Quantum Engineering (IQUPS), and the Paris Île-de-France Région in the framework of DIM SIRTEQ. **Author contributions:** J.F.R., T.P., and P.L. designed and supervised the project; M.L., T.P., L.T., and J.F.R. performed the ODMR experiments; T.P., F.O., and P.L. assembled the high-pressure cells; L.T., T.D., L.R., T.P., and J.F.R. developed the wide-field magnetic imaging system and the magnetization measurement; J.R., M.L., O.S., A.D, M.S., and J.F.R. customized the Orsay Physics FIB; M.L., J.R., M.S., and O.S. performed the nitrogen implantation in the diamond anvils; L.T., M.L., T.P., and J.F.R. performed the data analysis. All authors contributed to manuscript preparation. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data described here are available at Zenodo (*39*).

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1359/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
References (*40*–*42*)

21 December 2018; accepted 6 November 2019
10.1126/science.aaw4329

MARTIAN ATMOSPHERE

# Global circulation of Mars' upper atmosphere

M. Benna[1,2]*, S. W. Bougher[3], Y. Lee[1,2], K. J. Roeten[3], E. Yiğit [4], P. R. Mahaffy[1], B. M. Jakosky[5]

**The thermosphere of Mars is the interface through which the planet is continuously losing its reservoir of atmospheric volatiles to space. The structure and dynamics of the thermosphere is driven by a global circulation that redistributes the incident energy from the Sun. We report mapping of the global circulation in the thermosphere of Mars with the Mars Atmosphere and Volatile Evolution (MAVEN) spacecraft. The measured neutral winds reveal circulation patterns simpler than those of Earth that persist over changing seasons. The winds exhibit pronounced correlation with the underlying topography owing to orographic gravity waves.**

Exploration of the atmospheres of the Solar System terrestrial planets (Venus, Earth, and Mars) has shown that global circulation of neutral and ionized gases in the high-altitude regions of the atmosphere (including the thermosphere and ionosphere) dominates the dynamical state and evolution of these planetary environments [e.g., (*1*)]. Thermospheric circulation locally and globally distributes the deposited energy and momentum imparted both from below, by lower atmosphere waves, and from above, by the solar wind, and subsequently determines how the lower atmosphere and magnetosphere combine to interact with the thermosphere [e.g., (*2*)]. Although Earth's thermospheric circulation has been the target of extensive studies [e.g., (*3*, *4*)], the nature of Mars' thermospheric transport and its variability remain largely uncharacterized owing to the scarcity of direct observations. The Mars Global Surveyor (MGS), Mars Odyssey, and Mars Reconnaissance Orbiter spacecraft collected data on the dynamics of Mars' global circulation and its seasonal variability [e.g., (*5*–*8*)]. However, those observations provided very limited and incomplete constraints for global circulation models [e.g., (*9*–*12*)]. We report observations from the Neutral Gas and Ion Mass Spectrometer (NGIMS) (*13*) on the Mars Atmosphere and Volatile Evolution (MAVEN) spacecraft to investigate Mars' planet-wide thermospheric circulation.

Since entering Mars' orbit in 2015, NGIMS has collected in situ measurements of the composition of the thermosphere and ionosphere of Mars. In April 2016, the instrument's operation mode was modified to measure both speed and direction of neutral flows along the spacecraft path through the atmosphere (*14*). These observations took advantage of MAVEN's articulated payload platform (APP), on which the instrument is mounted, to rapidly and periodically swing NGIMS' boresight direction by ±8° in the spacecraft's local horizontal plane. This technique allows the instrument to measure neutral winds at altitudes of ~140 to 240 km with typical uncertainty of 20 m $s^{-1}$ for the along-track component and 6 m $s^{-1}$ for the cross-track component, with a measurement frequency of ~0.03 Hz. Flow directions and speeds are reconstructed during data processing of the modulated $CO_2$ abundance in the thermosphere using spacecraft pointing and trajectory information (*14*). In a typical orbit, neutral wind observations are conducted along a spacecraft track that extends over 2000 km and an altitude change of up to 80 km. Two successive wind measurements are separated in time by the 30 s it takes for the APP to complete a single swing and in space by ~130 km horizontally and ~10 km vertically as a result of the spacecraft motion.

To mitigate the effect of short- and long-term variabilities of the neutral flow patterns (on scales of minutes and hours, respectively), most wind observations were collected over monthly campaigns. During most campaigns, NGIMS conducted a set of 5 to 10 successive neutral wind observations at nearly the same local time–latitude (LT-LAT) combination. These observations were separated by the

[1]Solar System Exploration Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA. [2]Center for Space Sciences and Technology, University of Maryland Baltimore County, Baltimore, MD, USA. [3]Department of Climate and Space Science and Engineering, University of Michigan, Ann Arbor, MI, USA. [4]Department of Physics and Astronomy, George Mason University, Fairfax, VA, USA. [5]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO, USA.
*Corresponding author. Email: mehdi.benna@nasa.gov

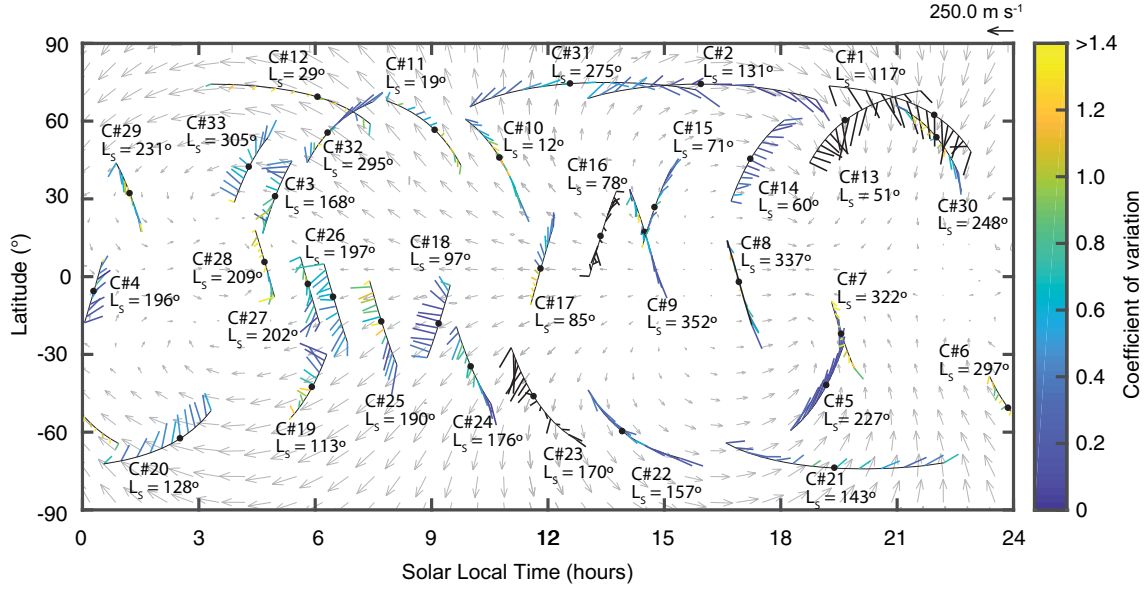



**Fig. 1. Dominant neutral winds observed using NGIMS during 33 monthly campaigns from April 2016 to December 2018.** The average neutral winds are shown as whiskers. The magnitude of their orbit-to-orbit variability are reflected in the whisker colors. The orbit-to-orbit variability of the winds for a given location is captured by the GCV (color bar). No variability is assessed for campaigns that involved a single orbit (black whiskers of C#1, 13, 16, and 23). The MAVEN ground tracks are shown as black traces, with periapsis locations indicated by black dots. The season of each campaign is indicated by the solar longitude ($L_S$). A neutral wind flow map generated by the M-GITM model (*9*) is shown in gray arrows for comparison. This map represents the expected winds at 150 km for $L_S$ = 180° season and under moderate solar EUV input. More detailed and season-specific comparisons between winds produced by the M-GITM model and a select number of neutral wind campaigns are provided elsewhere (*16*).

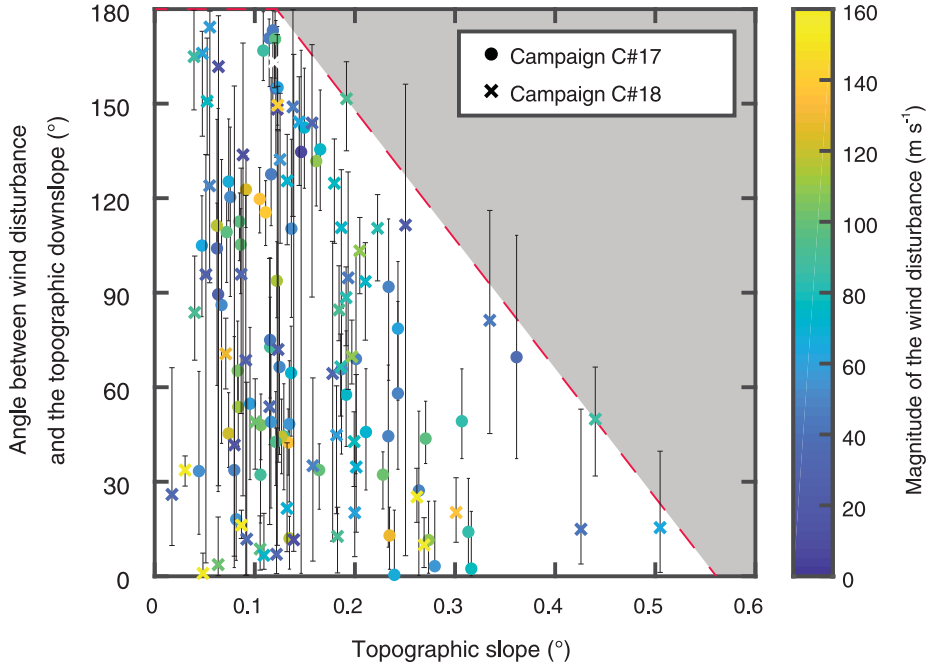



**Fig. 2. Topographic correlation of the deviations from the mean wind for campaigns C#17 and 18.** Wind disturbances tend to follow the downslope of the local terrain (low angular deviations) when its topographic slope is high. This preferential alignment creates the exclusion region depicted in gray. The bounds of the exclusion region (dashed red line) are used as a measure of the degree of alignment of the wind disturbances. This degree of alignment increases linearly with the sloping of the terrain, independently of the disturbance magnitude. For flat terrains (slopes <0.15°), the disturbance ceases to follow a preferential direction. The error bars reflect the propagated 1σ uncertainties of the measured disturbances.

4.5-hour orbital period of MAVEN. Between April 2016 and December 2018, a period of 1.5 Mars seasons, 33 campaigns (designated C#1 to 33) were completed, spanning latitudes of ±70° and nearly all local times. The list of the individual orbits that make up these campaigns is provided in Data S1. For each campaign, the dominant wind direction and magnitude were assessed by averaging the wind vectors at a fixed LT-LAT location and altitude. The magnitude of the orbit-to-orbit variability of the winds was parametrized by the generalized coefficient of variation (GCV), a consolidated measure of changes in both flow direction and magnitude (*14*). The GCV metric is a multivariate extension to the classical single-variable coefficient of variation, which provides a consolidated measure of wind variability across all campaigns. Dominant neutral winds (average wind vectors assessed for each campaign) measured over the 33 campaigns are depicted in Fig. 1 and provided in numerical form in Data S2.

These average neutral wind observations provide a global view of thermospheric circulation on Mars. On the day side, winds originate from the equatorial postnoon region, corresponding to the location of the expected high–atmospheric density and high-temperature region [e.g., (*9*)]. The stagnant winds of this region were captured during campaigns C#8 and 9. The observations show that this thermospheric mass is transported from day to night either through transpolar eastward and westward flows (e.g., C#1 and 2, 11 and 12, 21 to 23, and 30 and 31) or through an equatorial westward path (e.g., C#17 and 18 and 25 to 27). The high-latitude transpolar flows on the dusk side (C#1 and 2, 21 and 22, and 30 and 31) exhibit small orbit-to-orbit variations and have a mean speed of 211 m $s^{-1}$. Those on the dawn side are slower (an average of 97 m $s^{-1}$) and more variable. In the postdusk area (LT = 19 to 23 hours), the observed rotation of the flow westward (C#1, 13, 5, 21, and 30) is consistent with the presence of a convergent zonal wind region where dayside eastward flowing and nightside westward flowing winds meet. The location of the convergence region was captured at LT = 19.5 hours as the very slow but highly variable winds of C#7 during solar longitude ($L_S$) = 332°. The location of this convergence region seems to vary with seasons, because it was not present when the same LT-LAT was mapped 4.5 months earlier during C#5 ($L_S$ = 227°). In the predawn, the observed winds (C#3 and 4, 19 and 20, 25 to 29, and 33) are consistent with the presence of a second region of flow convergence near LT = 2 to 6 hours and LAT = ±30° where the thermospheric gas mass is most likely descending (*15*). The center of this convergence region was most likely captured as the slow and variable winds of C#28.

Figure 1 also provides a qualitative comparison between the average measured winds and the neutral wind flow map predicted by the Mars Global Ionosphere-Thermosphere Model (M-GITM) (*9*) at 150 km for $L_S$ = 180° season and under moderate solar extreme ultraviolet (EUV) input. The northern hemisphere autumn equinox and moderate solar EUV settings correspond to the average conditions encountered by NGIMS over the course of the 33 campaigns. This comparison is notional because the NGIMS observations span multiple seasons and solar activities and are taken over a wide range of altitudes (~140 to 240 km). A quantitative comparison between the measured winds and those modeled by M-GITM has been performed by Roeten *et al.* for five selected campaigns (*16*). They found that, in some cases, the magnitudes and/or directions of the M-GITM–simulated winds match those of the observed winds, whereas in others, large differences occur. The role of normal solar forcing [i.e., net radiative heating largely owing to the difference between the absorbed solar EUV and ultraviolet (UV) irradiance and which radiated back to space as $CO_2$ 15-μm emission] in driving thermospheric winds at Mars seems to be variable. In cases with large differences between the observed and simulated winds, processes other than normal solar forcing may have altered the expected circulation pattern.

All NGIMS wind observations were taken under conditions of moderate or minimum solar EUV activity. The global-scale thermospheric wind pattern is predicted to remain much the same during time periods when higher solar fluxes occur [e.g., (*9*)]. However, the magnitude of the winds should increase substantially as solar maximum conditions are approached and in situ EUV and UV heating is enhanced. Simulations with the Mars Thermosphere General Circulation Model predicted that horizontal wind magnitudes at equinox should increase by nearly ~125 m $s^{-1}$ at ~200 km as solar fluxes rise from solar minimum to maximum conditions [10.7-cm solar radio flux index (F10.7) = 70 to 200 solar flux units] (*17*). Likewise, M-GITM predicts terminator horizontal winds to vary by ~100 m $s^{-1}$

(equinoxes) and ~140 to 200 m $s^{-1}$ (solstices) for the same increase of solar activity (*9*). Thus, the combination of solar activity (primary) and seasonal variations is expected to contribute substantially to the change in magnitude of the martian thermospheric winds when in situ EUV-UV heating is the primary driver.

Owing to planetary rotation, the MAVEN spacecraft overflies different martian terrains on successive orbits. Analysis of orbit-to-orbit variations in the measured winds during campaigns C#17 and 18 revealed a pronounced correlation between the observed wind variability and the underlying topography of the planet. These two campaigns are unlike the others in two aspects. First, they covered the latitudinal band of ±30° over which Mars exhibits its topographic transition between the southern highlands and the northern lowlands. This band also encompasses the pronounced elevated terrains of the Tharsis plateau and its shield volcanoes. Second, these two campaigns captured the strong and steady equatorial westward flows that advect thermospheric gases from day to night. These steady flows exhibit the lowest measured orbit-to-orbit GCV of our equatorial campaigns. The inherent stability of the neutral flows in this region allows us to detect the faint disturbances that are induced by the varying martian topography. By contrast, campaigns that occurred in the equatorial eastward regions (C#7 to 9) exhibited a high intrinsic variability owing to the presence of the stagnant wind region (LT = 15 to 17 hours) and the convergence region at LT = 19.5 hours where flows are inherently slow and unorganized. Campaigns C#25 and 26, although also equatorial and with strong winds, took place during the planet-encircling dust event of June 2018 (*18*) that disturbed the thermosphere.

The correlation between the measured winds and the topography was assessed for campaigns C#17 and 18 by comparing the measured wind vectors for a given orbit with the mean vectors derived by averaging measurements taken at a given altitude from all orbits of the campaign (as shown in Fig. 1). The average wind provides a close approximation of the thermospheric flow over a featureless terrain (devoid of any topographic relief). The deviation from the average of the wind vectors that were observed on a given orbit were compared to topographic slopes of the terrain reconstructed from a 1°-resolution MGS elevation map (*19*).

Figure 2 shows that the wind disturbances of C#17 and 18 tend to preferentially align with the downslope direction of the underlying terrain. This alignment occurs for topographic slopes higher than 0.122° and is more pronounced over regions that exhibit large topographic gradients. However, over locations where the terrain has no pronounced sloping, the directions of the wind disturbances are nearly randomly distributed about the direction of the downslope. The disturbances that are shown in Fig. 2 have an average magnitude of 75 m $s^{-1}$ compared with average wind speeds of 195 m $s^{-1}$. A statistical hypothesis test (*14*) shows that this correlation has a probability of ~$1.8 \times 10^{-8}$ of resulting from a combination of statistical randomness and the undersampling of observations over terrains with high topographic elevation. We therefore conclude that the correlation is statistically significant.

An overlay of the wind disturbance vectors on a surface elevation map are shown in Fig. 3 for three orbits of campaigns C#17 and 18. The wind disturbances measured during MAVEN orbits 6047 and 6199 flow away from the elevated terrains of the Tharsis plateau. However, the same wind pattern is disturbed from its

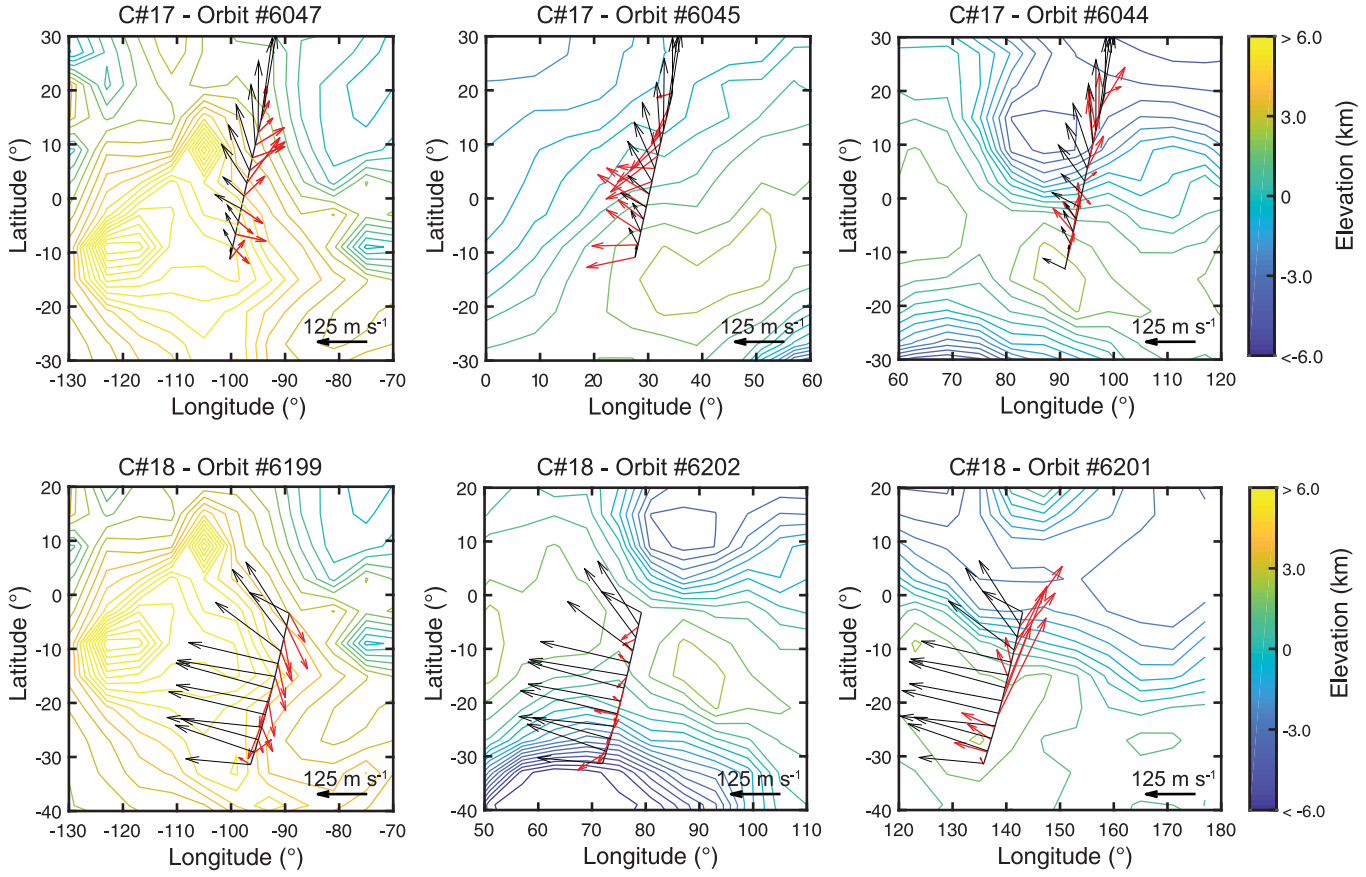



**Fig. 3. Relationship between surface topography, mean flows, and wind disturbances observed during three orbits of campaigns C#17 and 18.** The mean flow (black arrows) is derived by averaging measurements taken at the same altitude from all orbits of the campaign. The shown wind disturbances (red arrows) flow away from high elevation following the terrain local gradient.

average direction during MAVEN orbits 6045 and 6202, flowing down toward the low elevations of Arabia Terra and the Hellas basin. These winds flow downslope toward the Isidis basin and Hellas Planitia on orbits 6044 and 6201.

The apparent alignment of the observed wind disturbances with the underlying terrain is most likely the result of the winds flowing around and past standing features in the thermosphere. We interpret these standing thermospheric features as perturbations in both density and temperature that are linked to the underlying topography. The ionosphere can be ruled out as a possible source of these topographically linked thermospheric features for two reasons. First, the wind disturbances were observed at altitudes well above the dynamo region (120 to 160 km) where ion-neutral collisions dominate (*20*) and where ion drag can exert some influence on the thermosphere. Second, the crustal magnetic fields that can generate ionospheric structures with strong connection to the surface exhibit a topology that is weakly linked to the topography of the planet. Because the topographic relief that is responsible for these thermospheric standing features lies 150 to 200 km below the locations where NGIMS is sensing the winds, they most likely exert their influence in the lower thermosphere by means of upward propagating gravity waves.

Flow over terrain with a stably stratified atmosphere is known to produce vertically propagating gravity waves (*21*). Such gravity waves are often referred to as orographic gravity waves (or mountain waves) to distinguish them from convectively generated gravity waves. Gravity waves transport their energy and momentum upward to the middle and upper atmosphere, where they dissipate into the mean flow through wave breaking, diffusion, nonlinear interactions, and viscosity (*22*). This localized energy and momentum deposition is the most likely source of the observed correlation. However, gravity wave–driven transport was not forecast (*23*) to reach the high altitudes of the MAVEN observations (up to 240 km).

Depending on the background winds and temperature, orographic waves are predicted to rise nearly vertically to the thermosphere. Their signature can also reach indirectly higher altitudes through nonlinear interaction (*24*, *25*) with upward propagating nonorographic gravity waves. A combination of the two processes may be responsible for the observed correlation. The ability of gravity waves to propagate upward is controlled to a large extent by their intrinsic phase speed—specifically, horizontal phase speed relative to the background atmospheric flow. Harmonics with larger intrinsic phase speeds have a greater potential to penetrate deeper into the thermosphere (*26*). Terrain-generated gravity waves have zero phase speeds with respect to surface features and appear fixed in space for a stationary observer. However, relative to the background circulation, these waves are propagating upwind at the same speed as the background wind, which allows them to reach higher altitudes. The spectrum of the waves generated in response to the irregular topographical features is expected to be a function of the flow conditions and topographical properties.

Overall, the global circulation pattern detected by MAVEN appears simpler and more stable over martian seasons than predicted by models of Earth [e.g., (*27*)]. The persistence of this global pattern over changing seasons indicates the stability of the martian climate on seasonal time scales. By contrast, the observed high variability of the winds on a time scale of hours is consistent with the observed short-term fluctuations and extensive structures of thermospheric densities and temperatures (*28*–*30*). This short-term variability of the neutral flows reflects the high sensitivity of the thermosphere to small changes in the underlying or overlying sources of forcing (e.g., gravity waves, pressure changes in the lower atmosphere, solar and magnetospheric forcing, etc.). The mapped flow patterns contain no circulation cells at high latitudes (>60°) like those observed at Earth (*31*). This is most likely due to the lack of a strong dipolar intrinsic magnetic field at Mars. Polar circulation cells are the signatures of strong momentum transfer from convecting ions along magnetic field lines with the neutral gas of the upper atmosphere (*32*). At Earth, this process provides a pathway for energy transfer from the magnetosphere to the thermosphere, especially during geomagnetic disturbances (*33*). The absence of this pathway at Mars most likely prevents the thermosphere from experiencing heating and upwelling at high latitudes from particle precipitation and collisions between ions and neutrals, limiting atmospheric escape.

Our observations constrain the current state of Mars' upper-atmosphere dynamics. The observed global thermospheric circulation must be consistent with the seasonal patterns of the underlying density and temperature structure as well as the short-term (time scale of hours) variations of this underlying structure. The latter may be driven by surface topography and the resulting lower- to upper-atmosphere coupling by means of waves.

## REFERENCES AND NOTES


1. S. W. Bougher, R. G. Roble, T. Fuller-Rowell, in *Atmospheres in the Solar System: Comparative Aeronomy*, M. Mendillo, A. F. Nagy, J. H. Waite, Eds. (American Geophysical Union, vol. 130, 2002), pp. 261–288.
2. D. E. Siskind, S. W. Bougher, in *Heliophysics: Active Stars, Their Astrospheres, and Impacts on Planetary Environments*, C. J Schrijver, F. Bagenal, J. J. Sojka, Eds. (Cambridge Univ. Press, 2016), pp. 201–225.
3. H. G. Mayr, I. Harris, N. W. Spencer, *Rev. Geophys. Space Phys.* **16**, 539–565 (1978).
4. T. L. Killeen, *Rev. Geophys.* **25**, 433–454 (1987).
5. S. W. Bougher, J. M. Bell, J. R. Murphy, M. A. Lopez-Valverde, P. G. Withers, *Geophys. Res. Lett.* **33**, L02203 (2006).
6. S. W. Bougher, T. E. Cravens, J. Grebowsky, J. Luhmann, *Space Sci. Rev.* **195**, 423–456 (2015).
7. D. T. Baird, R. Tolson, S. Bougher, B. Steers, *J. Spacecr. Rockets* **44**, 1180–1187 (2007).
8. J. M. Bell, S. W. Bougher, J. R. Murphy, *J. Geophys. Res.* **112**, E12002 (2007).
9. S. W. Bougher *et al*., *J. Geophys. Res.* **120**, 311–342 (2015).
10. S. W. Bougher, D. A. Brain, J. L. Fox, G.-G. Francisco, C. Simon-Wedlund, P. G. Withers, in *The Atmosphere and Climate of Mars*, R. H. Haberle, R. T. Clancy, F. Forget, M. D. Smith, R. W. Zurek, Eds. (Cambridge Planetary Science Series, Cambridge Univ. Press, 2017), pp. 433–463.
11. F. González-Galindo, F. Forget, M. López-Valverde, M. Angelats i Coll, E. Millour, *J. Geophys. Res.* **114**, E04001 (2009).
12. A. S. Medvedev, E. Yiğit, P. Hartogh, E. Becker, *J. Geophys. Res.* **116**, E10004 (2011).
13. P. R. Mahaffy *et al*., *Space Sci. Rev.* **195**, 49–73 (2014).
14. Materials and methods are available as supplementary materials.
15. R. W. Zurek *et al*., *J. Geophys. Res.* **122**, 3798–3814 (2017).
16. K. J. Roeten *et al*., *J. Geophys. Res. Planets* **124**, 1–21 (2019).
17. S. W. Bougher, S. Engel, R. G. Roble, B. Foster, *J. Geophys. Res.* **104**, 16591–16611 (1999).
18. M. K. Elrod, S. W. Bougher, K. Roeten, R. Sharrar, J. Murphy, *Geophys. Res. Lett.* 2019GL084378 (2019).
19. D. E. Smith *et al*., *J. Geophys. Res.* **106**, 23689–23722 (2001).
20. M. O. Fillingim *et al*., *Icarus* **206**, 112–119 (2010).
21. D. R. Durran, in *Encyclopedia of Atmospheric Sciences* (Academic Press, ed. 2, 2015), pp. 95–102.
22. E. Yiğit, A. S. Medvedev, *J. Geophys. Res.* **122**, 4846–4864 (2017).
23. E. Yiğit *et al*., *Geophys. Res. Lett.* **42**, 8993–9000 (2015).
24. E. Yiğit, A. D. Aylward, A. S. Medvedev, *J. Geophys. Res. Atmos.* **113**, D19106 (2008).
25. E. Yiğit, A. S. Medvedev, P. Hartogh, *Ann. Geophys.* **36**, 1631–1646 (2018).
26. A. S. Medvedev, E. Yiğit, *Atmosphere* **10**, 531 (2019).
27. H. Rishbeth *et al*., *Ann. Geophys.* **18**, 945–956 (2000).
28. S. Bougher *et al*., *Science* **350**, aad0459 (2015).
29. S. W. Bougher *et al*., *J. Geophys. Res.* **122**, 1296–1313 (2017).
30. S. W. Stone, R. V. Yelle, M. Benna, M. K. Elrod, P. R. Mahaffy, *J. Geophys. Res.* **123**, 2842–2867 (2018).
31. T. L. Killeen, R. G. Roble, *Rev. Geophys.* **26**, 329–367 (1988).
32. J. P. Thayer, T. L. Killeen, *Geophys. Res. Lett.* **18**, 701–704 (1991).
33. J. M. Forbes, *J. Meteorol. Soc. Jpn.* **85B**, 193–213 (2007).


## ACKNOWLEDGMENTS


We are grateful for support from all components of the MAVEN project, including the spacecraft, operations, instrument, science, and project teams. Tests and calibrations of the NGIMS instrument were completed at the Planetary Environment Laboratory of NASA's Goddard Space Flight Center. **Funding:** The MAVEN mission is supported by NASA. **Author contributions:** M.B. conceived the measurement technique. M.B., P.R.M., and B.M.J. planned the observations and collected the data. M.B. and Y.L. processed the data. M.B., S.W.B., K.J.R., and Y.L. analyzed and interpreted the global circulation map. M.B., S.W.B., K.J.R., and E.Y. analyzed and interpreted the gravity wave signatures. M.B. wrote the manuscript with input from all co-authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The NGIMS data are available at the Planetary Data System, in the NGIMS bundle at https://atmos.nmsu.edu/data_and_services/atmospheres_data/MAVEN/ngims.html. The orbits and times we analyzed are listed in Data S1. The numerical values of the wind vectors plotted in Fig. 1 are provided in Data S2. The 10 wind data files that went into plotting Figs. 2 and 3 are located in the NGIMS bundle directories /l3/2017/11/ and /l3/2017/12/, respectively.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1363/suppl/DC1
Materials and Methods
Figs. S1 to S7
References (*34*–*38*)
Data S1 and S2


26 February 2019; accepted 28 October 2019
10.1126/science.aax1553

BODY SIZE

# Why whales are big but not bigger: Physiological drivers and ecological limits in the age of ocean giants

J. A. Goldbogen[1]*, D. E. Cade[1], D. M. Wisniewska[1], J. Potvin[2], P. S. Segre[1], M. S. Savoca[1], E. L. Hazen[1,3,4], M. F. Czapanskiy[1], S. R. Kahane-Rapport[1], S. L. DeRuiter[5], S. Gero[6], P. Tønnesen[6], W. T. Gough[1], M. B. Hanson[7], M. M. Holt[7], F. H. Jensen[8], M. Simon[9], A. K. Stimpert[10], P. Arranz[11], D. W. Johnston[12], D. P. Nowacek[13], S. E. Parks[14], F. Visser[15,16,17], A. S. Friedlaender[4], P. L. Tyack[18], P. T. Madsen[6,19], N. D. Pyenson[20,21]

**The largest animals are marine filter feeders, but the underlying mechanism of their large size remains unexplained. We measured feeding performance and prey quality to demonstrate how whale gigantism is driven by the interplay of prey abundance and harvesting mechanisms that increase prey capture rates and energy intake. The foraging efficiency of toothed whales that feed on single prey is constrained by the abundance of large prey, whereas filter-feeding baleen whales seasonally exploit vast swarms of small prey at high efficiencies. Given temporally and spatially aggregated prey, filter feeding provides an evolutionary pathway to extremes in body size that are not available to lineages that must feed on one prey at a time. Maximum size in filter feeders is likely constrained by prey availability across space and time.**

Large body size can improve metabolic and locomotor efficiency. In the oceans, extremely large body size evolved multiple times, especially among edentulous filter feeders that exploit dense patches of small-bodied prey (*1*, *2*). All of these filter feeders had smaller, toothed ancestors that targeted much larger, single prey (*3*, *4*). The ocean has hosted the rise and fall of giant tetrapods since the Triassic, but the largest known animals persist in today's oceans, comprising multiple cetacean lineages (*5*–*8*). The evolution of specialized foraging mechanisms that distinguish the two major whale clades—biosonar-guided foraging on individual prey in toothed whales (Odontoceti) and engulfment filter feeding on prey aggregations in baleen whales (Mysticeti)—likely led to the diversification of crown cetaceans during the Oligocene (~33 to 23 million years ago). The origin of these foraging mechanisms preceded the recent evolution of the largest body sizes (*9*, *10*), and the diversification of these mechanisms across this body size spectrum was likely enhanced by scale-dependent predator-prey processes (*11*). It is hypothesized that toothed whales evolved larger body sizes to enhance diving capacity and exploit deep-sea prey using more powerful biosonar (*12*), whereas baleen whales evolved larger sizes for more efficient exploitation of abundant, but patchily distributed, small-bodied prey (*13*). Cetacean foraging performance is constrained by diving physiology because cetaceans must balance two spatially decoupled resources: oxygen at the sea surface and higher-quality food at depth (*14*). In both lineages, large body size confers an ecological benefit that arises from the scaling of fundamental physiological processes; in some species, anatomical, molecular, and biochemical adaptations further enhance diving capacity (*13*). As animal size increases, mass-specific oxygen storage is constant yet mass-specific oxygen usage decreases (*13*). Therefore, larger air-breathers should have greater diving capacity and thus be capable of feeding for longer periods at a given depth, leading to higher feeding rates overall. In theory, this leads to relatively greater dive-specific energy intake with increasing body size; and, with unlimited prey at the scale of foraging grounds and seasons, larger divers will also exhibit greater energetic efficiencies (i.e., energy intake relative to energy use) while foraging. We hypothesized that the energetic efficiency of foraging will increase with body size because larger animals will have greater diving capacities and more opportunities to feed more frequently per dive. Filter-feeding baleen whales will exhibit relatively higher efficiencies compared with single-prey–feeding toothed whales, because they can exploit greater biomass at lower trophic levels. This study uses whale-borne tag data to provide a comparative test of these fundamental predictions.

Our direct measures of foraging performance using multisensor tags (Fig. 1) show that the largest odontocetes, such as sperm whales (*Physeter macrocephalus*) and beaked whales (Ziphiidae), exhibited high feeding rates during long, deep dives (Fig. 2). By investing time and energy in prolonged dives, these whales accessed deeper habitats that contained less mobile and potentially more abundant prey (*15*), such as weakly muscularized, ammoniacal squid. Conversely, rorqual whales performed fewer feeding events per dive despite their large body size, because they invested large amounts of energy to engulf larger volumes of prey-laden water (*16*). The energetic efficiency ($E_E$, defined as the energy from captured prey divided by the expended energy, including diving costs and postdive recovery) is determined largely by the number of feeding events per dive (Fig. 2) and the amount of energy obtained during each feeding event (Fig. 3). This amount of energy obtained per feeding event was calculated from prey type and size distributions historically found in the stomachs of odontocetes (except for killer whales, for which we used identified prey remains from visually confirmed prey capture events), as well as the acoustically measured biomass, density, and distribution of krill at rorqual foraging hotspots (*17*). Our results show that although larger odontocetes appear to feed on larger prey relative to the prey of smaller, toothed whales, these prey were not disproportionally larger (Fig. 3 and table S11), and toothed whales did feed more frequently on this smaller prey type. Thus, the energy obtained from prey in a dive did not outweigh the increased costs associated with larger body size and deeper dives (fig. S2), thereby causing a decrease in $E_E$ with increasing body size in odontocetes (Fig. 4). In contrast, the measured distribution and density of krill biomass suggests that larger rorquals are not prey-limited at the scale of individual dives. Because larger

[1]Hopkins Marine Station, Department of Biology, Stanford University, Pacific Grove, CA, USA. [2]Department of Physics, Saint Louis University, St. Louis, MO, USA. [3]Environmental Research Division, National Oceanic and Atmospheric Administration, Southwest Fisheries Science Center, Monterey, CA, USA. [4]Institute of Marine Sciences, University of California, Santa Cruz, Santa Cruz, CA, USA. [5]Mathematics and Statistics Department, Calvin University, Grand Rapids, MI, USA. [6]Zoophysiology, Department of Bioscience, Aarhus University, Aarhus, Denmark. [7]Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA. [8]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. [9]Greenland Climate Research Centre, Greenland Institute of Natural Resources, Nuuk, Greenland. [10]Moss Landing Marine Laboratories, Moss Landing, CA, USA. [11]Biodiversity, Marine Ecology and Conservation Group, Department of Animal Biology, University of La Laguna, La Laguna, Spain. [12]Nicholas School of the Environment, Duke University Marine Laboratory, Beaufort, NC, USA. [13]Pratt School of Engineering, Duke University, Durham, NC, USA. [14]Department of Biology, Syracuse University, Syracuse, NY, USA. [15]Department of Freshwater and Marine Ecology, IBED, University of Amsterdam, Amsterdam, Netherlands. [16]Department of Coastal Systems, NIOZ and Utrecht University, Utrecht, Netherlands. [17]Kelp Marine Research, Hoorn, Netherlands. [18]Sea Mammal Research Unit, School of Biology, Scottish Oceans Institute, University of St Andrews, St Andrews, UK. [19]Aarhus Institute of Advanced Studies, Aarhus University, DK-8000 Aarhus C, Denmark. [20]Department of Paleobiology, National Museum of Natural History, Washington, DC, USA. [21]Department of Paleontology and Geology, Burke Museum of Natural History and Culture, Seattle, WA, USA.
*Corresponding author. Email: jergold@stanford.edu

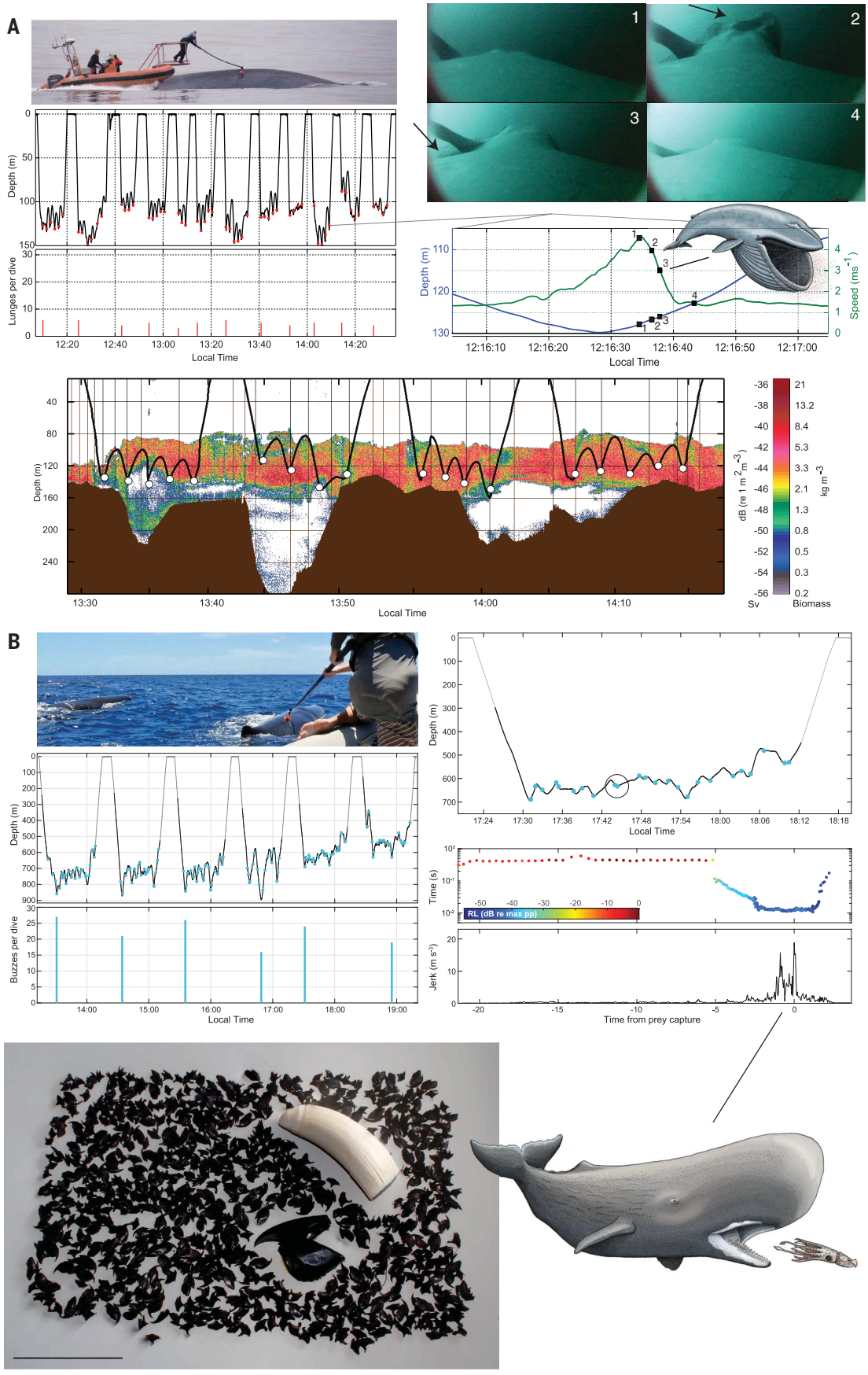



**Fig. 1. Whale tag data quantifies foraging performance.** (**A**) Blue whale suction-cup tagging using a rigid-hulled inflatable boat and a carbon fiber pole (upper left). Tag data from a blue whale showing 12 consecutive foraging dives and the number of lunge-feeding events per dive (left). Inset (right) shows the kinematic signatures used to detect lunge-feeding events (with an increase in speed and upward movement before lunging) and simultaneous video frames that directly confirm engulfment [images 1 to 4: 1, prior to mouth opening; 2, maximum gape (shown by arrow); 3, maximum extension of the ventral groove blubber (shown by arrow); and 4, after mouth closure during the filter phase]. (Bottom) Example of time-synchronized dive profile and the estimated biomass as a function of depth (*17*), grid lines are 147 m by 40 m. Prey mapping data were used to estimate the distribution of krill densities targeted by tagged whales. (**B**) Sperm whale suction-cup tagging (upper left) and six foraging dives with feeding events (thicker lines denote echolocation activity). Middle right panels show the acoustic interclick interval (ICI) and kinematic signatures (jerk, or rate of acceleration) used to infer feeding events at depth. The photograph on the bottom left shows examples of cephalopod beaks (single large beak, *Mesonychoteuthis hamiltoni*; many small beaks, *Gonatus fabricii*) found in the stomachs of sperm whales (lower left) that were used to estimate the size distributions of captured prey (sperm whale tooth and 10 cm line are also shown for scale, photo by Per Henriksen). Illustrations by Alex Boersma.

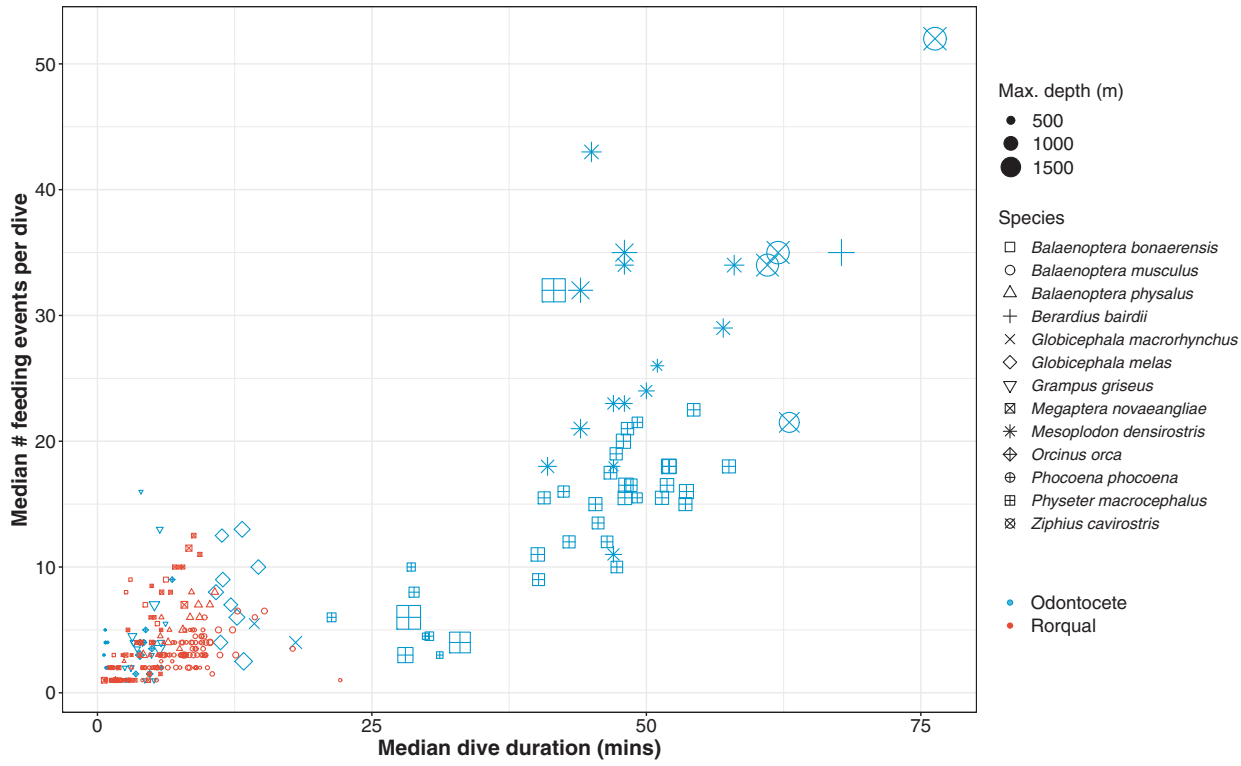



**Fig. 2. Number of feeding events per foraging dive.** Beaked whales (Ziphiidae) and some sperm whales (*P. macrocephalus*) exhibit high feeding rates during long, deep dives, whereas rorquals and delphinids feed less frequently during shorter, shallower dives. Balaenids were excluded from this analysis because they are continuous-ram filter feeders and do not exhibit discrete feeding events like rorquals and odontocetes.

rorquals have relatively larger engulfment capacities (*16*), rorquals exhibited much more rapid increases in energy captured from prey with increasing body size (Fig. 3). If they can detect and exploit the densest parts of an individual krill patch, as evidenced by their ability to maneuver more and increase feeding rates per dive when krill density is higher (*14*), then $E_E$ should increase with body size (Fig. 4). These results were robust to assumptions about trait similarity from shared ancestry as well as the scaling of metabolic rate (MR), which we simulated over a wide range as ($MR \propto Mc^{0.45:0.75}$, where Mc is cetacean body mass).

The divergence in energetic scaling between rorquals and odontocetes that results from available prey has major implications for understanding the ecology and evolution of gigantism in marine ecosystems. For toothed whales, increasing body size leads to hyperallometric investment in biosonar structures that increase prey detection range (*12*). The largest living toothed whales today, sperm whales and beaked whales, independently evolved large body size to push their physiological limits for dive duration to spend more time feeding in the deep sea. The mesopelagic and bathypelagic realms are not only among the largest ecosystems on the planet, they also provide less competitive niches with fewer endothermic predators, providing opportunities to capture high-value prey (*18*). Although sperm whales foraging on giant squids (Architeuthidae) persists as an iconic motif, giant squid beaks are rare in sperm whale stomachs at a global scale (*19*). However, sperm whale biosonar, owing to a hypertrophied nasal complex, is more powerful than beaked whale biosonar by approximately two orders of magnitude (*12*). This allows sperm whales to scan larger volumes of water and, in some regions, to find and chase very large prey. Sperm whales have higher attack speeds and reduced feeding rates per dive when foraging on giant squid (*20*), which contrasts with how sperm whales feed with slower speeds and higher feeding rates on smaller squid in other regions (*21*). This discrepancy suggests that larger prey will incur greater foraging costs, which partially offset the increased energetic gain. Smaller prey are usually more abundant than larger prey (*22*), so efforts to optimize foraging efficiency require the ability to detect the distribution of prey size, which favors the evolution of powerful sonar. Both beaked whales and many sperm whales in our study may have adopted a less risky strategy by targeting more reliable patches of cephalopods often at depths greater than 1000 m, thereby yielding up to 50 feeding events per dive (Fig. 2). Nevertheless, the ability of sperm whales to forage on the largest squid, when available, highlights an advantage of their large size compared with beaked whales, which feed on smaller prey. Regardless of whether odontocetes target a few large prey or many small prey in individual dives, the energy gained from these deep-sea resources is ultimately constrained by the total amount of prey biomass that can be captured during a breath-hold dive. Therefore, prey availability is a key ecological factor that constrains body size and population density in these lineages.

By contrast, gigantism in mysticetes is advantageous because they exhibit positive allometry in filter-feeding adaptations that enable bulk consumption of dense prey patches (*16*). For the largest rorquals, each lunge captured a patch of krill with an integrated biomass and energetic content that exceeded, on average, those of the largest toothed whale prey by at least one order of magnitude (Fig. 3). This ability to process large volumes of prey-laden water, calculated as 100 to 160% of the whale's own body volume in the largest rorquals, underlies the high energetic efficiency of foraging, even when accounting for differences in body size (fig. S1). During lunge feeding, water and prey are engulfed in a matter of seconds and at speeds several times those of steady swimming (*16*). However, whales in a separate mysticete clade (Balaenidae), represented by bowhead whales (*Balaena mysticetus*) and right whales (*Eubalaena* spp.), do not feed in discrete events

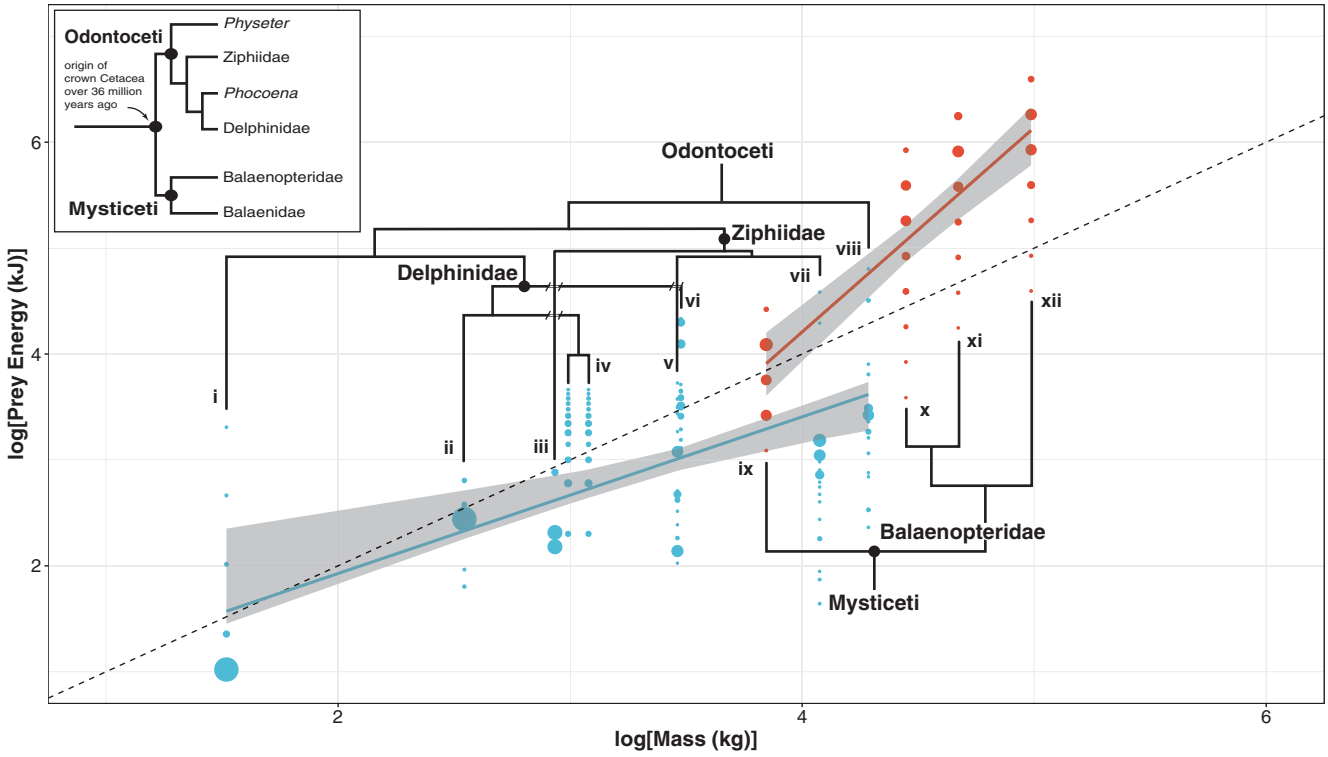



**Fig. 3. Scaling of prey energy captured by toothed whales and rorqual whales during each feeding event.** Estimates for prey energy (prey mass multiplied by prey energy density) obtained from each feeding event. For rorquals, the values indicate the integrated energy of all krill captured for each engulfment event. Symbol size indicates the relative frequency of occurrence based on stomach content data and prey mapping data for odontocetes and mysticetes, respectively. Symbol color is as in Fig. 2. The vertical spread of the data reflects the distribution of prey data for each species. This data was used to weight the regression fitted to species-specific means. The dashed line denotes isometry, indicating that larger toothed whales capture disproportionally less energy from prey ($y = 2.81x^{0.74}$, where $y$ represents energy intake and $x$ represents cetacean body mass), whereas larger rorquals capture disproportionally larger prey energy, with increasing body size ($y = 0.000309x^{1.93}$). Generalized least squares regressions are shown with 95% confidence intervals (CI) (gray bands; see also table S11). The phylogenetic tree inset (with arbitrary branch lengths) shows evolutionary relationships (*32*) among species [(i) harbor porpoise, *Phocoena phocoena*; (ii) Risso's dolphin, *Grampus griseus*; (iii) Blainville's beaked whale, *Mesoplodon densirostris*; (iv) pilot whales, *Globicephala* spp.; (v) Cuvier's beaked whale, *Ziphius cavirostris*; (vi) killer whale, *Orcinus orca*; (vii) Baird's beaked whale, *Berardius bairdii*; (viii) sperm whale, *P. macrocephalus*; (ix) Antarctic minke whale, *Balaenoptera bonaerensis*; (x) humpback whale, *Megaptera novaeangliae*; (xi) fin whale, *Balaenoptera physalus*; (xii) blue whale, *Balaenoptera musculus*]. Balaenids were excluded from this analysis because they are continuous-ram filter feeders and do not exhibit discrete feeding events like rorquals and odontocetes.

but rather continuously ram prey-laden water through their baleen for up to several minutes at a time (*23*). The speed-dependent drag associated with continuous-ram filtration necessitates slow swimming speeds to minimize energy expenditure (*23*). This strategy may be optimized for foraging on smaller copepods that form less dense patches, thereby resulting in lower energetic efficiencies relative to similarly sized rorquals (Fig. 4). The high-speed dynamics of rorqual lunge feeding also generate high drag (*16*), but the rapid engulfment of dense krill patches yields higher efficiencies. Both continuous-ram filter-feeding and lunge-feeding mysticetes appeared to have independently evolved gigantism (>12 m body length) during an era of intensified wind-driven upwelling and glacial cycles, processes that characterize productive whale foraging hotspots in the modern oceans (*9*). Coastal upwelling intensity increases the number and density of aggregations of the relatively small-bodied forage species (*24*) that make filter feeding energetically efficient (*14*). Our analyses point to filter feeding as a mechanism that explains the evolutionary pathway to gigantism because it enabled the high-efficiency exploitation of large, dense patches of prey.

The largest comparable vertebrates, sauropod dinosaurs, reached their maximum size on land about midway through their 140-million-year history, and their evolutionary patterns show no real limits to extreme size (*25*). If sauropod size was not limited by physical factors, such as gravity, hemodynamics, and bone mechanics (*26*), then it may have been ultimately constrained by energetics and food availability (*27*) rather than by an ability to access available food. In the marine environment, the combination of filter feeding and greater abundance of food likely facilitated the evolution of not only gigantic filter-feeding whales, but also that of several independent lineages of large filter-feeding elasmobranchs (*3*, *6*). Both filter-feeding sharks and mesothermic single-prey–feeding sharks exhibit greater body size compared with single-prey–feeding ectothermic sharks (*3*), suggesting parallel evolutionary trajectories with cetaceans in terms of gigantism and morphological adaptations that increase foraging capacity and net energy intake (*4*). The largest filter-feeding sharks are larger than mesothermic raptorial-feeding sharks, which may reflect either a lack of large prey as a limiting factor in today's oceans or an additional temperature-dependent metabolic constraint. Similarly, the larger size of baleen whales compared with filter-feeding sharks suggests an overall advantage for animals that exhibit both endothermy and filter-feeding adaptations, particularly in cold, productive habitats. The combination of high metabolic rates and the ability to short-circuit the food web with filter-feeding adaptations may have enabled high-efficiency exploitation of low trophic levels (*28*), thereby facilitating

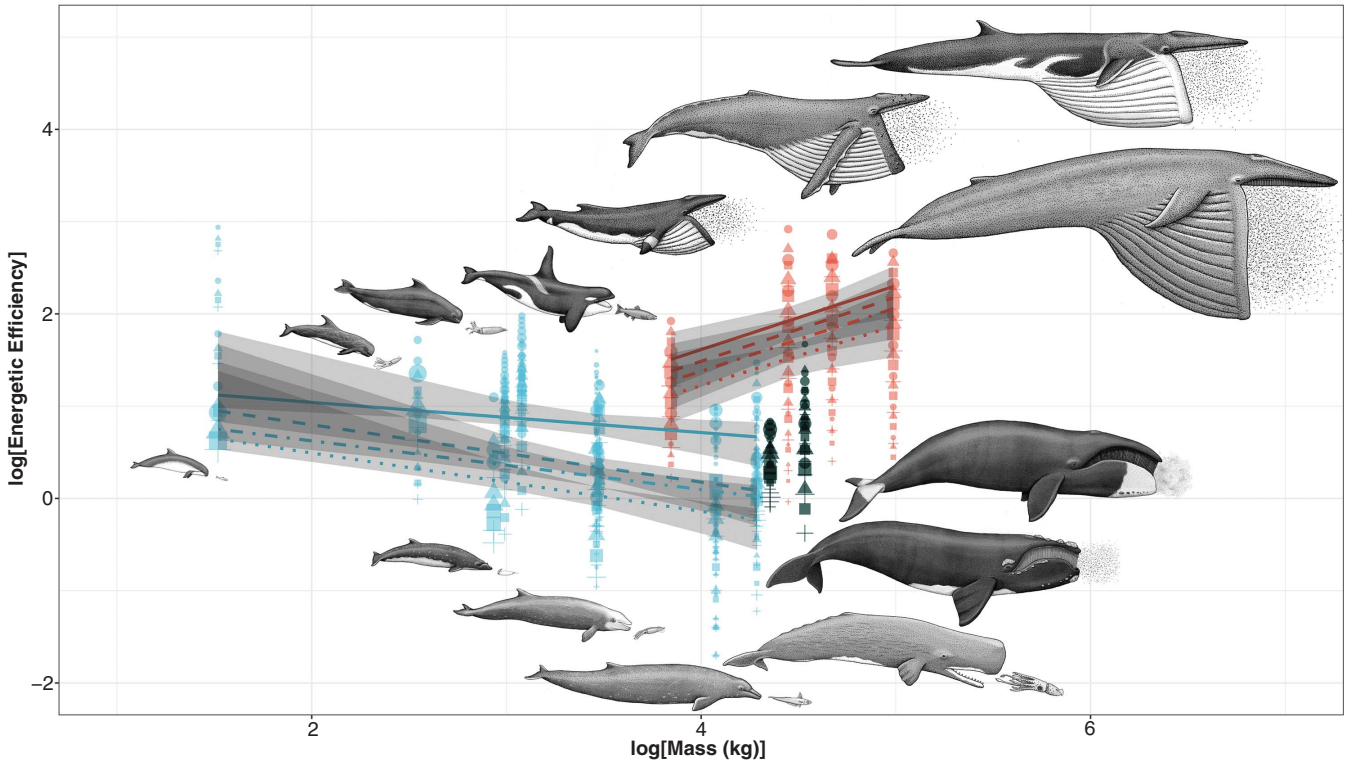



**Fig. 4. Scaling of energetic efficiency for foraging dives and corresponding surface intervals.** The energetic efficiency ($E_E$, defined as the energy from captured prey divided by the expended energy, including diving costs and postdive recovery) of foraging decreases in toothed whales (blue) but increases in rorqual whales lunge filter feeding on krill (red). Bowhead whales and right whales, which continuous-ram filter feed on copepods (green), exhibit lower energetic efficiencies compared with rorqual whales of similar size. These scaling relationships (table S11) are robust to assumptions about metabolic rate (plus symbols and dotted line, MR $\propto$ $Mc^{0.75}$; squares and dot-dash line, MR $\propto$ $Mc^{0.68}$; triangles and dashed line, MR $\propto$ $Mc^{0.61}$; circles and solid line, MR $\propto$ $Mc^{0.45}$) that modulate the rate of energy expenditure of foraging. Regressions are shown with 95% CI (gray bands). The vertical spread of the data corresponds to prey quality distribution data (as in Fig. 3), with larger icons denoting greater proportions of observed values. The vertical spread of the data also reflects the distribution of prey data for each species. Log energetic efficiencies less than zero suggest that whales will be unable to survive on that prey type and quality alone. Illustrations by Alex Boersma.

the evolution of large body size in multiple lineages.

We have shown that cetacean gigantism is driven by the hyperallometry of structures that increase prey capture rates and energy intake in clades with divergent feeding mechanisms, despite the potential constraints to size. However, to maintain a high energetic efficiency at larger sizes, cetaceans must exploit either large individual prey or dense patches of small prey. Although the lack of large prey and the increasing costs of capturing such prey limits energetic efficiency of the largest toothed whales, our analyses suggest that large rorquals are not limited by the size and density of krill patches at the productive apex of their foraging seasons. How long these dense krill patches are available during the summer feeding season at higher latitudes, or throughout the rest of the year (*29*), may ultimately determine the amount of lipid reserves that can be used to fuel ocean basin–scale migrations as well as reproductive output at lower latitudes (*30*, *31*). The size of the largest animals does not seem to be limited by physiology (*5*), but rather is limited by prey availability and the rate at which that prey can be exploited using the foraging mechanisms these whales have evolved.

## REFERENCES AND NOTES


1. G. J. Vermeij, *PLOS ONE* **11**, e0146092 (2016).
2. C. R. McClain *et al.*, *PeerJ* **3**, e715 (2015).
3. C. Pimiento, J. L. Cantalapiedra, K. Shimada, D. J. Field, J. B. Smaers, *Evolution* **73**, 588–599 (2019).
4. M. Friedman, *Proc. R. Soc. B* **279**, 944–951 (2012).
5. W. Gearty, C. R. McClain, J. L. Payne, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4194–4199 (2018).
6. M. Friedman *et al.*, *Science* **327**, 990–993 (2010).
7. N. P. Kelley, N. D. Pyenson, *Science* **348**, aaa3716 (2015).
8. R. E. Fordyce, F. G. Marx, *Curr. Biol.* **28**, 1670–1676.e2 (2018).
9. G. J. Slater, J. A. Goldbogen, N. D. Pyenson, *Proc. R. Soc. B* **284**, 20170546 (2017).
10. O. Lambert *et al.*, *Nature* **466**, 105–108 (2010).
11. P. Domenici, *Comp. Biochem. Physiol. A* **131**, 169–182 (2001).
12. F. H. Jensen, M. Johnson, M. Ladegaard, D. M. Wisniewska, P. T. Madsen, *Curr. Biol.* **28**, 3878–3885.e3 (2018).
13. J. A. Goldbogen, P. T. Madsen, *J. Exp. Biol.* **221**, jeb166033 (2018).
14. E. L. Hazen, A. S. Friedlaender, J. A. Goldbogen, *Sci. Adv.* **1**, e1500469 (2015).
15. K. J. Benoit-Bird, B. L. Southall, M. A. Moline, *Proc. R. Soc. B* **283**, 20152457 (2016).
16. J. A. Goldbogen *et al.*, *Funct. Ecol.* **26**, 216–226 (2012).
17. See supplementary materials.
18. M. Clarke, C. Lu, *J. Mar. Biol. Assoc. U. K.* **55**, 165–182 (1975).
19. M. R. Clarke, *Philos. Trans. R. Soc. Lond. Ser. B* **351**, 1053–1065 (1996).
20. K. Aoki *et al.*, *Mar. Ecol. Prog. Ser.* **444**, 289–301 (2012).
21. A. Fais, M. Johnson, M. Wilson, N. Aguilar Soto, P. T. Madsen, *Sci. Rep.* **6**, 28562 (2016).
22. E. P. White, S. K. M. Ernest, A. J. Kerkhoff, B. J. Enquist, *Trends Ecol. Evol.* **22**, 323–330 (2007).
23. J. Potvin, A. J. Werth, *PLOS ONE* **12**, e0175220 (2017).
24. K. J. Benoit-Bird, C. M. Waluk, J. P. Ryan, *Geophys. Res. Lett.* **46**, 1537–1546 (2019).
25. R. B. Benson, G. Hunt, M. T. Carrano, N. Campione, *Palaeontology* **61**, 13–48 (2018).
26. P. M. Sander *et al.*, *Biol. Rev. Camb. Philos. Soc.* **86**, 117–155 (2011).
27. G. P. Burness, J. Diamond, T. Flannery, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14518–14523 (2001).
28. M. A. Tucker, T. L. Rogers, *Proc. R. Soc. B* **281**, 20142103 (2014).
29. C. K. Geijer, G. Notarbartolo di Sciara, S. Panigada, *Mammal Rev.* **46**, 284–296 (2016).
30. E. Pirotta *et al.*, *Am. Nat.* **191**, E40–E56 (2018).
31. R. Williams *et al.*, *ICES J. Mar. Sci.* **70**, 1273–1280 (2013).
32. J. H. Geisler, M. R. McGowen, G. Yang, J. Gatesy, *BMC Evol. Biol.* **11**, 112 (2011).


## ACKNOWLEDGMENTS

We thank C. Taylor for Echoview processing and active acoustic data collection, A. Boersma for providing illustrations of cetacean species and their prey, and P. Henriksen for the photograph of squid beaks and the sperm whale tooth. We also thank W. Gearty for assistance with comparative phylogenetic analyses. **Funding:** This research was funded in part by grants from the National Science Foundation (IOS-1656676 and IOS-1656656; OPP-1644209 and 07-39483); the Office of Naval Research (N000141612477); and a Terman Fellowship from Stanford University. All procedures in the United States were conducted under approval of the National Marine Fisheries Service (permits

781-1824, 16163, 14809, 16111, 19116, 15271, and 20430); Canada DFO SARA/MML 2010-01/SARA-106B; National Marine Sanctuaries (MULTI-2017-007); Antarctic Conservation Act (2009-014 and 2015-011); and institutional IACUC committee protocols. Fieldwork, data collection, and data processing for *M. densirostris* were funded by the Office of Naval Research grants N00014-07-10988, N00014-07-11023, N00014-08-10990, N00014-18-1-2062, and 00014-15-1-2553, and the U.S. Strategic Environmental Research and Development Program Grant SI-1539. P.L.T. gratefully acknowledges funding from the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland). MASTS is funded by the Scottish Funding Council (HR09011) and contributing institutions. Fieldwork, data collection, and data processing for *Globicephala melas* and data collection at the Azores were funded by the Office of Naval Research (ONR grants N00014-12-1-0410, N000141210417, N00014-15-1-2341, and N00014-17-1-2715); the Danish Council for Independent Research (award number 0602-02271B); the Dutch Research Council (award number 016.Veni.181.086); and a Semper Ardens Grant from the Carlsberg Foundation. For SRKW field work, we thank the NOAA Ocean Acoustics Program for providing funding and C. Emmons, D. Giles, and J. Hogan for assistance in the field. For *P. macrocephalus*, fieldwork off Dominica was supported through a FNU fellowship from the Danish Council for Independent Research, supplemented by a Sapere Aude Research Talent Award, a Carlsberg Foundation expedition grant, a grant from Focused on Nature, and a CRE Grant from the National Geographic Society to S.G.; a FNU large frame grant; as well as a Villum Foundation Grant (to P.T.M.) with supplementary grants from Dansk Akustisk Selskab (to P.T.), Oticon Foundation (to P.T.), and Dansk Tennis Foundation (to P.T.). The Greenland data collection and analysis were funded by grants from the Oticon Foundation and the Carlsberg Foundation to M.S. Tagging work on *P. phocoena* was funded in part by the German Federal Agency for Nature Conservation (BfN) under contract Z1.2-530/2010/14 and the BfN-Cluster 7 "Effects of underwater noise on marine vertebrates." **Author contributions:** Overall idea, concept, and approach developed by J.A.G. Bioenergetic models developed by J.P. and implemented by J.A.G. Integration of data analysis and interpretation directed by J.A.G. Data analysis conducted and implemented by J.A.G., D.E.C., D.M.W., J.P., and P.S.S., with statistical contributions by D.M.W., S.L.D., M.S.S., and E.L.H.. Manuscript written by J.A.G. and N.D.P. with contributions by D.E.C., D.M.W., J.P., P.L.T., P.T.M., and F.H.J. All authors read, edited, and discussed the manuscript and participated in data collection. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data analyzed in this study will be available at the Stanford Data Repository (sdr.stanford.edu) immediately upon publication at https://purl.stanford.edu/zk778rt5347.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1367/suppl/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S11
References (*33*–*185*)

View/request a protocol for this paper from *Bio-protocol*.

3 May 2019; accepted 31 October 2019
10.1126/science.aax9044

GLYCOBIOLOGY

# Cryo–electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B

Ana S. Ramírez*, Julia Kowal*, Kaspar P. Locher†

**Oligosaccharyltransferase (OST) catalyzes the transfer of a high-mannose glycan onto secretory proteins in the endoplasmic reticulum. Mammals express two distinct OST complexes that act in a cotranslational (OST-A) or posttranslocational (OST-B) manner. Here, we present high-resolution cryo–electron microscopy structures of human OST-A and OST-B. Although they have similar overall architectures, structural differences in the catalytic subunits STT3A and STT3B facilitate contacts to distinct OST subunits, DC2 in OST-A and MAGT1 in OST-B. In OST-A, interactions with TMEM258 and STT3A allow ribophorin-I to form a four-helix bundle that can bind to a translating ribosome, whereas the equivalent region is disordered in OST-B. We observed an acceptor peptide and dolichylphosphate bound to STT3B, but only dolichylphosphate in STT3A, suggesting distinct affinities of the two OST complexes for protein substrates.**

N-glycosylation is one of the most prevalent posttranslational modifications of secretory proteins (*1*, *2*). It is estimated that >10% of human proteins carry N-glycans (*3*), which are essential for intracellular processes such as protein folding and trafficking, and for extracellular recognition and signaling (*1*, *2*). The process is initiated in the endoplasmic reticulum (ER) with the transfer of the preassembled, high-mannose oligosaccharide $GlcNac_2$-$Man_9$-$Glc_3$ from a dolichyl-pyrophosphate carrier to asparagine residues in the glycosylation sequon Asn-X-Thr/Ser (*4*). This reaction is catalyzed by oligosaccharyltransferase (OST), a multisubunit complex located in the ER membrane (*1*, *2*, *5*). Unlike yeast, human cells contain two distinct OST complexes, OST-A and OST-B (*2*), and their combined functions result in a glycoproteome roughly 10× larger than that of yeast (*6*). The OST-A and OST-B complexes share six subunits (ribophorin-I, ribophorin-II, OST48, TMEM258, DAD1, and OST4) (*7*–*9*) but feature distinct paralogous catalytic subunits, STT3A (in OST-A) or STT3B (in OST-B) (*10*). In addition, OST-A contains the adaptor protein DC2, which mediates the association with the protein translocation channel Sec61 (*11*, *12*). When bound to translating, membrane-associated ribosomes and the translocon-associated protein complex, a super-complex is formed that couples the synthesis, translocation into the ER, and N-glycosylation of secretory proteins (*11*, *12*). In contrast, OST-B does not contain DC2 but instead contains either MAGT1 or TUSC3 (OST3 or OST6 in yeast, respectively) (*13*, *14*), two paralogous redox chaperones. These allow OST-B to catalyze posttranslocational processing of acceptor sites even near folded protein elements or formed disulfide bridges (*14*–*16*).

High-resolution cryo–electron microscopy (cryo-EM) structures of yeast OST have provided insight into the architecture of multimeric OST systems (*12*, *17*, *18*), and a medium-resolution cryo-EM structure visualized how OST-A associates with the translocon and the ribosome. However, the structural basis of the distinct functions of human OST-A and OST-B requires high-resolution structures of both complexes. Whereas bacterial and archaeal single-subunit OST enzymes, including PglB and AglB, have been visualized in the presence of substrates (*19*–*21*), no structures of eukaryotic OST complexes have been determined in substrate-bound states. To tackle these questions, we separately purified human OST-A and OST-B from human embryonic kidney cells by expressing affinity-tagged versions of DC2 (for OST-A) or MAGT1 (for OST-B) (fig. S1). The purified OST complexes were homogeneous and functional, as shown by an in vitro glycosylation assay using a fluorescently labeled peptide and a synthetic, lipid-linked oligosaccharide (LLO) analog containing a disaccharide ($GlcNAc_2$) moiety (Fig. 1). Intriguingly, we found that glycan transfer catalyzed by OST-B

Institute of Molecular Biology and Biophysics, Eidgenössische Technische Hochschule (ETH), CH-8093 Zürich, Switzerland.
*These authors contributed equally to this work.
†Corresponding author. Email: locher@mol.biol.ethz.ch

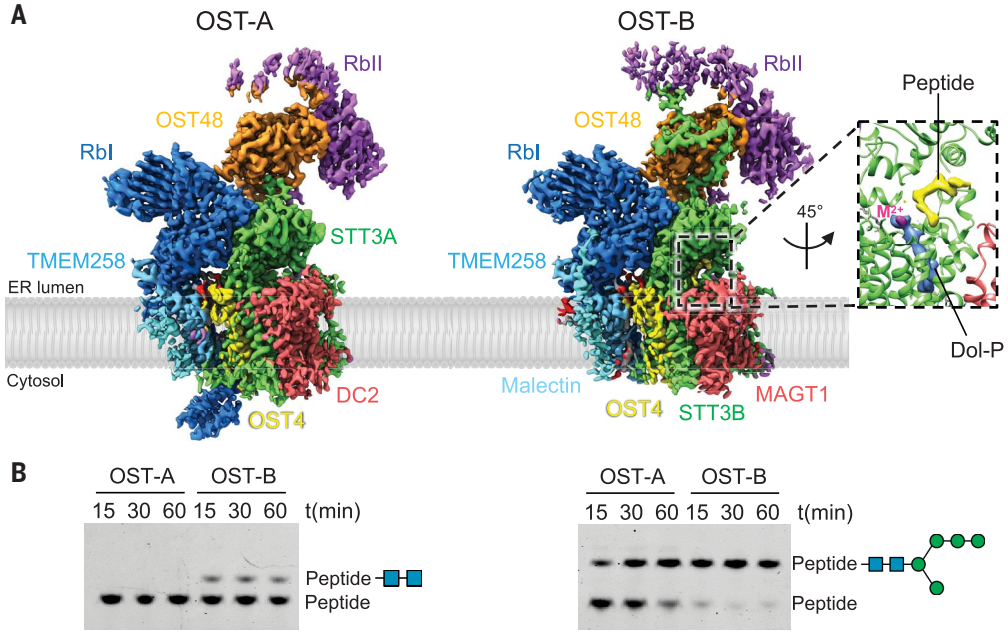



**Fig. 1. Biochemical and cryo-EM studies of the human OST complexes.** (**A**) Cryo-EM maps of OST-A (left) and OST-B (right) are shown and colored according to the subunits, as indicated. Inset on the right shows the EM density for ligands bound to STT3B. In this view, DAD1 is located behind STT3A and STT3B. RbI, ribophorin-I; RbII, ribophorin-II. (**B**) In vitro glycosylation assays with purified OST-A and OST-B complexes. Fluorescently labeled peptide TAMRA-GSGNSTVT was used as the acceptor substrate, and farnesyl-citronellyl-PP-$GlcNAc_2$ or farnesyl-citronellyl-PP-$GlcNAc_2Man_5$ as the donor substrate. Glycosylated and nonglycosylated peptides (top and bottom rows, respectively) were separated using SDS–polyacrylamide gel electrophoresis Tricine gels.

was faster than that catalyzed by OST-A. Furthermore, the reactions proceeded more efficiently when a chemoenzymatically extended LLO analog carrying five additional mannoses ($GlcNAc_2Man_5$) (*22*, *23*) was transferred, suggesting that the mannose units increase the affinity of both human OST complexes for the LLO.

Single-particle cryo-EM analysis of the purified OST-A and OST-B complexes resulted in three-dimensional reconstructions with overall resolutions of 3.5 Å in both cases (Fig. 1). The local resolution of the transmembrane (TM) domains and of the central parts of the luminal regions was substantially better and allowed unambiguous building of atomic models (figs. S2 to S5). Despite the low sequence similarity between human and yeast OST subunits (between 30 and 50%) (table S1), the overall architectures of OST-A and OST-B resemble that of the yeast enzyme (root mean square deviation: $RMSD_{OST\text{-}A\text{-}yOST}$ = 2.75 Å for 1660 Cα atoms and $RMSD_{OST\text{-}B\text{-}yOST}$ = 1.99 Å for 1730 Cα atoms) (*17*, *18*). Thus, similar subcomplexes can be identified: Subcomplex I consists of ribophorin-I and TMEM258 and is identical in OST-A and OST-B. Subcomplex II is distinct in the two OST versions: In OST-A, it consists of OST4, the catalytic STT3A subunit, and the translocon adaptor protein DC2. In OST-B, it features OST4, the catalytic STT3B subunit, and the redox chaperone MAGT1 (fig. S6). Although Western blots demonstrate that full-length MAGT1 was present in our OST-B preparations (Fig. 2A), the EM map only revealed clear densities for TM2 to TM4, which showed a remarkable structural similarity to DC2 (Fig. 2B). Finally, subcomplex III consists of DAD1, OST48, and ribophorin-II (fig. S6). Human ribophorin-II contains an N-terminal extension of ~300 residues compared with its yeast homolog SWP1. Whereas this region is highly flexible in OST-A, it is visible in the EM map of OST-B. This difference is likely attributable to interactions with a C-terminal extension of STT3B that is absent in STT3A. As a consequence of this stabilization, ribophorin-II displays two main conformations in OST-B (fig. S7 and movie S1). In the TM interfaces between the three subcomplexes, we observed several ordered phospholipids and digitonin molecules. Ordered N-glycans on the surface of OST were observed for the catalytic subunits STT3A (N537 and N548) and STT3B (N616, N627, and N641), and for ribophorin-I (N299) (figs. S8 and S9).

Various additional proteins have been proposed to interact with mammalian OST. We found strong density consistent with a TM helix and a short luminal stretch of malectin in close proximity to the TM domain of TMEM258 and the luminal domain of ribophorin-I. Western blot analysis of purified OST-A and OST-B complexes revealed a strong signal for full-length malectin in OST-B but a lower signal in OST-A preparations, suggesting that malectin is bound in substoichiometric amounts in OST-A (fig. S10). Although malectin was proposed to have a role in quality control during the folding of glycoproteins, its precise molecular mechanism remains to be clarified (*24*). The observation that malectin is bound and partially ordered illustrates that human OST complexes can serve as platforms for interactions that help modulate protein N-linked glycosylation by controlling attached glycans and the status of glycoprotein folding.

The pronounced similarity between human OST-A and OST-B (RMSD: 0.96 Å for 1715 Cα atoms) suggests that subtle structural differences are responsible for their specific subunit interactions and their distinct cellular functions. We noted two key differences: First, despite the similar arrangement of transmembrane helices TM10–13, there are differences in the protein surfaces where STT3A and STT3B interact with DC2 or MAGT1 (Fig. 2B). A detailed analysis revealed that MAGT1 could not associate with STT3A, nor could DC2 associate with STT3B, owing to steric clashes (fig. S11). As a consequence, only OST-A can interact with the translocon. The second structural difference is that, even in the absence of bound ribosome, the C-terminal domain of ribophorin-I forms a cytoplasmic four-helix bundle in OST-A.

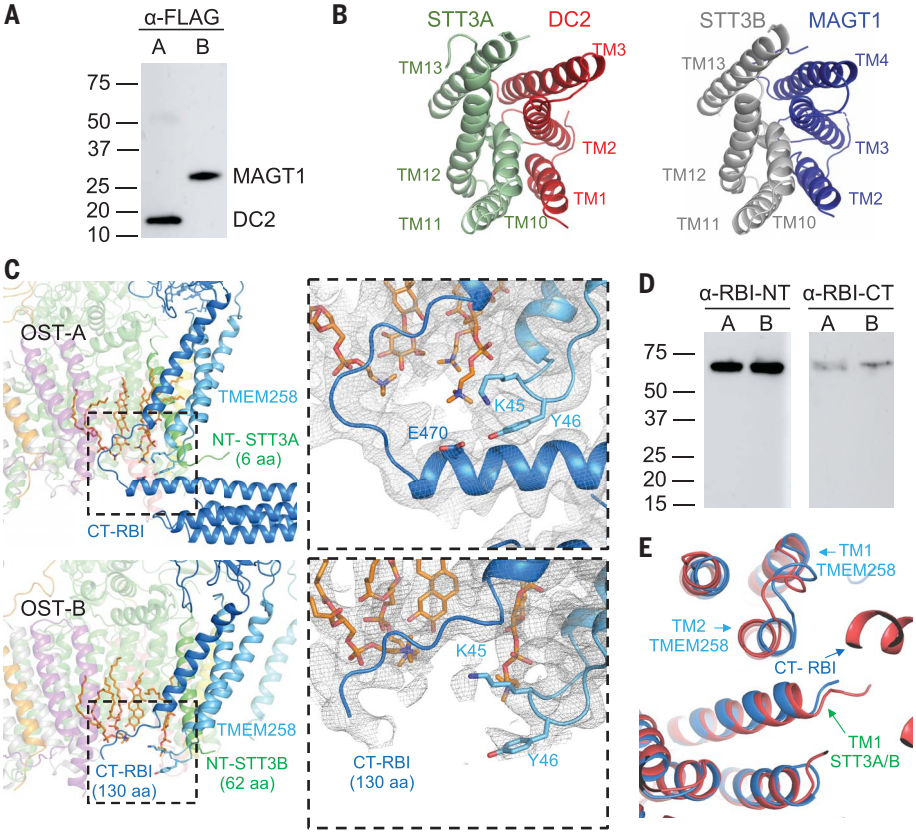



**Fig. 2. Interaction of STT3 paralogs with their neighbor subunits.** (**A**) Western blot analysis of purified OST-A and OST-B using anti-Flag antibody to detect Flag-tagged DC2 and MAGT1. (**B**) Cytosolic view onto the TM regions of STT3A and DC2 (left) and STT3B and MAGT1 (right). (**C**) Structural arrangement of ribophorin-I, TMEM258, and STT3 in OST-A (top) and OST-B (bottom). Insets show closed-up views of the interaction between ribophorin-I and TMEM258. The residues involved in contacts are shown as sticks and labeled with single-letter amino acid abbreviations (E, Glu; K, Lys; Y, Tyr). EM density is shown as a gray mesh (map levels: 4.0σ, as defined by PyMOL). Subunits are colored as in Fig. 1. CT, C-terminal region; NT, N-terminal region; aa, amino acids. (**D**) Western blot analysis of purified OST-A and OST-B using antibodies against the N- or C-terminal regions of ribophorin-I (RBI). (**E**) Cytoplasmic view onto the superimposed structures of OST-A (red) and OST-B (blue), focusing on the TM helices of ribophorin-I, TMEM258, and TM1 of the STT3 subunits.

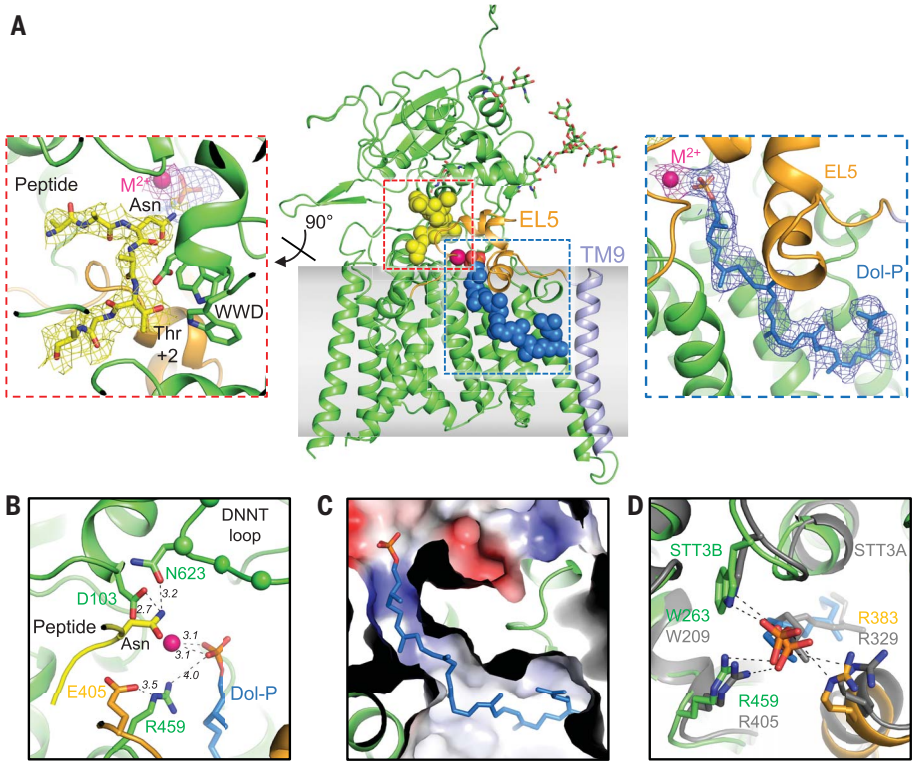



**Fig. 3. Active site of human OST.** (**A**) Structure of STT3B shown in cartoon representation [EL5, orange; transmembrane helix 9 (TM9), light blue]. Close-up of binding sites for bound peptide (left inset) and DolP (right inset). Bound peptide is shown as yellow sticks. Residues interacting with +2 position (WWD motif) are shown as sticks and labeled. DolP is shown as sticks with carbon atoms colored blue, and the divalent metal ion is shown as a pink sphere. The EM density covering bound substrates is shown as a mesh and colored according to substrate color (map level: 5.0σ, as defined by PyMOL). (**B**) Close-up of active site of STT3B. Residues presumably involved in catalysis are shown as sticks and labeled (D, Asp; N, Asn; R, Arg). Bound peptide is shown as cartoon, and the acceptor asparagine as sticks. Divalent ion is shown as a pink sphere. (**C**) Electrostatic surface potential of the LLO binding site, color coded by electrostatic potential ranging from blue (most positive) to red (most negative). DolP is shown as in (A). (**D**) LLO binding site, as viewed from the ER lumen. STT3B residues interacting with the phosphate group of DolP are shown as sticks, colored as in (A) and labeled (W, Trp). A superposition of STT3A (gray) with bound DolP is shown.

This bundle was found to be in contact with translating ribosome in a recent tomography study (*12*). In contrast, no such feature was visible in the EM of OST-B (Fig. 2C). To rule out that a truncation of ribophorin-I had occurred in our OST-B preparation, we conducted Western blot analysis using antibodies recognizing the C- or N-terminal regions of ribophorin-I. This analysis revealed that full-length ribophorin-I was present in both OST complexes (Fig. 2D), suggesting that the C-terminal region of ribophorin-I is disordered in OST-B. Our high-resolution structures provide an explanation for this key difference: In OST-A, the first helix of the C-terminal region of ribophorin-I interacts with the cytoplasmic loop connecting TM1 and TM2 of TMEM258. Specifically, $Glu^{470}$ of ribophorin-I forms a hydrogen bond with the side chain of $Tyr^{46}$ of TMEM258, which is simultaneously involved in a cation–π interaction with its neighboring residue $Lys^{45}$ (Fig. 2C). In contrast, such an interaction is absent in OST-B, where TM1 of STT3B and TM2 of TMEM258 are tilted compared with their position in OST-A (Fig. 2E). Furthermore, the cytoplasmic N-terminal segment of STT3B is *56* residues longer than that of STT3A. Given that this region is in the immediate vicinity of the C-terminal region of ribophorin-I, it is more likely than the shorter segment of STT3A to interfere with the folding of the four-helix bundle (Fig. 2C).

The catalytic subunits STT3A and STT3B form similar active sites. Previously reported structures of yeast OST were visualized in apo states (*17*, *18*), but we observed well-defined density for two ligands bound to STT3B (Fig. 3A and fig. S12). Furthermore, whereas TM9 and external loop EL5 (connecting TM9 and TM10) were disordered in the yeast apo-OST structures, both adopted an "engaged" conformation in human OST-B, reminiscent of the conformational changes observed in bacterial PglB upon binding to substrates (*19*, *21*). It is of interest that, with the exception of divalent metal ions present in the purification buffers, no cofactors, substrates, or substrate analogs were added to our OST-B preparations. The bound substrate analogs had thus been retained from the native cellular environment during purification, which suggests relatively high affinities. The density in the peptide-binding pocket could be interpreted as an average of polypeptides containing an N-glycosylation sequon. As was observed in PglB, the polypeptide backbone formed a 180° turn (Figs. 1A and 3A), suggesting that only unfolded polypeptide regions can be processed by OST-B. The side chain of the +2 threonine is in close proximity to the conserved WWD motif, which has been demonstrated to be essential for sequon recognition (*25*). The side chain of the acceptor asparagine is located in a tunnel formed by the packing of EL5 against the ER-luminal domain of STT3B and is in close proximity to the catalytic $Asp^{103}$ residue and the side chain of $Asn^{623}$, which is part of the conserved DNNT loop (Fig. 3B) (*19*, *21*, *26*). We further observed density consistent with a bound divalent metal ion in the active site (Fig. 3A) in the vicinity of conserved, negatively charged side chains including $Asp^{103}$, as well as residues $Asp^{167}$ and $Glu^{169}$ of the DxE motif (*19*, *21*, *25*).

A hydrophobic groove formed by TM6 and TM11 is present on the surface of STT3B (Fig. 3C), similar to grooves observed in PglB and the apo structures of yeast OST (*17*–*19*, *25*). In human OST-B, this groove contained well-defined density of a bound molecule featuring the characteristic shape of a polyprenyl chain, as well as strong density for a phosphate group in contact with the divalent ion and a conserved arginine side chain ($Arg^{459}$). However, no density was visible for a second phosphate group or for ordered glycans, ruling out the possibility that a functional LLO, containing a pyrophosphate group and a glycan, was bound. We therefore interpreted the observed density as a bound dolichylphosphate (DolP) molecule. Although surprising, DolP was recently shown to copurify with the ER-localized O-mannosyltransferase complex PMT1-PMT2 (*27*). In our OST-B structure, the phosphate group of bound DolP forms hydrogen bonds with $Arg^{459}$, $Arg^{383}$, and $Trp^{263}$ of the STT3B subunit (Fig. 3D). $Arg^{459}$ is a conserved residue in OST enzymes that appears to stabilize the leaving group DolPP during the nucleophilic substitution reaction (*25*). We found that the guanidinium group of $Arg^{459}$ was within hydrogen bonding distance of $Glu^{403}$, a catalytic residue located within EL5, and shown to be essential for activity both in bacterial PglB and yeast OST (*17*, *28*).

In contrast to OST-B, no density for a bound peptide was observed in OST-A. However, density for a bound DolP molecule was clearly visible in the LLO-binding groove. The interactions of bound DolP with STT3A are very similar to those observed in STT3B, suggesting a conserved mechanism for LLO recognition in both complexes (Fig. 3D). Despite the absence of bound peptide, EL5 of STT3A was fully ordered and adopted a similar conformation as in STT3B (fig. S12), demonstrating that DolP binding is sufficient for EL5 to adopt an engaged conformation. This is notably different than in bacterial PglB, where an equivalent conformational change requires the binding of both LLO and acceptor peptide (*19*, *21*). The absence of bound peptide in our OST-A structure suggests that OST-B has a higher affinity for acceptor peptides.

Our structural observations correlate well with the distinct cellular functions of OST-A and OST-B: OST-A is responsible for processing of most of the N-glycosylation sites. Given its proximity to the translocon, OST-A is exposed to the unfolded polypeptide sequences of all secretory proteins and therefore able to process even suboptimal sequons (*29*). In contrast, OST-B processes proteins that are already partially folded and may even contain disulfide bridges. This requires OST-B to have substantially stronger affinity for glycosylation sequons than does OST-A, which is in line with our structural observation that copurified peptides were bound to the OST-B complex. Our finding that purified OST-A had a lower in vitro activity than OST-B is both intriguing and counterintuitive. The challenge of OST-A is to keep up with protein translation and translocation and prevent the protein folding in the ER that removes target glycosylation sites. One would thus expect OST-A to be a faster enzyme than OST-B. Given that the opposite is observed in vitro, it is possible that, in the cell, interactions with the translocon and the translating ribosome may stimulate OST-A and thus increase its catalytic rate, either through altered protein dynamics or by inducing conformational changes.

Our OST-A and OST-B structures provide insight into substrate binding of eukaryotic OST complexes and establish that the distinct functions of the two human complexes are based on structural differences of their catalytic subunits STT3A and STT3B, which result in interactions with distinct subunits and different affinities for acceptor peptides. These structures provide a basis for the design of inhibitors of N-glycosylation that modulate the maturation and activation of protein markers involved in tumor formation (*30*).

## REFERENCES AND NOTES


1. J. Breitling, M. Aebi, *Cold Spring Harb. Perspect. Biol.* **5**, a013359 (2013).
2. N. Cherepanova, S. Shrimal, R. Gilmore, *Curr. Opin. Cell Biol.* **41**, 57–65 (2016).
3. R. Ganetzky, F. J. Reynoso, M. He, in *Biomarkers in Inborn Errors of Metabolism*, U. Garg, L. D. Smith, Eds. (Elsevier, 2017), pp. 343–360.
4. R. G. Spiro, *Adv. Protein Chem.* **27**, 349–467 (1973).
5. D. F. Zielinska, F. Gnad, J. R. Wiśniewski, M. Mann, *Cell* **141**, 897–907 (2010).
6. M. Aebi, *Biochim. Biophys. Acta* **1833**, 2430–2437 (2013).
7. D. J. Kelleher, R. Gilmore, *Glycobiology* **16**, 47R–62R (2006).
8. A. Dumax-Vorzet, P. Roboti, S. High, *J. Cell Sci.* **126**, 2595–2606 (2013).
9. D. B. Graham *et al.*, *Cell Rep.* **17**, 2955–2965 (2016).
10. D. J. Kelleher, D. Karaoglu, E. C. Mandon, R. Gilmore, *Mol. Cell* **12**, 101–111 (2003).
11. S. Shrimal, N. A. Cherepanova, R. Gilmore, *J. Cell Biol.* **216**, 3625–3638 (2017).
12. K. Braunger *et al.*, *Science* **360**, 215–219 (2018).
13. B. L. Schulz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11061–11066 (2009).
14. N. A. Cherepanova, S. Shrimal, R. Gilmore, *J. Cell Biol.* **206**, 525–539 (2014).
15. S. Shrimal, S. F. Trueman, R. Gilmore, *J. Cell Biol.* **201**, 81–95 (2013).
16. S. Shrimal, R. Gilmore, *J. Cell Sci.* **126**, 5513–5523 (2013).
17. R. Wild *et al.*, *Science* **359**, 545–550 (2018).
18. L. Bai, T. Wang, G. Zhao, A. Kovach, H. Li, *Nature* **555**, 328–333 (2018).
19. M. Napiórkowska *et al.*, *Nat. Struct. Mol. Biol.* **24**, 1100–1106 (2017).
20. S. Matsumoto, Y. Taguchi, A. Shimada, M. Igura, D. Kohda, *Biochemistry* **56**, 602–611 (2017).
21. M. Napiórkowska, J. Boilevin, T. Darbre, J.-L. Reymond, K. P. Locher, *Sci. Rep.* **8**, 16297 (2018).
22. A. S. Ramírez *et al.*, *Glycobiology* **27**, 525–535 (2017).
23. A. S. Ramírez *et al.*, *Glycobiology* **27**, 1–8 (2017).
24. K. Takeda, S. Y. Qin, N. Matsumoto, K. Yamamoto, *Biochem. Biophys. Res. Commun.* **454**, 436–440 (2014).
25. C. Lizak, S. Gerber, S. Numao, M. Aebi, K. P. Locher, *Nature* **474**, 350–355 (2011).
26. M. Igura *et al.*, *EMBO J.* **27**, 234–243 (2008).
27. L. Bai, A. Kovach, Q. You, A. Kenny, H. Li, *Nat. Struct. Mol. Biol.* **26**, 704–711 (2019).
28. S. Gerber *et al.*, *J. Biol. Chem.* **288**, 8849–8861 (2013).
29. N. A. Cherepanova, S. V. Venev, J. D. Leszyk, S. A. Shaffer, R. Gilmore, *J. Cell Biol.* **218**, 2782–2796 (2019).
30. A. S. Puschnik *et al.*, *Cell Rep.* **21**, 3032–3039 (2017).


## ACKNOWLEDGMENTS

We thank the staff at the Scientific Center for Optical and Electron Microscopy (ScopeM, ETH Zurich, Switzerland) for support during cryo-EM data collection; J. Zürcher for help with protein expression and cell culture work; and J. Boilevin and J.-L. Reymond for providing the LLO analog farnesyl-citronellyl-PP-GlcNAc$_2$. We thank M. Aebi, T. Darbre, and J.-L. Reymond for helpful discussions. **Funding:** This work was supported by the Swiss National Science Foundation (Transglyco and GlycoSTART Sinergia program funds to K.P.L.). **Author contributions:** A.S.R. and K.P.L. conceived of the project and designed the experiments. A.S.R. performed cloning, preparation of stable cell lines, and expression, purification, and biochemical characterization of OST complexes. J.K. prepared cryo-EM grids and collected and processed EM data. A.S.R. built, refined, and validated the structures with help from K.P.L. All authors contributed to the writing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Atomic coordinates of the OST-A and OST-B models were deposited in the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) under IDs 6S7O for OST-A and 6S7T for OST-B. The three-dimensional cryo-EM maps were deposited in the Electron Microscopy Data Bank (EMDB) under IDs EMD-10110 for OST-A and EMD-10112 for OST-B. All other data are presented in the paper and supplementary materials.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1372/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 and S2
References (*31*–*44*)
Movie S1
View/request a protocol for this paper from *Bio-protocol*.

2 September 2019; accepted 13 November 2019
10.1126/science.aaz3505

POLYMERS

# High strength in combination with high toughness in robust and sustainable polymeric materials

Xiaojian Liao[1], Martin Dulle[2], Juliana Martins de Souza e Silva[3], Ralf B. Wehrspohn[3,4], Seema Agarwal[1], Stephan Förster[2,5], Haoqing Hou[6], Paul Smith[7], Andreas Greiner[1]*

In materials science, there is an intrinsic conflict between high strength and high toughness, which can be resolved for different materials only through the use of innovative design principles. Advanced materials must be highly resistant to both deformation and fracture. We overcome this conflict in man-made polymer fibers and show multifibrillar polyacrylonitrile yarn with a toughness of 137 ± 21 joules per gram in combination with a tensile strength of 1236 ± 40 megapascals. The nearly perfect uniaxial orientation of the fibrils, annealing under tension in the presence of linking molecules, is essential for the yarn's notable mechanical properties. This underlying principle can be used to create similar strong and tough fibers from other commodity polymers in the future and can be used in a variety of applications in areas such as biomedicine, satellite technology, textiles, aircrafts, and automobiles.

Synthetic materials with a combination of high strength and high toughness are rare and represent an important technological challenge (*1*). Engineered metal alloys (*2*), metallic glass composites (*3*), nanocellulose paper (*4*), glass (*5*), and carbon-based materials (*6*–*9*) are some examples of materials with this property combination. High resistance to both deformation (high strength) and fracture (high toughness) is achieved in man-made single-polymer nanofibers of very small diameter by electrospinning (*10*–*14*). However, these single nanofibers are not robust enough for handling for real-world applications. Natural fibers, like dragline spider silks (*15*, *16*) and recombinant spider silks (*17*), achieve the combination of high strength and high toughness as well, but their applicability is restricted by either low availability or high prices for various applications.

We discovered a straightforward concept for combination of high strength and high toughness through the preparation of polymeric fibers by yarn electrospinning, which creates fibers consisting of thousands of aligned nanofibrils in combination with a specified amount of a linker molecule. Simple alignment of the nanofibrils in electrospun yarns in combination with a high degree of crystallization does not result in high toughness in conjunction with high strength. Rather, this combination is achievable through the addition of a small amount of an interconnecting molecule during yarn electrospinning and annealing after the heat stretching of the nanofibrils.

We fabricated the high-strength and high-toughness polymer yarns in three steps (fig. S1). First, continuous yarns were obtained by yarn electrospinning a solution of commercial polyacrylonitrile (PAN) {Dolan, copolymer with 4.18 mole % [6.35 weight % (wt %)] methyl acrylate according to our $^1$H-NMR (proton nuclear magnetic resonance) analysis, number average molar mass ($M_n$) = 120,000; molar mass dispersity ($Đ$) = 2.79} and different amounts of the bifunctional poly(ethylene glycol) bisazide (PEG-BA) as an interconnecting molecule. Azides were applied in "click" reactions (*18*). These yarns were then stretched at 160°C in air. The stretched yarns were further annealed at 120°, 130°, and 140°C under tension for several hours, which finally resulted in high-strength and high-toughness yarns, depending on the applied conditions and the composition of PAN and PEG-BA (we tested amounts ranging from 0 to 6 wt % of

[1]Macromolecular Chemistry and Bavarian Polymer Institute, University of Bayreuth, 95440 Bayreuth, Germany. [2]JCNS-1/ICS-1, Forschungszentrum Jülich, 52425 Jülich, Germany. [3]Institute of Physics, Martin Luther University Halle-Wittenberg, Heinrich-Damerow-Straße 4, 06120 Halle (Saale), Germany. [4]Fraunhofer Institute for Microstructure of Materials and Systems (IMWS), Walter-Hülse-Straße 1, 06120 Halle (Saale), Germany. [5]Physical Chemistry, Rheinisch-Westfälische Technische Hochschule Aachen University, 52074 Aachen, Germany. [6]College of Chemistry and Chemical Engineering, Jiangxi Normal University, Nanchang, Jiangxi 330022, People's Republic of China. [7]ETH Zürich, HCP F41.2, 8093 Zürich, Switzerland.
*Corresponding author. Email: greiner@uni-bayreuth.de

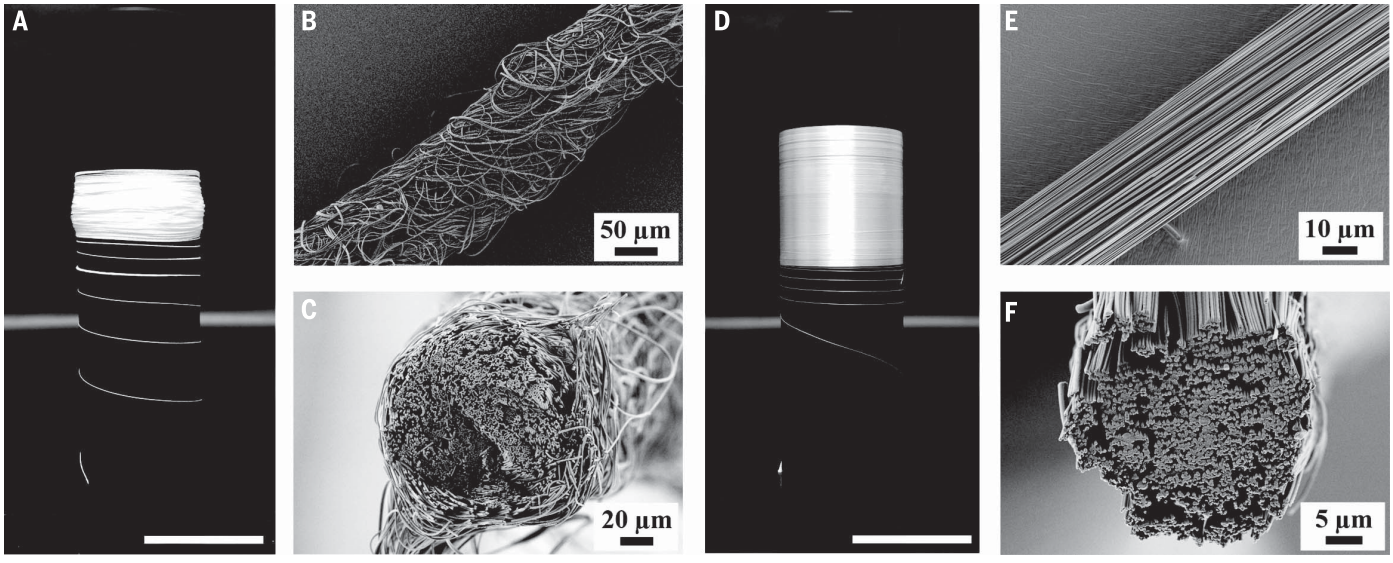



**Fig. 1. Photographs and scanning electron microscopy (SEM) images of the yarns.** (**A**) Photograph of continuous as-spun yarns. (**B**) SEM image of the long axis of the as-spun yarns. (**C**) SEM image of a cross-section of the as-spun yarns. (**D**) Photograph of stretched yarns. (**E**) SEM image of the long axis of stretched (at SR 8 at 160°C) and annealed (130°C for 4 hours) yarns (EASY). (**F**) SEM image of a cross-section of the stretched (at SR 8 at 160°C) and annealed (130°C for 4 hours) yarns (EASY). The scale bars in the photographs of the as-spun yarns (A) and stretched yarns (D) are 20 mm.

PEG-BA relative to PAN). The electrospun, stretched, and annealed yarns are hereafter abbreviated as EASYs; the yarns processed with PEG-BA are abbreviated as i-EASYs.

In a model study for the systematic understanding of the effect of stretching with commercial PAN without PEG-BA, we found that as-spun yarns had an average diameter of 130 ± 12 μm and consisted of ~3000 nonoriented individual fibrils with 1.17 ± 0.12 μm diameter (Fig. 1, A to C, and fig. S2, A and B). Heat stretching of the yarns for several minutes resulted in their manifold elongation accompanied by a change in the macroscopic appearance (Fig. 1D) and the alignment of the fibrils in the yarns (Fig. 1E). Stretching needed to be conducted above the glass transition temperature ($T_g$) of PAN ($T_g$ = 103°C according to our differential scanning calorimetry measurement) but below the onset of oxidation at 180°C (*19*). We investigated yarns with different stretch ratios (SRs) ranging from 1 to 9 (where SR is the length of stretched yarn divided by the length of as-spun yarn, and a SR of 1 is unstretched) at stretch temperatures of 130° and 160°C and determined the alignment factor. The alignment factor [orientation of the fibrils of the yarn, with values ranging from 0 for an isotropic orientation to 100% for a perfect alignment (*20*)] increased from ~46.0% (at SR 1) to 99.6% at SR 9 and stretch temperature 160°C (fig. S3A). The orientation of the fibrils can also be seen in the three-dimensional (3D) x-ray images of the yarn samples (fig. S4 and movies S1 and S2). The tortuosity estimation, calculated using the 3D images, is equal to 1.00 for stretched yarns (at SR 8 at 160°C) and agrees well with the uniform orientation of the fibrils. The stretching of yarns naturally caused a reduction of their diameter, from 130 ± 12 μm (unstretched yarns) to 50 ± 3.3 μm (at SR 5 at 130°C) and 36 ± 1.3 μm (at SR 9 at 160°C) (fig. S2, C and D). Simultaneously, the diameters of the fibrils reduced from 1.17 ± 0.12 μm to 0.57 ± 0.01 μm and to 0.37 ± 0.07 μm at 130° and 160°C, respectively (Fig. 1F and fig. S3B). The reduction in diameter of the yarns after stretching can be explained by the untwisting and alignment of the fibrils. Stretching also reduced the linear densities of the yarns, which changed from 3.74 ± 0.14 tex [mass of fiber (g)/1000 m] in the as-spun yarns to 0.39 ± 0.04 tex at SR 9 at 160° C (fig. S3C). After heat stretching, annealing under tension (about 15 to 20 cN) was applied to achieve high toughness and high strength of the yarns. This annealing step (130°C for 4 hours in air) did not result in any further changes in the diameters of the yarns (EASY) or of the fibrils (fig. S3D).

The results of the model studies were transferred to yarns composed of PAN and the interconnecting molecule PEG-BA. Bisazides were reported to undergo the [2+3] click azide cycloaddition reaction (*18*) with the acrylonitrile groups of PAN, which could either favorably lead to bridging between the fibrils in the yarns or cause reaction with PAN in the bulk of the yarns. PEG-BA contents in yarns in the range from 0 to 4 wt % had no significant effect on the diameter of stretched and annealed yarns (fig. S3D). To analyze the effect of PEG-BA on the mechanical properties of the yarns, the toughness and the specific strength were analyzed for stretched yarns with different contents of PEG-BA relative to PAN. The maximum stress (fig. S5A) and modulus (fig. S5B) increased with the SR, whereas the toughness did not linearly increase with the SR (fig. S5C). The increase of PEG-BA content decreased the maximum stress and modulus slightly (fig. S5, D and E). In contrast, the toughness increased slightly (fig. S5F). However, the subsequent annealing step had a significant effect on the toughness of the yarns when PEG-BA was present. An annealing time of 4 hours was found to be optimal for the maximum strength, modulus, and toughness at an annealing temperature of 130°C (fig. S5, G to I). Optimum values were obtained with 4 wt % PEG-BA, at SR 8 at 160°C, with subsequent annealing at 130°C for 4 hours (Fig. 2 and Fig. 3, A and C). These yarns (i-EASY) have a tensile strength of 1236 ± 40 MPa, a modulus of 13.5 ± 1.1 GPa, and a toughness of 137 ± 21 J/g, which are similar properties to those of dragline spider silk (*15*, *16*). They also have a value for the tensile modulus of 13.5 GPa, which is close to the theoretical limit calculated for atactic crystalline PAN fibers (*21*). For comparison, the strength of an aligned, electrospun PAN nonwoven is 110 ± 12 MPa and its toughness is 57 ± 3.0 J/g. Untreated as-electrospun PAN yarn has a strength of 72 ± 3.0 MP and its toughness is 76 ± 10 J/g. The linear density of i-EASY was only 0.4 ± 0.06 tex and it had an alignment factor of the fibrils of 99.4%. This yarn, weighing 0.008 mg, could lift a total mass of up to 30 g repeatedly without breaking (movie S3) and could even be used to sew a button to a shirt (movie S4). After repeatedly lifting 30-g weights, the yarn elongated slightly, most likely because of the elongation at the yielding point (strain of ~2.5%). Even after 5000 cycles of loading and unloading at a maximum tensile strength of 400 MPa, only a negligible change in the final tensile strength (~5.3%) and plastic deformation (~4.2%) was observed (fig. S6, A and B). The plastic deformation occurred mainly in the first 10 cycles. Owing to minor structural changes after 10 cycles, the energy loss coefficient with cycling showed a slight decrease, with values in the range of 0.23 to 0.15 (fig. S6C).

On the basis of our findings, we postulate that only the combination of high-fibril orientation, caused by stretching and annealing in the presence of a certain amount of PEG-BA as an interlinking molecule, yielded high strength and high toughness in combination with high crystallinity (Fig. 3, A to C). Clearly, the crystallinity of the PAN in the yarns is insufficient on its own for the achievement of high strength in combination with high toughness. Polarized Raman spectroscopy confirmed that the heat stretching procedure caused the orientation of the PAN macromolecules along the yarn's main axis (fig. S7A), with the percentage of aligned yarns increasing from 66.1% at no-stretch to

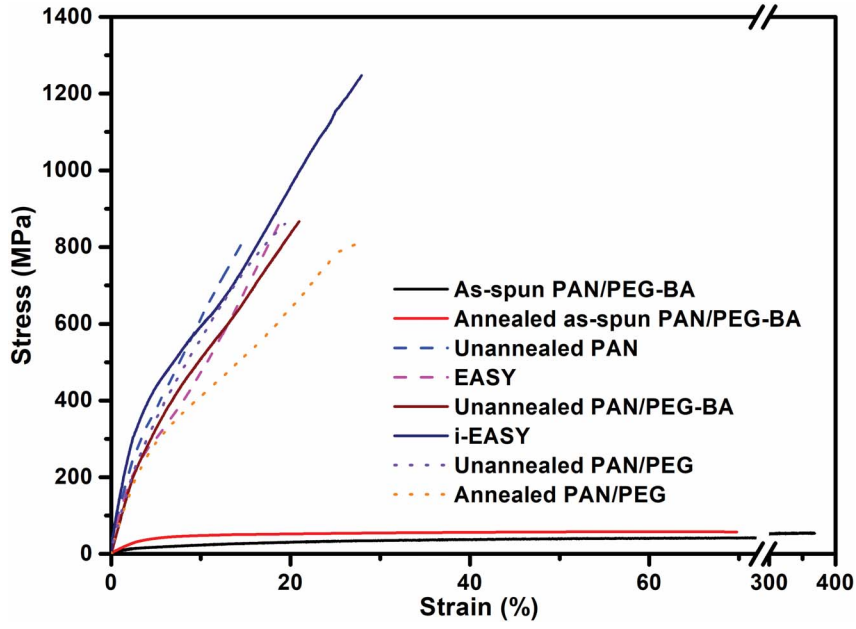



**Fig. 2. Comparison of different yarns.** Stress-versus-strain curves of unannealed and annealed (130°C for 4 hours) yarns (at SR 8) with 0 wt % PEG-BA, 4 wt % PEG-BA, and 4 wt % PEG.

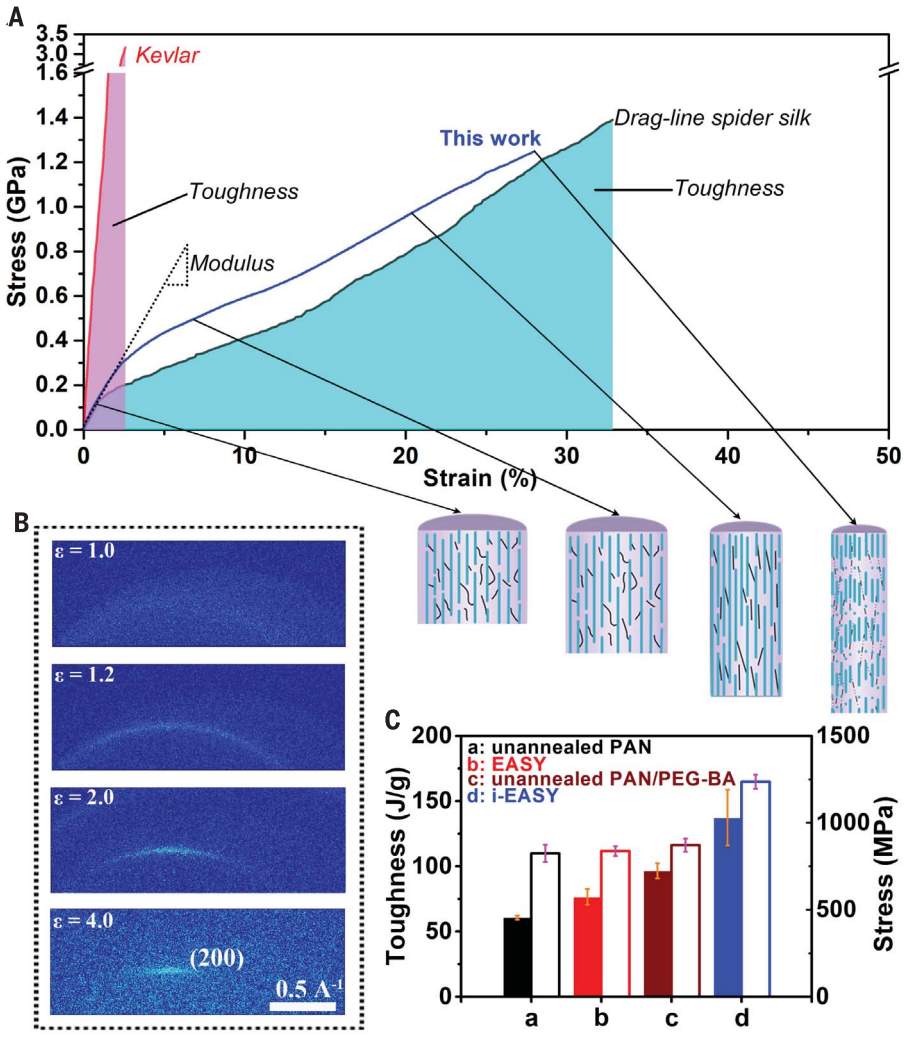



**Fig. 3. Comparison of tensile strength and toughness of stretched and annealed yarns and their crystallization during heat stretching.** (**A**) Comparison of stress-strain behavior and toughness of yarns (at SR 8 at 160°C, annealed at 130°C for 4 hours) with dragline spider silk and Kevlar [silk and Kevlar data are taken from the literature (*16*, *25*)]. A model for the stress/strain behavior of yarns is presented beneath the graph. The blue lines represent PAN fibrils, and the black lines denote PEG-BA moieties. (**B**) In situ 2D-WAXS patterns recorded during the stretching process of a single yarn at 160°C. With increasing extension, we observe the development of a sharp Debye-Scherrer ring, subsequently followed by the development of a sharp (200) reflection, indicating crystal formation and alignment with high-orientation orders. (**C**) Comparison of the toughness (filled columns) and strength at break (open columns) of unannealed and annealed yarns with a SR 8 at 160°C. Error bars indicate the standard deviation of toughness and strength.

83.3% at SR 8 (stretched at 160°C). Wide-angle x-ray scattering experiments demonstrated that heat stretching resulted in a marked increase in the crystallinity from ~56.9% (at no stretch) to ~92.4% (at SR 9; fig. S7B), whereas annealing alone did not significantly increase crystallinity (fig. S7C). The size of the crystallites increased during stretching, from ~3.4 nm (as-spun) to ~12.9 nm and accompanied the increase in the degree of crystallinity (at SR 9; fig. S7D). Tensional forces during annealing are necessary to preserve the high degree of crystalline orientation. This is also supported by in situ x-ray diffraction measurements of the crystalline orientation during stretching at 160°C (Fig. 3B). The crystallinity orientational order parameter (*22*) increased from 0.37 to 0.96 by heat stretching, but no significant increase was observed upon annealing (Fig. 3B and fig. S7, E to I). If no tension force was applied, the orientation parameters dropped from 0.96 to 0.82 during annealing because of thermal motion, which caused macromolecular relaxation and a reduction in the mechanical properties.

We analyzed the development of the specific strength and toughness throughout different steps of preparation, which are: as-spun (green star in fig. S8), stretched (purple star in fig. S8), and finally annealed (blue oval in fig. S8). The strength primarily increased because of the stretching, whereas the toughness increased because of the annealing after stretching, which caused alignment of the fibrils and crystallization of PAN.

Too much PEG-BA can have an adverse impact on fibril resilience. Yarns with 5 wt % and 6 wt % PEG-BA featured lower strength and toughness than those with 4 wt % PEG-BA (fig. S5, E and F). Simultaneously, control experiments were performed by obtaining yarns with PAN and pure PEG ($M_n$ = 1000, 4 wt %) (Fig. 2), which show a slight decrease in tensile strength (from 858 ± 103 MPa to 784 ± 77 MPa) and modulus (from 7.6 ± 1.0 GPa to 7.3 ± 1.1 GPa), a big increase in elongation at break (from 18.0 ± 1.6 % to 27 ± 2.5 %), and a slight increase in toughness (from 83 ± 12.2 J/g to 89 ± 15.0 J/g) after annealing. The yarns consisting of PAN and PEG (4 wt % but without azide groups) have almost the same elongation at break as yarns made from PAN and PEG-BA (4 wt % with azide groups), but they have lower strength. This suggests that the flexible oligomer PEG makes the yarns have higher elongation at break but lower strength than the PEG-BA yarns, due to the missing interaction between the fibrils. The interaction between cyano groups and PEG in the presence of azide groups was confirmed by $^{1}$H-NMR, $^{13}$C-NMR (carbon-13 nuclear magnetic resonance), and ATR-FTIR (attenuated total reflectance–Fourier transform infrared spectroscopy) spectra (figs. S9 to S12). A new peak at about 4.2 parts per million (ppm) in $^{1}$H-NMR and 162 ppm in $^{13}$C-NMR spectra represents the carbon in the tetrazole after reaction between nitrile and azide groups. Further, a decreased intensity of the peak at 2100 cm$^{-1}$ represents azide groups in the ATR-FTIR spectra, which suggests that the nitrile group in the PAN and azide groups in the PEG-BA can have a reaction in the yarns during the annealing process. i-EASYs are still soluble in *N*,*N*′-dimethylformamide and do not show any increase in molecular weight according to gel permeation chromatographic analysis, which strongly supports the idea that intermolecular cross-linking reactions of PAN molecules in the bulk of the fibrils did not occur. The most plausible explanation for these findings is that the reaction occurs only on the surface of the fibrils. Therefore, we postulate that under the presented conditions, interfibrillar reactions via PEG-BA are the dominating reaction.

A possible model for understanding the mechanical properties of i-EASY is shown in Fig. 3A. It highlights a reduction in the fractional-free volume after strain-induced crystallization and shrinking. As a result, PEG-BA macromolecules in the PAN matrix probably diffuse

to the fibril-free surface, which is the optimal position for efficient interfibrils reaction. Too much PEG-BA may lead to PEG microphases that represent defect structures that deteriorate the mechanical properties. Starting from pristine yarns, the PAN fibrils in the yarns begin to disentangle, resulting in a yield point. Beyond the yield point, the PEG-BA moieties bridging the PAN fibrils are responsible for stress transfer, impeding their ability to slide along and over each other. At a critical stress, the PEG-BA bridges might rupture, causing the yarns to break. This mechanism can be illustrated assuming the case of a single craze propagating ahead of the crack during stable crack growth. Then the local crack driving force (*23*, *24*) is $\sim\frac{1}{\sqrt{E_1E_2}}$, where, $E_1$ is the modulus in the fibril direction and $E_2$ is the modulus perpendicular to the fibrillar direction, which can be related to the PEG-BA elastic modulus. Thus, increasing $E_2$ should result in delaying the crack propagation. This mechanism is also supported by the fact that, at higher cross-linking density, the mechanical properties worsen.

To achieve the high strength and high toughness combination, annealing was required in addition to orientation and crystallization. Though single-polymer nanofibers could yield very high strength and high toughness, the handling difficulty of these single nanofibers has prevented their use in real-world applications. The presented process yields ultrafine electrospun yarns robust enough for practical use, through a combination of thousands of fibrils with high alignment and interfibrillar reactions. We are convinced that ultrafine polymer yarns, as we show here, are not restricted to the present system. An important precondition for other systems is the extensibility of the yarn for orientation of the fibrils and reactive groups for interlinking of the fibrils.

## REFERENCES AND NOTES


1. R. O. Ritchie, *Nat. Mater.* **10**, 817–822 (2011).
2. M. F. Ashby, *Materials Selection in Mechanical Design*, (Butterworth & Heinemann, ed. 2, 1999).
3. D. C. Hofmann *et al*., *Nature* **451**, 1085–1089 (2008).
4. H. Zhu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8971–8976 (2015).
5. M. D. Demetriou *et al*., *Nat. Mater.* **10**, 123–128 (2011).
6. A. B. Dalton *et al*., *Nature* **423**, 703 (2003).
7. M. Naraghi *et al*., *ACS Nano* **4**, 6463–6476 (2010).
8. Z. Xu *et al*., *Adv. Mater.* **28**, 6449–6456 (2016).
9. J. Cai, M. Naraghi, *Carbon* **137**, 242–251 (2018).
10. A. Arinstein, M. Burman, O. Gendelman, E. Zussman, *Nat. Nanotechnol.* **2**, 59–62 (2007).
11. D. Papkov *et al*., *ACS Nano* **7**, 3324–3331 (2013).
12. Y. Ding, H. Hou, Y. Zhao, Z. Zhu, H. Fong, *Prog. Polym. Sci.* **61**, 67–103 (2016).
13. J. H. Park, G. C. Rutledge, *J. Mater. Sci.* **53**, 3049–3063 (2018).
14. D. Papkov *et al*., *ACS Nano* **13**, 4893–4927 (2019).
15. D. A. Tirrell, *Science* **271**, 39–40 (1996).
16. F. Vollrath, D. P. Knight, *Nature* **410**, 541–548 (2001).
17. K. Spiess, A. Lammel, T. Scheibel, *Macromol. Biosci.* **10**, 998–1007 (2010).
18. Z. P. Demko, K. B. Sharpless, *Angew. Chem. Int. Ed.* **41**, 2113–2116 (2002).
19. N. Yusof, A. F. Ismail, *J. Anal. Appl. Pyrolysis* **93**, 1–13 (2012).
20. For details of calculation, see supplementary materials.
21. T. Shen, C. Li, B. Haley, S. Desai, A. Strachan, *Polymer* **155**, 13–26 (2018).
22. Q. Ouyang *et al*., *J. Macromol. Sci. B* **50**, 2417–2427 (2011).
23. H. R. Brown, *Macromolecules* **24**, 2752–2756 (1991).
24. H. H. Kausch, C. J. G. Plummer, *Polymer (Guildf.)* **35**, 3848–3857 (1994).
25. D. Zhu, X. Zhang, Y. Ou, M. Huang, *J. Compos. Mater.* **51**, 2449–2465 (2016).


## ACKNOWLEDGMENTS


We thank R. Schneider for the GPC measurements, H. Schmalz and L. Benker for measurements of polarized Raman spectroscopy, P. Schmidt and T. Braun for support in construction of the yarn electrospinning set-up, S. Jiang for advice in electrospinning, and A. Wambach for technical support with figures and data arrangement. We acknowledge the use of the research facilities of the University of Bayreuth, the Bavarian Polymer Institute, the Institute of Physics - Martin-Luther-Universität Halle-Wittenberg, and the Forschungszentrum Jülich for this project. **Funding:** The work was supported by the University of Bayreuth (A.G.) and the Deutsche Forschungsgemeinschaft (WE4051/21-1 and 4051/22-1, to J.M.d.S.e.S.). **Author contributions:** A.G., X.L., and S.A. conceived and supervised the project. X.L. prepared the yarns; measured WAXS, TGA, NMR, and IR; and performed stress-strain tests. X.L., A.G., and S.A. wrote the manuscript. M.D. measured the crystal data and evaluated and discussed with S.F. X-ray computed tomography was performed by J.M.d.S.e.S. and was evaluated and discussed by J.M.d.S.e.S. and R.B.W. P.S. and H.H. evaluated and discussed the mechanical data of the yarns with X.L. and A.G. All authors contributed to the analysis and discussion of the data. The manuscript writing was led by A.G. with contributions from all other authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Primary data are available from https://myfiles.uni-bayreuth.de/ssf/s/readFile/share/31206/3130562511270201132/publicLink/Raw%20data%2020191018.zip.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1376/suppl/DC1
Materials and Methods
Figs. S1 to S12
Table S1
References (*26*–*46*)
Movies S1 to S4

27 July 2019; accepted 13 November 2019
10.1126/science.aay9033

NANOMATERIALS

# Wafer-scale synthesis of monolayer two-dimensional porphyrin polymers for hybrid superlattices

Yu Zhong[1]*, Baorui Cheng[1]*, Chibeom Park[1,2], Ariana Ray[3], Sarah Brown[1,2], Fauzia Mujid[1], Jae-Ung Lee[4,5], Hua Zhou[6], Joonki Suh[1], Kan-Heng Lee[4,7], Andrew J. Mannix[2], Kibum Kang[1,2,8], S. J. Sibener[1,2], David A. Muller[7], Jiwoong Park[1,2,4]†

The large-scale synthesis of high-quality thin films with extensive tunability derived from molecular building blocks will advance the development of artificial solids with designed functionalities. We report the synthesis of two-dimensional (2D) porphyrin polymer films with wafer-scale homogeneity in the ultimate limit of monolayer thickness by growing films at a sharp pentane/water interface, which allows the fabrication of their hybrid superlattices. Laminar assembly polymerization of porphyrin monomers could form monolayers of metal-organic frameworks with $Cu^{2+}$ linkers or covalent organic frameworks with terephthalaldehyde linkers. Both the lattice structures and optical properties of these 2D films were directly controlled by the molecular monomers and polymerization chemistries. The 2D polymers were used to fabricate arrays of hybrid superlattices with molybdenum disulfide that could be used in electrical capacitors.

Monolayer two-dimensional polymers (2DPs), which are one-molecule-thick, freestanding films composed of periodically linked monomers (*1*–*4*), offer an ideal material system with two key advantages. First, their properties can be tuned at the molecular level by using different monomers and polymerization chemistries (*5*, *6*). Second, as the molecular analogs of 2D atomic crystals [e.g., graphene and transition metal dichalcogenides (TMDs)] (*7*–*9*), 2DPs can be assembled through van der Waals (vdW) interactions into heterostructures and superlattices, layer by layer. vdW heterostructures generated from 2D atomic crystals have produced properties not observed in individual building blocks (*10*, *11*). Adding the chemical tunability of the 2DPs to such vdW heterostructures will lead to the properties and functionalities designed at the molecular level and further tuned by the interlayer interactions. However, it has remained an unmet challenge to scalably synthesize monolayer 2DP films and subsequently integrate them with other materials with monolayer precision (*12*, *13*). This is due to the difficulty of controlling reactions in the monolayer limit with large-scale uniformity and to the lack of facile methods for the transfer and integration of monolayer 2DPs because of their fragility. Previous experiments have reported progress toward large-scale synthesis of 2DPs (*14*–*22*) but with limited success with regard to wafer-scale homogeneity, microscopic characterization of crystalline structures, and scalable thin-film integration (*23*).

Here, we report the wafer-scale synthesis and integration of monolayer 2DPs for the fabrication of their hybrid heterostructures with monolayer precision. We developed an interfacial synthesis technique, laminar assembly polymerization (LAP), that is compatible with various molecular building blocks and two primary polymerization chemistries (coordination and covalent). This approach incorporated key features necessary for scalable and facile processing, including large-area synthesis, ambient growth conditions, and compatibility with established patterning and integration methods. These characteristics enabled the fabrication of superlattices based on monolayer 2DPs and 2D atomic crystals.

The design approach for the 2DP monolayers was based on porphyrin building blocks (Fig. 1A). These molecules had two variation sites: one at the center of the porphyrin ring [M = 2H, Fe(III), or Pt(II)] that tuned the optical spectra (Fig. 1B) and the other on the phenyl groups (R = -COOH or $-NH_2$) that controlled the monomer-monomer bonds. The monomers were cross-linked either through coordination bonds via a copper paddle wheel structure in the presence of $Cu^{2+}$ ions (Fig. 1A, left, and fig. S1; R = -COOH) (*14*) or through covalent bonds via the Schiff base reaction in the presence of terephthalaldehyde (TPA) (Fig. 1A, right, and fig. S1; R = $-NH_2$) (*21*). The former forms coordination 2DPs (**2DP I**, **-II**, and **-III**; M = 2H, $Pt^{2+}$, or $Fe^{3+}$, respectively), also known as monolayer metal-organic frameworks (MOFs), whereas the latter forms covalent 2DPs (**2DP IV**; M = 2H), also known as monolayer covalent organic frameworks (COFs). The linkage chemistry for all the 2DPs was confirmed by Fourier-transform infrared spectroscopy (FTIR) (fig. S2).

Wafer-scale 2DP films were all produced at a sharp pentane-water interface and then transferred onto a substrate (e.g., fused silica in Fig. 1) placed underneath by slowly draining the bottom solution (more details are in the supplementary text). We visualized these films using a custom color-coding scheme based on the hyperspectral optical transmission images (Fig. 1C and fig. S3). Images of four transferred 2DP monolayers that covered an entire 2-inch (5-cm) fused silica substrate are shown in Fig. 1D. The films displayed uniform contrast over entire wafers, suggesting macroscopic continuity and homogeneity (a higher-resolution analysis is shown in fig. S4). The MOF-based **2DP I**, **-II**, and **-III** with different M had distinct absorption spectra (Fig. 1B), resulting in markedly different colors (shown in Fig. 1D). The absorption spectra of the 2DPs resembled those of the corresponding porphyrin monomers (figs. S5 and S6 and discussions in the supplementary text), indicating that the optical properties of the 2DP films could be directly tuned at the molecular level.

These monolayer 2DPs were synthesized using the LAP explained in Fig. 2. It is based on monomer self-assembly and polymerization at the sharp interface formed between two immiscible solvents (pentane/water) that strictly confined the monomers in a monolayer limit, which was critical for precise control of the thickness (Fig. 2, A and B). Laminar flow of the assembled monomers led to large-scale continuity and homogeneity in thickness (Fig. 2A and fig. S7 describe the LAP synthesis and the in situ optical characterization apparatus; additional details are in the supplementary text).

There are three phases in the LAP process (illustrated in Fig. 2, A and B): injection, self-assembly, and polymerization. During injection, the monomers were introduced from the edge of the reactor (width $W$) and directly delivered onto the sharp pentane/water interface by a continuous stream of carrier solution through the pentane layer (within 1 cm from the edge, movie S1). The pentane-mediated delivery has two key advantages. First, the mass flow of the precursor is continuous at the interface, which is achieved by using microsyringe pumps and by carefully choosing the combination of the carrier solvents. Second, the pentane/water interface is steady during the growth, resulting in minimal disturbance. This contrasts with dropwise delivery through the air, which disturbs the interface. Once delivered to the interface, the porphyrin-based monomers self-assembled at the interface because of their amphiphilicity and then spread, while being restricted by the longer sidewalls

[1]Department of Chemistry, University of Chicago, Chicago, IL 60637, USA. [2]James Franck Institute, University of Chicago, Chicago, IL 60637, USA. [3]Department of Physics, Cornell University, Ithaca, NY 14853, USA. [4]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637, USA. [5]Department of Physics, Ajou University, Suwon 16499, Republic of Korea. [6]Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA. [7]School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA. [8]Department of Materials Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea.
*These authors contributed equally to this work.
†Corresponding author. Email: jwpark@uchicago.edu

(length $L$). This process generated laminar flow of the monomers away from the injection region and resulted in a continuous monolayer assembly. Then, the polymerization of assembled monomers takes place gradually through the reaction with the reagents present in water ($Cu^{2+}$ ions for MOF-based **2DP I**, **-II**, and **-III**; TPA for COF-based **2DP IV**).

The monolayer nature of the 2DPs was confirmed by optical images that showed unidirectional movement of the monolayer assembly parallel to the longer sidewalls (Fig. 2C and movie S2) with little mixing perpendicular to this direction (Fig. 2F), confirming a laminar flow. This monolayer remained intact upon solvent washing after a complete polymerization (~30 min); in contrast, unpolymerized films were washed away (Fig. 2D). Quantitative measurements of the synthesized area of **2DP I** (R = -COOH and M = 2H with $Cu^{2+}$ ions) as a function of the injected volume of the monomer solution closely followed a linear growth model consistent with a near-unity monomer-to-monolayer yield (Fig. 2E and movie S3).

The LAP synthesis offers several advantages important for thin-film processing and integration of 2DP monolayers. First, the growth can be easily scaled up by injecting more monomers (for greater $L$) and by adding an array of nozzles in parallel (for greater $W$) (fig. S8). For example, the 2-inch films shown in Fig. 1D were synthesized with three nozzles in a 2-inch ($W$) by 5-inch ($L$) reactor. Second, lateral heterojunctions of monolayer 2DPs can be grown with tunable compositions and widths by introducing different monomers from each nozzle and by controlling the relative injection rates (Fig. 2, F and G, and fig. S9). Sharp interfaces between adjacent monolayer stripes were observed without voids. Finally, the 2DP films were compatible with a wide range of patterning and transfer techniques. For instance, they could be transferred to various substrates without tearing, distortion, or buckling after evaporating pentane (e.g., $SiO_2$/Si in Fig. 3B and gold in Fig. 3D), and they could be patterned by using a laser marker while still on the water surface with a scanning laser (fig. S10). Multiple patterning and transfer steps can be combined to fabricate laterally patterned and vertically stacked heterostructures while maintaining the integrity of the intricate patterns, as shown in Fig. 2H and fig. S10.

The 2DP films were mechanically robust and homogeneous in thickness. On the large scale, they exhibited considerable mechanical strength and could be transferred onto various substrates as continuous films, as shown in Fig. 1D. As an additional example, a scanning electron microscopy (SEM) image of a **2DP I** film transferred and suspended over a holey transmission electron microscope (TEM) grid

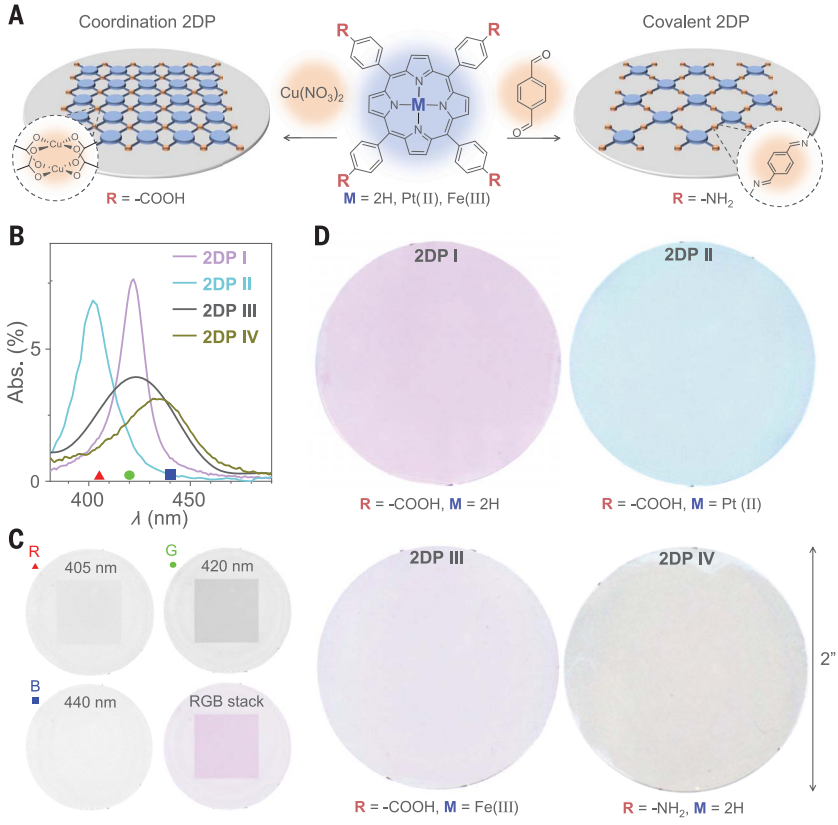



**Fig. 1. Wafer-scale monolayer 2DPs.** (**A**) Schematic of monolayer 2DPs and corresponding chemical structures of the molecular precursors. (**B**) Absorption spectra of monolayer 2DPs on fused silica substrates. (**C**) Hyperspectral transmission images and resulting false-color images of 1-inch-square **2DP I** on a 2-inch fused silica substrate. Transmission images taken at the wavelengths of 405, 420, and 440 nm were assigned red, green, and blue channels, respectively, to generate the false-color image. A linear transmission scale of 50 to 95% was applied to all of the channels. (**D**) False-color images of monolayer 2DPs covering entire 2-inch fused silica wafers. The same color code was applied in (C) and (D).

(2-μm-diameter holes) (Fig. 3A) displays an array of freestanding 2DP membranes. These membranes were suspended with a near-perfect yield (> 99%; one broken membrane, denoted by an arrow) and appeared uniform and continuous over the entire area without cracks or voids. These 2DP films were close to 1 nm in height, near the expected thickness of a monolayer (*24*), with a uniform and smooth surface as measured by atomic force microscopy (Fig. 3B and fig. S11).

The MOF-based **2DP I**, **-II**, and **-III** showed a polycrystalline structure, which was confirmed by synchrotron grazing incidence x-ray diffraction (GIXRD) (Fig. 3C and figs. S12 and S13). Using **2DP II** as an example, the in-plane XRD pattern showed all of the main peaks predicted on the basis of the structure model (inset in Fig. 3C and table S1), and the average lateral domain size was estimated to be ~20 nm according to the Scherrer equation. As additional evidence, the crystalline structure of **2DP II** monolayers transferred onto flat Au(111) surfaces was confirmed with scanning tunneling microscopy (STM) performed under ultrahigh vacuum. The STM topography image (Fig. 3D) showed a square lattice with a single-crystalline domain that fully covered the 30-nm by 30-nm area [see the 2D fast Fourier transform (FFT) image (inset in Fig. 3D)]. Another STM image (Fig. 3E; 60 nm by 60 nm) displayed three primary crystalline orientations [lattice constant $a$ = 1.66 ± 0.03 nm (mean ± standard error of the mean), measured from Fig. 3F], suggesting that the **2DP II** films are polycrystalline with domain structures similar to those in previously imaged 2DPs (*25*, *26*). The lattice constant extracted from these microscopic STM analyses is close to that from GIXRD measurements (1.64 nm) performed on the macroscopic scale (0.1 mm by 10 mm) with a mismatch less than 2% (Fig. 3C and figs. S12 and S13). In the composite inverse 2D FFT image in Fig. 3G, each region is colored according to the lattice orientation. We used this map to estimate the sizes of domains (between 10 and 40 nm) and locate domain boundaries (marked by dashed lines in Fig. 3E). For the

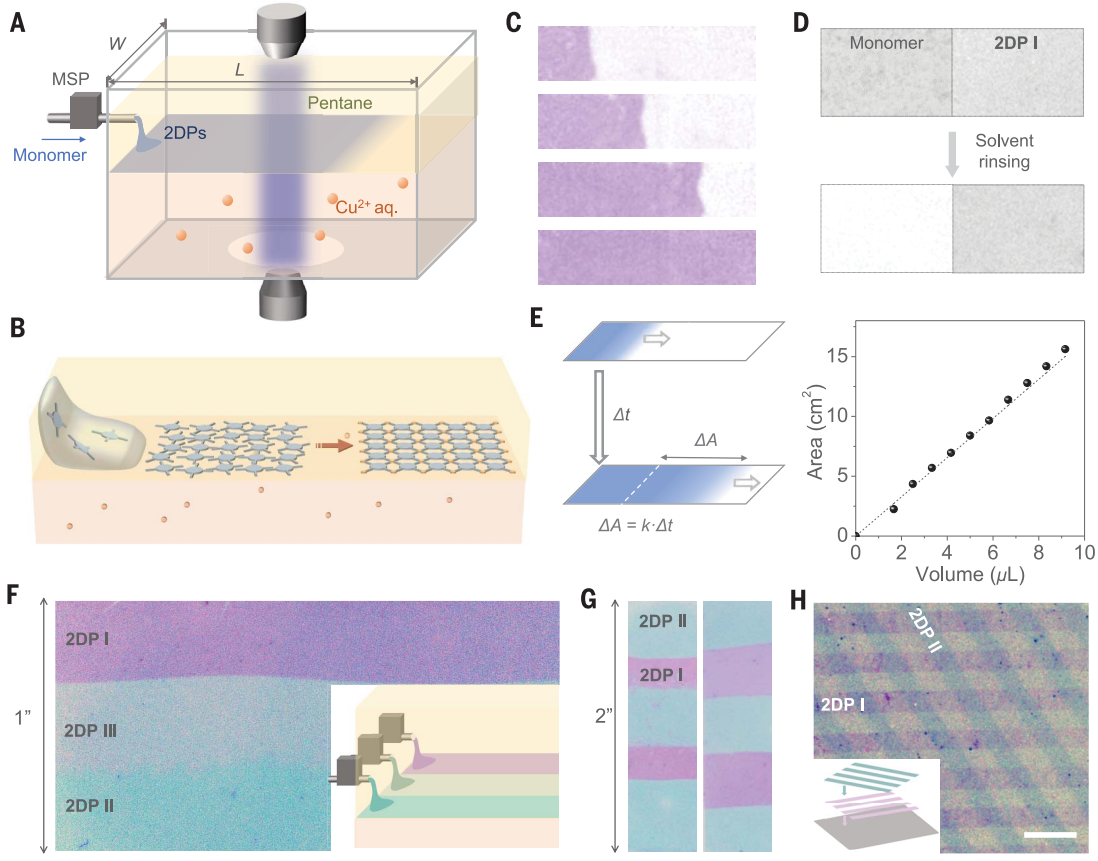



**Fig. 2. Laminar assembly polymerization.** (**A**) Schematic of a LAP reactor and in situ optical characterization apparatus. MSP, microsyringe pump. (**B**) Schematic of the LAP synthesis that involves three phases. (**C**) False-color images of **2DP I** film at four different stages during growth. Images are individual frames extracted from movie S2 measured at the wavelength of 425 nm. The precursor was injected from the left side. The film was colored with purple. The view size is 6 mm by 24 mm. (**D**) Optical transmission images comparing monolayer films produced with and without $Cu^{2+}$ ions before and after rinsing, measured at the wavelength of 425 nm. Image size is 0.67 cm by 1 cm. (**E**) (Left) Schematic of a linear growth model based on LAP. The film area increases linearly over time with a rate constant $k = C_N \cdot A_0 \cdot v \cdot \eta / N_{eff}$, where $C_N$ is number concentration (i.e., the number of molecules per microliter) of the molecular precursor, $A_0$ is the unit cell area of the **2DP I** lattice, $v$ is the volumetric injection rate, $\eta$ is the monomer-to-monolayer yield, and $N_{eff}$ is the effective layer number. (Right) Relation between film area and volume of the injected precursor, measured for **2DP I**. The dashed line indicates the theoretical curve for 100% monomer-to-monolayer conversion based on the lattice structure of **2DP I** ($\eta$ = 100%, $N_{eff}$ = 1). The data points were collected from movie S3. (**F**) False-color image of **2DP I/2DP III/2DP II** lateral junctions. (Inset) Schematic of generating lateral heterostructures of **2DP I/2DP III/2DP II** generated using three nozzles in LAP. (**G**) False-color images of **2DP I/2DP II** lateral junctions with tunable stripe widths. (**H**) False-color image of overlapped **2DP I** and **2DP II** stripes. Scale bar, 500 μm.

COF-based **2DP IV**, no evidence for long-range order could be collected (through GIXRD or selected area electron diffraction), similar to other monolayer covalent 2DPs reported previously (*26*).

In Fig. 4, we further demonstrate the potential of LAP by presenting an array of vertically programmed hybrid vdW superlattices. These superlattices were produced by repeatedly stacking in vacuum hybrid 2D building units 2DP/$(MoS_2)_n$, each made of a 2DP monolayer and $n$ monolayers of $MoS_2$. Examples of a **2DP II**/$(MoS_2)_3$ superlattice and a **2DP II**/$MoS_2$ film are shown in Fig. 4, A and B, respectively (detailed methods are shown in figs. S14 and S15) (*27*, *28*). Figure 4A shows a cross-sectional annular dark field (ADF) scanning transmission electron microscope (STEM) image of a representative **2DP II**/$(MoS_2)_3$ superlattice—an 11-layer stack—constructed by alternating one layer of **2DP II** and three layers of $MoS_2$. The image shows three bright bands separated by two dark lines. Each of the bright bands consisted of three layers of $MoS_2$, and the dark layer in between corresponds to a **2DP II** monolayer, as confirmed by the composite ADF and electron energy loss spectroscopy (EELS) mapping (Fig. 4A and fig. S16). The films ran parallel to each other with sharp interfaces and a uniform layer thickness over the entire 100-nm view of the ADF STEM image, indicating a high level of uniformity. In addition, the composition of the superlattice could be tuned by using a different 2DP, as demonstrated with the **2DP III**/$(MoS_2)_2$ superlattice shown in Fig. 4C. EELS data confirmed the chemical composition of each constituting layer, where **2DP III** was identified by a strong carbon signal and $MoS_2$ by a strong sulfur signal (Fig. 4C and fig. S16).

The vertical structure and composition of the hybrid vdW superlattices could be directly engineered by using different hybrid building units. Figure 4D (left) shows a series of vdW superlattices with varied superlattice periodicity $d$ made of **2DP II**/$(MoS_2)_n$ repeating units, with $n$ = 1, 2, or 3. The grazing incidence wide-angle x-ray scattering (GIWAXS) data presented in Fig. 4D (middle and right) and figs. S17 and S18 show the distinctive diffraction peak for each superlattice. By radially integrating the 2D GIWAXS images along the out-of-plane direction, 1D spectra were obtained in reciprocal space and used to measure $d$ (fig. S19). For example, the superlattice with $n$ = 3 showed $d$ = 3.5 nm and a vdW thickness of 1.5 nm for **2DP II**, which was close to the

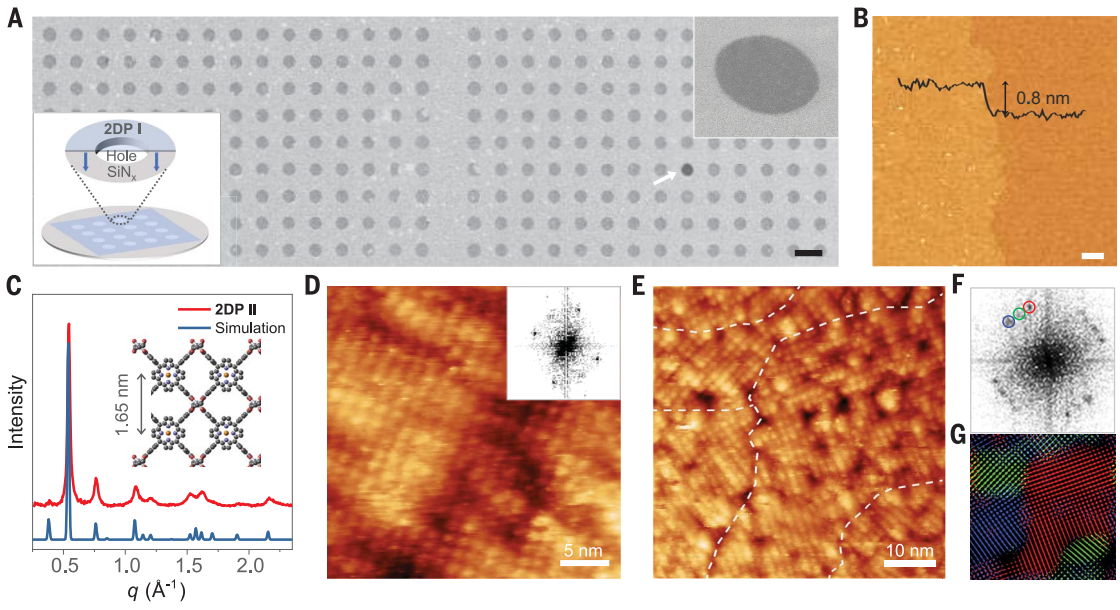



**Fig. 3. Structural characterizations of 2DPs.** (**A**) SEM image of monolayer **2DP I** on a holey silicon nitride TEM grid. The white arrow indicates a hole not covered by monolayer **2DP I**. Scale bar, 5 μm. (Bottom left inset) Schematic of monolayer **2DP I** suspended over a hole on a silicon nitride TEM grid. (Top right inset) Magnified SEM image of monolayer **2DP I** suspended over a 2-μm hole. (**B**) AFM height image of monolayer **2DP I**. Scale bar, 500 nm. (Inset) AFM height profile. (**C**) Experimental and calculated in-plane XRD profiles for **2DP II**. The experiment was conducted on a stacked **2DP II** of 147 layers on sapphire. (Inset) Crystal structure of **2DP II**. (**D**) Constant-current STM topography image of a single-crystalline domain of monolayer **2DP II** on a thin film of Au(111) on mica. (Inset) 2D FFT of the image. (**E**) Constant-current STM topography image of multiple-crystalline domains of monolayer **2DP II**. Boundaries between different domains are manually identified by the white dashed line. (**F**) 2D FFT of (E) showing square lattices of three major orientations. (**G**) Color-coded inverse 2D FFT image generated using the three sets of square lattice spots in (F). One spot from each set is circled with the corresponding color in (F).

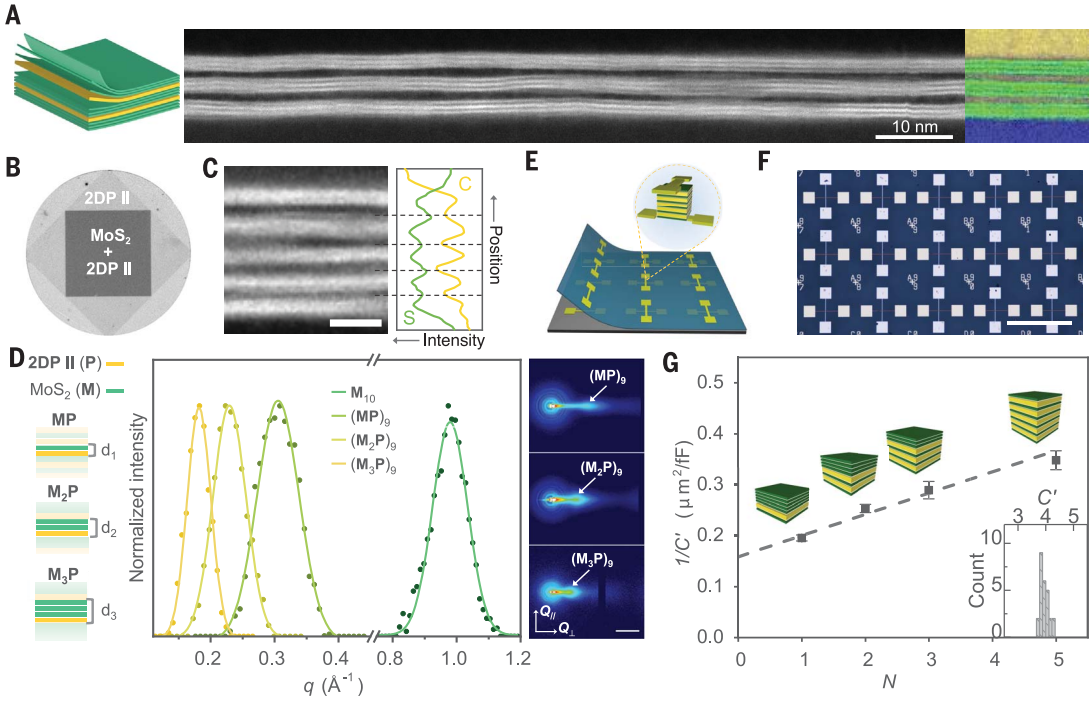



**Fig. 4. 2DP/TMD vertical superlattices.** (**A**) (Left) Schematic of a 2DP/$(MoS_2)_3$ superlattice. (Middle) Cross-sectional ADF STEM image of a **2DP II**/$(MoS_2)_3$ superlattice film transferred onto a $SiO_2$/Si substrate. Each bright band consists of three $MoS_2$ monolayers, and each dark layer between the bands is a monolayer **2DP II**. (Right) Composite image of carbon (yellow) and oxygen (blue) EELS mapping and ADF STEM signal (green) taken from a different area on the sample shown in fig. S16. (**B**) Optical transmission image of a **2DP II**/$MoS_2$ heterostructure on fused silica taken at the wavelength of 405 nm. The diameter of the wafer is 1 inch. (**C**) (Left) Cross-sectional ADF STEM image of a **2DP III**/$(MoS_2)_2$ superlattice film transferred onto a $SiO_2$/Si substrate. Each bright layer consists of two layers of $MoS_2$ stacked, and each dark layer is a **2DP III** monolayer. (Right) EELS profiles of carbon and sulfur taken from a different area on the sample shown in fig. S16. Scale bar, 5 nm. (**D**) (Left) Structures of **2DP II**/$(MoS_2)_n$ vertical superlattices. (Middle) Normalized diffraction peaks corresponding to **2DP II** /$(MoS_2)_n$ superlattices measured by GIWAXS. (Right) 2D GIWAXS scattering patterns of **2DP II**/$(MoS_2)_n$ superlattices. Scale bar, 0.2 Å$^{-1}$. (**E**) Schematic of vertical capacitor device arrays and individual device geometry. (**F**) Optical image of a 3-by-5 capacitor device array. Scale bar, 500 μm. (**G**) Reciprocal of area-normalized capacitance, *1/C'*, as a function of *N*, the number of **2DP II** layers in stacked $(MoS_2$/**2DP II**$)_N(MoS_2)_{6-N}$ films. Each data point is averaged from 10 devices with corresponding stacked film structures shown above. Green, $MoS_2$; yellow, **2DP II**. The inset shows a capacitance histogram of 25 devices of $N = 2$.

value measured from Fig. 4A. The results from other superlattices agreed very well with the predicted values (fig. S14 and table S2). In addition, x-ray reflectivity (XRR) measurements conducted on similar superlattice structures and the scattering length density profiles generated from fitting the XRR spectra clearly revealed oscillations of electron density consistent with the alternating structures of $(MoS_2)_n$ and **2DP II** (fig. S20 and supplementary text). The thickness of the repeating units extracted from the XRR analysis matched those obtained from GIWAXS and cross-sectional STEM within a 2-Å mismatch (table S3). Both the GIWAXS and the XRR data were taken from a macroscopic area randomly chosen from 1-cm by 1-cm superlattice films, illustrating the homogeneity of the vdW superlattices on a large scale.

Vertically programmed vdW superlattices and heterostructures provide a powerful platform for fabricating uniform arrays of devices whose properties are engineered layer by layer (*10*, *11*, *29*–*32*). To demonstrate such potential, we chose to fabricate arrays of electrical capacitors (photo shown in Fig. 4F) from vdW heterostructures of **2DP II** and $MoS_2$ (Fig. 4E), as such a process involves integration of both top and bottom electrodes and the capacitance can be directly tuned by the thickness of the superlattices. Each device in an array consisted of two gold electrodes sandwiching the vdW heterostructure dielectric (*27*). Figure 4G shows the results measured from a series of heterostructures, $(MoS_2/\mathbf{2DP\ II})_N(MoS_2)_{6-N}$, where $N$ monolayers of **2DP II** films were inserted in between $MoS_2$ layers of a six-layer $MoS_2$ stack (schematics in Fig. 4G). Thus, the dielectric thickness and the capacitance are directly tuned by varying the layer number of monolayer **2DP II**. The measured inverse capacitance, $1/C'$, where $C'$ is the area-normalized capacitance, linearly increased as $N$ increased from 1 to 5. Using the classical capacitor model, we extracted the dielectric constants of **2DP II** (4.1) and $MoS_2$ (2.7), and these were in agreement with reported values (*33*, *34*). The measured capacitance from an array of devices exhibited a narrow distribution (lower inset in Fig. 4G), suggesting the spatial uniformity of the hybrid heterostructures. This spatial uniformity is comparable to what has been achieved with stacked $MoS_2$ films (*27*). Thus, this method offers a general platform to incorporate diverse molecular species into vdW hybrid thin films for functional devices.

## REFERENCES AND NOTES


1. M. Servalli, A. D. Schlüter, *Annu. Rev. Mater. Res.* **47**, 361–389 (2017).
2. P. Payamyar, B. T. King, H. C. Öttinger, A. D. Schlüter, *Chem. Commun.* **52**, 18–34 (2016).
3. D. F. Perepichka, F. Rosei, *Science* **323**, 216–217 (2009).
4. G. Franc, A. Gourdon, *Phys. Chem. Chem. Phys.* **13**, 14283–14292 (2011).
5. H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **341**, 1230444 (2013).
6. P. J. Waller, F. Gándara, O. M. Yaghi, *Acc. Chem. Res.* **48**, 3053–3063 (2015).
7. K. S. Novoselov *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10451–10453 (2005).
8. K. F. Mak, C. Lee, J. Hone, J. Shan, T. F. Heinz, *Phys. Rev. Lett.* **105**, 136805 (2010).
9. K. S. Novoselov *et al*., *Science* **306**, 666–669 (2004).
10. K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. Castro Neto, *Science* **353**, aac9439 (2016).
11. A. K. Geim, I. V. Grigorieva, *Nature* **499**, 419–425 (2013).
12. X. Feng, A. D. Schlüter, *Angew. Chem. Int. Ed.* **57**, 13748–13763 (2018).
13. J. W. Colson, W. R. Dichtel, *Nat. Chem.* **5**, 453–465 (2013).
14. R. Makiura *et al*., *Nat. Mater.* **9**, 565–571 (2010).
15. R. Dong, T. Zhang, X. Feng, *Chem. Rev.* **118**, 6189–6235 (2018).
16. D. Sheberla *et al*., *J. Am. Chem. Soc.* **136**, 8859–8862 (2014).
17. R. Dong *et al*., *Angew. Chem. Int. Ed.* **54**, 12058–12063 (2015).
18. H. Sahabudeen *et al*., *Nat. Commun.* **7**, 13461 (2016).
19. R. Dong *et al*., *Nat. Mater.* **17**, 1027–1032 (2018).
20. T. Bauer *et al*., *Angew. Chem. Int. Ed.* **50**, 7879–7884 (2011).
21. M. Matsumoto *et al*., *Chem* **4**, 308–317 (2018).
22. V. Müller *et al*., *Angew. Chem. Int. Ed.* **56**, 15262–15266 (2017).
23. W. Wang, A. D. Schlüter, *Macromol. Rapid Commun.* **40**, e1800719 (2019).
24. S. Motoyama, R. Makiura, O. Sakata, H. Kitagawa, *J. Am. Chem. Soc.* **133**, 5640–5643 (2011).
25. J. F. Dienstmaier *et al*., *ACS Nano* **5**, 9737–9745 (2011).
26. L. Cardenas *et al*., *Chem. Sci.* **4**, 3263 (2013).
27. K. Kang *et al*., *Nature* **550**, 229–233 (2017).
28. K. Kang *et al*., *Nature* **520**, 656–660 (2015).
29. Y. Liu, Y. Huang, X. Duan, *Nature* **567**, 323–333 (2019).
30. Y. Liu *et al*., *Nat. Rev. Mater.* **1**, 16042 (2016).
31. D. Jariwala, T. J. Marks, M. C. Hersam, *Nat. Mater.* **16**, 170–181 (2017).
32. C. Wang *et al*., *Nature* **555**, 231–236 (2018).
33. X. Chen *et al*., *Nat. Commun.* **6**, 6088 (2015).
34. E. Redel *et al*., *Appl. Phys. Lett.* **103**, 091903 (2013).


## ACKNOWLEDGMENTS


We thank D. V. Talapin and J. S. Anderson for helpful discussions. We thank S. E. Kim for helping with preparing the manuscript. We thank Z. Zhang for helping with the XRR measurements.
**Funding:** This work was primarily supported by the Air Force Office of Scientific Research (FA9550-16-1-0031, FA9550-16-1-0347, and FA9550-18-1-0480) and the National Science Foundation (NSF) through the University of Chicago Materials Research Science and Engineering Center (MRSEC; NSF DMR-1420709). Additional funding was provided by the Cornell Center for Materials Research (NSF DMR-1719875) and the Platform for the Accelerated Realization, Analysis, and Discovery of Interface Materials (PARADIM; NSF DMR-1539918). Material characterizations including electron microscopy were supported by the Cornell Center for Materials Research (NSF DMR-1719875) and the MRSEC Shared User Facilities at the University of Chicago (NSF DMR-1420709). Funding was provided by the National Science Foundation grant nos. NSF-CHE-1566364 and NSF-CHE-1900188 for UHV-STM imaging and instrumentation. This work made use of the Pritzker Nanofabrication Facility at the University of Chicago, which receives support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-1542205), a node of the National Science Foundation's National Nanotechnology Coordinated Infrastructure. Y.Z. acknowledges support by the Camille and Henry Dreyfus Foundation, Inc., under the Dreyfus Environmental Postdoc award EP-16-094. F.M. acknowledges support by the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1746045. A.J.M. was supported by the Kadanoff-Rice Postdoctoral Fellowship through the University of Chicago MRSEC. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. **Author contributions:** Y.Z., B.C., and J.P. conceived the experiments. Y.Z. developed the LAP synthesis. B.C. fabricated the 2DP/TMD heterostructures and the capacitor devices, conducted GIXRD simulation, and performed the GIWAXS measurements. Y.Z. and B.C. grew the 2DP films, fabricated 2DP heterostructures, conducted AFM and SEM imaging, and performed the device measurements. Y.Z., B.C., and J.-U.L. carried out the optical characterizations, and C.P. and F.M. synthesized monolayer TMD films. A.R. and D.A.M. conducted TEM imaging, STEM imaging, and FIB milling. S.B. and S.J.S. conducted STM imaging. Y.Z., B.C., H.Z., J.S., and F.M. performed the GIXRD experiments and the structural characterizations. K.-H.L., K.K., A.J.M., and B.C. developed the stacking methods. Y.Z., B.C., and J.P. wrote the manuscript. All authors discussed and commented on the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are reported in the main text and supplementary materials.


## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1379/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S20
Table S1 to S3
References (*35*–*45*)
Movies S1 to S3


6 May 2019; accepted 28 October 2019
Published online 7 November 2019
10.1126/science.aax9385

NEUROSCIENCE

# Sensory coding mechanisms revealed by optical tagging of physiologically defined neuronal types

Donghoon Lee, Maiko Kume, Timothy E. Holy*

Neural circuit analysis relies on having molecular markers for specific cell types. However, for a cell type identified only by its circuit function, the process of identifying markers remains laborious. We developed physiological optical tagging sequencing (PhOTseq), a technique for tagging and expression profiling of cells on the basis of their functional properties. PhOTseq was capable of selecting rare cell types and enriching them by nearly 100-fold. We applied PhOTseq to the challenge of mapping receptor-ligand pairings among pheromone-sensing neurons in mice. Together with in vivo ectopic expression of vomeronasal chemoreceptors, PhOTseq identified the complete combinatorial receptor code for a specific set of ligands.

Molecular markers are a powerful tool for labeling and analyzing neuronal cell types. However, in many cases, a single marker is insufficient to define a distinct cell type and may label a few, or a few hundred, physiologically distinguishable cell types (*1*–*3*). In such cases, one wishes to select specific physiological populations and discover their molecular identities. However, tools for proceeding from function to molecular markers are not fully mature. Markers such as *c-fos* label active neurons promiscuously across many cell types (*4*). Cells can be chosen for sequencing based on patch recording (*5*–*8*); however, this approach faces obstacles for rare cell types or when one needs many cells of the same type to profile low-abundance transcripts.

We reasoned that the key bottlenecks of this pipeline could be replaced by an all-optical approach. One prototype is CaMPARI, which uses the simultaneous presence of calcium and excitation by short-wavelength light under the control of the experimenter (*9*). Although CaMPARI allows one to control the labeling of excited neurons temporally, as with *c-fos*, any active neuron will be labeled, so this may comprise dozens or more cell types. There is considerable need for tools that provide the specificity of patch-seq with the convenience and high throughput of optical methods.

Here, we report a method called physiological optical tagging sequencing (PhOTseq). PhOTseq allows specific cell types to be defined intersectionally through their physiological activity under diverse conditions. We exploited two coexpressed fluorescent proteins, GCaMP and photoactivatable mCherry (PAmCherry) (*10*–*12*); GCaMP permits recording neuronal activity by large-scale calcium imaging, and PAmCherry enables long-term labeling of selected neurons by photoactivation (PA). Using these reporters, we performed calcium imaging with a diverse set of stimuli or experiences and then selected a specific collection of cells for tagging by PA. Last, the tagged neurons were harvested and profiled for their mRNA expression (Fig. 1A).

We created several plasmids and viruses, all expressing variants of GCaMP-2A-PAmCherry. The self-cleaving 2A peptide (*13*) allowed both proteins to be expressed from a single mRNA; the tight coupling facilitated ratiometric analysis to compensate for variability in expression. Initial in vitro and in vivo experiments demonstrated the feasibility of performing both calcium imaging and phototagging by using this reporter (figs. S1 and S2). To test the performance of the complete pipeline (Fig. 1A), we turned to mouse vomeronasal sensory neurons (VSNs), whose primary function is pheromone sensing and social communication. VSNs exhibit dense cell packing (*14*) (thus challenging phototagging accuracy), extreme functional heterogeneity, and an unambiguous relationship between molecular identity and physiological function (*15*–*18*). We generated a transgenic tetO-GCaMP5g-2A-PAmCherry mouse and drove expression in all VSNs through olfactory marker protein (OMP)–internal ribosomal entry site (IRES)–tetracyclin transactivator (tTA) (fig. S3A). We observed calcium responses from the VSNs, and two-photon PA resulted in phototagging at single-cell resolution (figs. S3 and S4).

We evoked combinatorial VSN responses using two ligands: 5-androsten-3β, 17β-diol disulfate (A7864) and 1,3,5(10)-estratrien-3, 17β-diol disulfate (E1050). A custom online algorithm automatically segmented responsive neurons; out of 430 neurons responding to at least one ligand, 192 responded to both ligands. Voxels corresponding to these dually responsive neurons were chosen to create a mask so that only these voxels were illuminated during PA (Fig. 1, B to D). Photoactivated neurons were readily distinguished from the background, and PA tagged an average of 1.32 cells per region (Fig. 1E and fig. S5). The photoactivated vomeronasal epithelia (VNEs) were subsequently subjected to dissociation and fluorescence-activated cell sorting (FACS) (Fig. 1, F and G). In the FACS analysis, four to five abundant clusters were observed, of which only one (P1) was almost entirely dependent on PA. This population represents the one selected later for sequencing.

We used PhOTseq to begin unraveling the logic of chemosensation. Fewer than a dozen vomeronasal receptors (VRs) have ligands of known structure (*5*, *19*, *20*), and these are scattered widely across the gene family. We reasoned that a saturation analysis of VRs responsible for encoding a focused region of "chemical space" will provide insights into the molecular logic of chemosensation. First, we examined functional types of VSNs broadly by light-sheet calcium imaging while presenting 15 different sulfated steroids, of which six were also delivered over a range of concentrations, for a total of 27 different chemosensory stimuli. Approximately 15,000 neurons, ~10% of the total VSN population per hemisphere, were visualized in each imaging volume (*3*). From three imaging volumes, more than 5000 neurons responded to at least one of the ligands, and they fell naturally into 20 different response types (Fig. 2, A and B). From this reference clustering, we chose to phototag the most abundant cell type, cluster 1, and three others showing similar chemoreceptive fields (clusters 2, 3, and 6) (fig. S6), among which cluster 3 comprised fewer than 2% of the responsive neurons and presumably ~0.2% of the total VSN population (Fig. 2B). As a control group, we selected one of the most dissimilar cell types, cluster 19, which strongly responded to sulfated pregnanolones.

For cell type selection under two-photon microscopy, on the basis of Fig. 2B, we used a subset of stimuli that distinguish cells among these five selected clusters (figs. S7 and S8 and movie S1). After online two-photon PA, we collected photoactivated or nonphotoactivated experimental control cells by FACS and performed single-cell RNA sequencing. We obtained a total of 622 qualified cells (fig. S9), among which 61, 44, 49, 40, and 65 photoactivated cells were from experiments aimed at functionally defined cluster 1, cluster 2, cluster 3, cluster 6, and cluster 19 cells, respectively (Fig. 2C).

Physiological responses of chemosensory neurons are thought to be largely determined by the VR genes they express (*21*). Therefore, our investigation was focused on the expression of VR genes. Among photoactivated cells, any given cluster exhibited substantial enrichment of at least one VR gene, with different VR genes being enriched in different clusters (Fig. 2C). By contrast, among control cells, VR gene expression was sporadic (fig. S10). Because our targeting accuracy was less than 100% (fig. S5C), we

Department of Neuroscience, Washington University School of Medicine in St. Louis, St. Louis, MO, USA.
*Corresponding author. Email: holy@wustl.edu

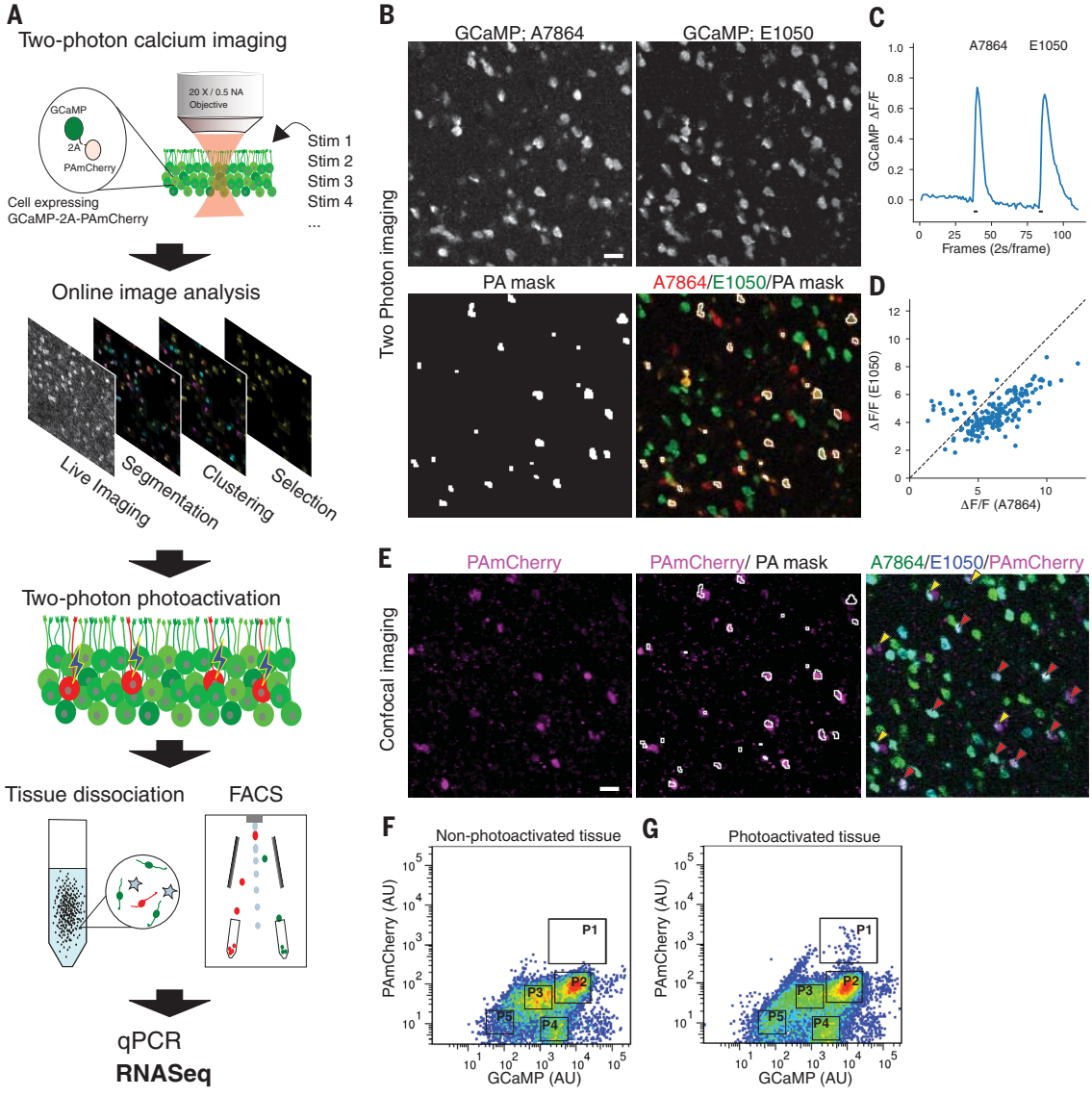



**Fig. 1. Two-photon PA can tag neurons chosen by activity pattern.** (**A**) Workflow of PhOTseq. NA, numerical aperture; qPCR, quantitative polymerase chain reaction; RNASeq, RNA-sequencing; Stim, stimulus. (**B** to **D**) Two-photon calcium imaging of the VNE explanted from a GCaMP5g-p2a-PAmCherry transgenic mouse. (B) Representative images of calcium chemosensory response evoked by either 10 μM A7864 (top left) or 10 μM E1050 (top right). PA mask (bottom) selects cells responsive to both stimuli. Scale bar, 20 μm. (C) The average GCaMP intensity obtained from masked cells. Black bars: delivery time course of stimuli. (D) Individual responses ($\Delta F/F$) among masked cells to 10 μM A7864 and 10 μM E1050. (**E**) Confocal imaging after online two-photon PA. PAmCherry signals (left), PAmCherry signals with two-photon PA mask (middle), and PAmCherry signals with calcium responses to either 10 μM A7864 (cyan) or 10 μM E1050 (green) (right). Red and yellow arrow heads: the examples of PA mask regions associated with single and double PAmCherry-positive neurons, respectively. Scale bar, 20 μm. (**F** and **G**) FACS analysis of cells from nonphotoactivated tissue (F) or photoactivated tissue (G). Clusters were marked as P1, P2, P3, P4, and P5. P1 and P2 represent a photoactivated and an experimental control group, respectively. Of the total population, P1 and P2 accounted for 0.013% and 33.9% (F) or 0.218% and 37.3% (G), respectively. AU, arbitrary unit.

also expected some sporadic expression among cluster-selected photoactivated neurons (Fig. 2C). When the type of a VSN was defined by its maximally expressed VR gene, only one abundant type was found among cluster-selected photoactivated neurons (Fig. 2D) [cluster 1: *Vmn1r89* type (25%); cluster 2: *Vmn1r86* type (30%); cluster 3: *Vmn1r78* type (27%); cluster 6: *Vmn1r237* type (50%); cluster 19: *Vmn1r58* type (58%)]. These five VR genes were the only significantly enriched receptor gene in each photoactivated group compared with all the other cells (Fig. 2E). During these analyses, we updated the *Vmn1r237* gene model because the read coverage of *Vmn1r237* and cloning revealed 3′ untranslated regions missing from existing gene models (fig. S11 and data file S1). We also unexpectedly observed coexpression of *Vmn1r85* in *Vmn1r86* neurons (Fig. 2F and fig. S12, A to C). Including data from our control and sporadic cells, coexpression was also observed for several other VR gene pairs, in each case consisting of a genomically adjacent pair (fig. S12, D to F). Taken together, we identified six putative VR genes mediating a focused set of responses.

To test whether PhOTseq identified true receptor-ligand pairs, we performed a gain-of-function study. Because of abnormal localization of VR proteins (*22*), in vitro heterologous expression systems have had little success, particularly for the V1R family (*23*). We reasoned that a VSN cell's endogenous machinery would allow functional expression of a VR gene; seeking a more efficient route than making a transgenic mouse (*5*), we explored a virus-mediated approach. Among the viral vectors we tested, intravenous injection of rAAV2/8-CAG–green fluorescent protein (GFP) was able to induce expression of GFP in a subset of VSNs (fig. S13). We delivered rAAV2/8-CAG-GCaMP-2A-VR, in which VR was one of the genes identified by PhOTseq, and for each obtained multiple neurons that showed statistically significant responses to at least one ligand (Fig. 3, A and B, and movie S2). In each case, the dominant response pattern matched the expected PhOTseq response pattern (Fig. 3, C and D). We also tested *Vmn1r85* because this gene was found to be coexpressed with *Vmn1r86* (Fig. 2C). The response pattern matched that of cluster 4, which was another group that was similar

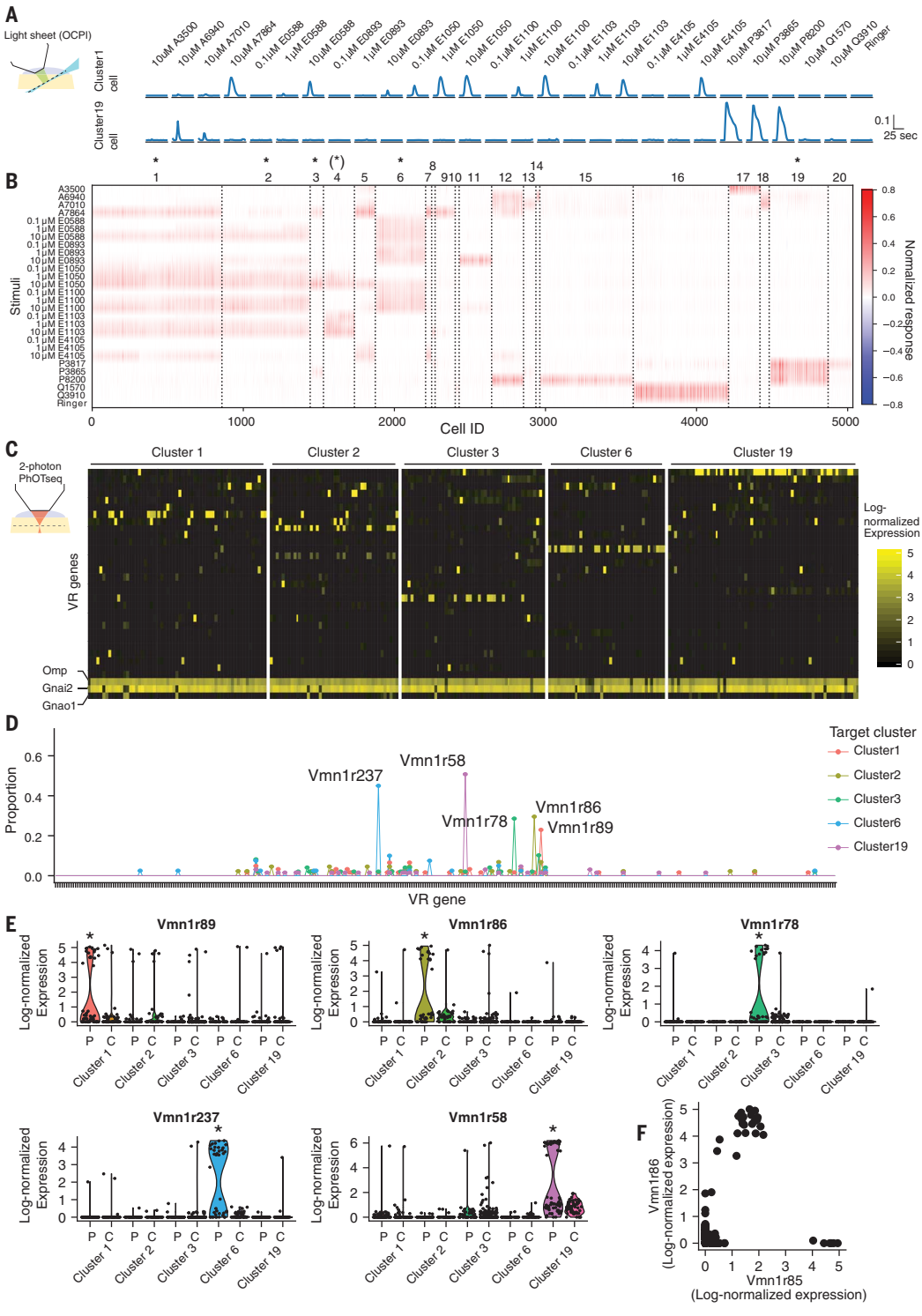



**Fig. 2. PhOTseq-identified VR genes of overlapping cell types.** (**A**) Example calcium traces (normalized $\Delta F$) from two cells. Amplitude was normalized by the maximum amplitude of all cells recorded. Ligands are listed at the top. OCPI, objective coupled planar illumination. (**B**) Neuronal responses to sulfated steroids. Cells are on columns, and stimuli are on rows. If not indicated, the ligand concentration is 10 μM. The color bar indicates normalized response. Cluster identities are reported at the top. Asterisks indicate PhOTseq target cell types. Asterisks in parentheses indicate functional type whose receptor identity was discovered during analysis by ectopic expression in Fig. 3. (**C**) Expression of the 30 most highly expressed VR genes when averaged across all sequenced cells; also shown are three marker genes (*Omp*, *Gnai2*, and *Gnao1*). Cells are on columns, and genes are on rows. PhOTseq-targeted functional types are shown at the top; cells belonging to these functional clusters were specifically photoactivated and sequenced. The color bar indicates the log-normalized expression level. ID, identification. (**D**) Proportion of a VSN type in each group shown in (C). Each tick on the *x* axis represents a different VR gene. Each functional type exhibited only one common VSN type. (**E**) Expression of five VR genes across different experiment groups. "P" indicates photoactivated cells, and "C" indicates nonphotoactivated control cells. *$P_{adj}$ < 0.01 (Wilcoxon rank sum test; $P_{adj}$ < 0.01) (Wilcoxon rank sum test; $P_{adj} = 2.7 \times 10^{-11}$, $2.4 \times 10^{-10}$, $1.1 \times 10^{-20}$, $1.3 \times 10^{-40}$, and $2.7 \times 10^{-30}$ ; average fold difference: 6.2, 23.8, 80.7, 75.7, and 45.9 for *Vmn1r89*, *Vmn1r86*, *Vmn1r78*, *Vmn1r237*, and *Vmn1r58*, respectively). (**F**) Single-cell expression of *Vmn1r85* and *Vmn1r86* is nonexclusive.

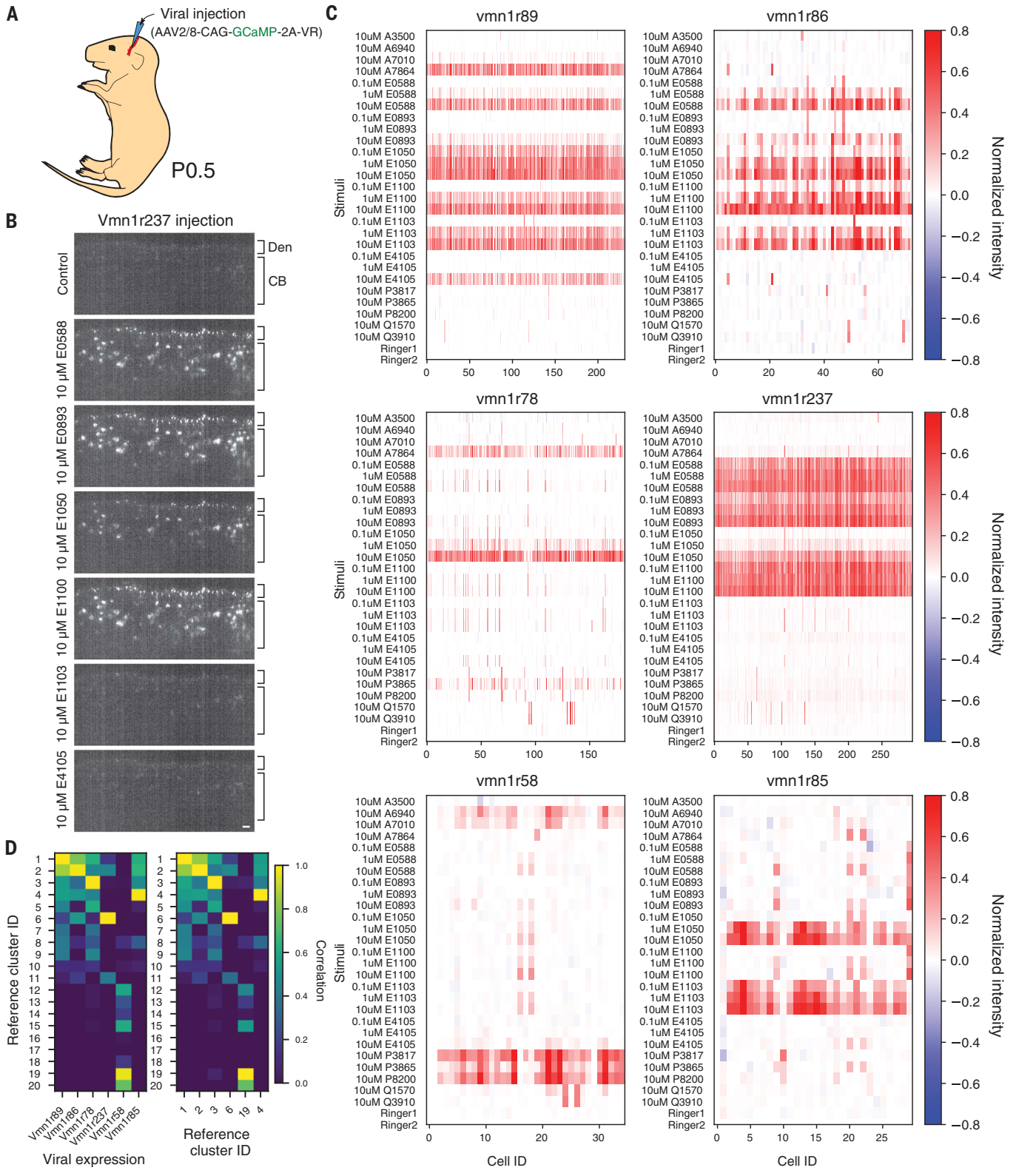
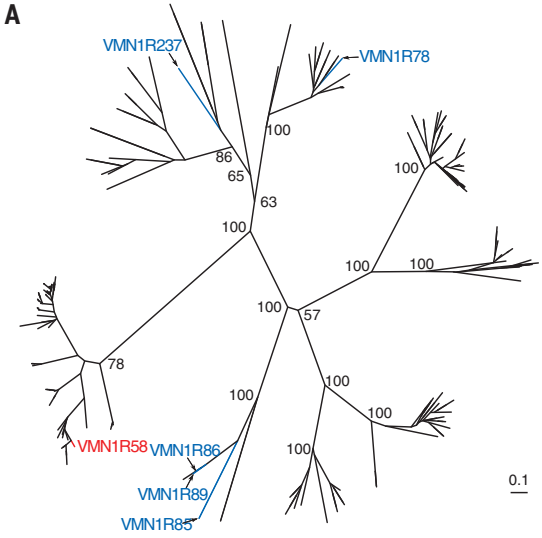
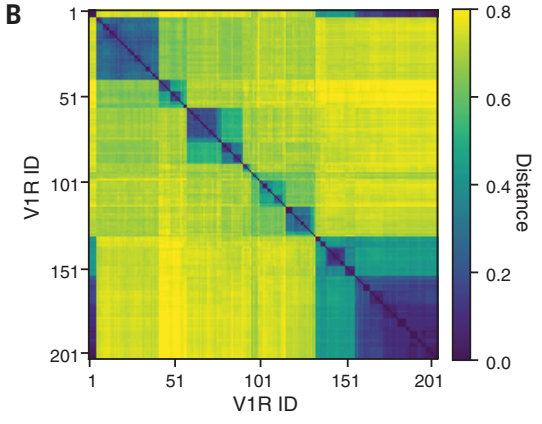



**Fig. 3. Ectopic expression–enabled functional analysis of VRs.** (**A**) Expression by means of AAV injection in the temporal vein of newborn pups. AAV, adeno associated virus; P0.5, postnatal day 0.5. (**B**) Optical section from light-sheet calcium imaging after the ectopic expression of GCaMP-2A-Vmn1r237 in response to different ligands. The apical layer is occupied by dendritic tips (Den) with cell bodies (CB) below. Scale bar, 20 μm. (**C**) Calcium response after ectopic expression of GCaMP-2A-Vmn1r89, -Vmn1r86, -Vmn1r78, -Vmn1r237, -Vmn1r58, or -Vmn1r85. Ligands are identical to those in Fig. 2B. (**D**) Pairwise correlation between the reference clustering and the ectopic responses shown in (C) (left) or autocorrelogram of the reference clustering (right).




C

| Rank | VMN1R89 | VMN1R86 | VMN1R85 | VMN1R78 | VMN1R237 | VMN1R58 |
|---|---|---|---|---|---|---|
| 1 | VMN1R86 | VMN1R89 | VMN1R86 | VMN1R83 | VMN1R236 | VMN1R62 |
| 2 | VMN1R88 | VMN1R88 | VMN1R89 | VMN1R75 | VMN1R233 | VMN1R63 |
| 3 | VMN1R85 | VMN1R85 | VMN1R88 | VMN1R80 | VMN1R234 | VMN1R60 |
| 4 | VMN1R87 | VMN1R87 | VMN1R87 | VMN1R72 | VMN1R235 | VMN1R57 |
| 5 | VMN1R81 | VMN1R81 | VMN1R236 | VMN1R76 | VMN1R225 | VMN1R61 |
| 6 | VMN1R236 | VMN1R236 | VMN1R237 | VMN1R81 | VMN1R226 | VMN1R59 |
| 7 | VMN1R83 | VMN1R82 | VMN1R233 | VMN1R74 | VMN1R66 | VMN1R56 |
| 8 | VMN1R82 | VMN1R84 | VMN1R78 | VMN1R79 | VMN1R224 | VMN1R64 |
| 9 | VMN1R84 | VMN1R215 | VMN1R11 | VMN1R73 | VMN1R78 | VMN1R55 |
| 10 | VMN1R77 | VMN1R15 | VMN1R31 | VMN1R3 | VMN1R67 | VMN1R65 |
| 11 | VMN1R237 | VMN1R237 | VMN1R77 | VMN1R238 | VMN1R69 | VMN1R171 |
| 12 | VMN1R73 | VMN1R77 | VMN1R81 | VMN1R2 | VMN1R184 | VMN1R167 |
| 13 | VMN1R75 | VMN1R83 | VMN1R28 | VMN1R77 | VMN1R228 | VMN1R169 |
| 14 | VMN1R78 | VMN1R73 | VMN1R26 | VMN1R82 | VMN1R232 | VMN1R175 |
| 15 | VMN1R80 | VMN1R70 | VMN1R224 | VMN1R84 | VMN1R76 | VMN1R170 |
| 16 | VMN1R70 | VMN1R233 | VMN1R75 | VMN1R237 | VMN1R81 | VMN1R176 |
| 17 | VMN1R233 | VMN1R79 | VMN1R83 | VMN1R236 | VMN1R82 | VMN1R90 |
| 18 | VMN1R79 | VMN1R201 | VMN1R22 | VMN1R70 | VMN1R68 | VMN1R177 |
| 19 | VMN1R76 | VMN1R216 | VMN1R234 | VMN1R185 | VMN1R185 | VMN1R125 |
| 20 | VMN1R74 | VMN1R78 | VMN1R79 | VMN1R184 | VMN1R79 | VMN1R168 |

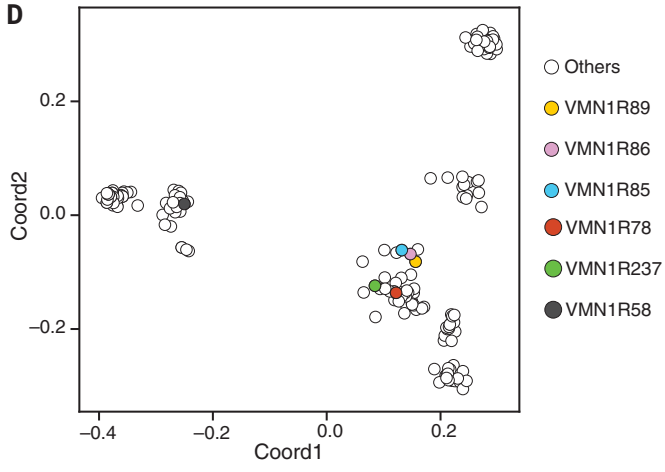



**Fig. 4. Sequence and chemoreceptive similarity are strongly correlated.** (**A**) Unrooted phylogenetic tree of V1R proteins. The five functionally similar (blue) and one distant (red) receptor studied here are marked. The scale bar indicates the number of amino acid substitutions per site. Numbers indicate bootstrap values. (**B**) Pairwise distances among 204 V1R protein sequences (see supplementary materials, materials and methods). (**C**) For these deorphanized VRs, the top 20 nearest VRs, based on (B), are rank ordered. The deorphanized VRs are colored if shown in the table. (**D**) Two-dimensional representation of the distance matrix shown in (B). Each dot represents a single VR protein. coord, coordinate.

to cluster 1 but was not chosen for analysis by PhOTseq (Fig. 3, C and D). Because sequencing revealed control neurons that expressed only *Vmn1r85* (Fig. 2F), functional cluster 4 cells likely expressed *Vmn1r85*.

Only a few studies have examined the relationship between VR sequence and chemosensory function (*5*, *19*) and never from the vantage point of complete knowledge of the VRs that underlie a set of nearest neighbors in terms of function. To investigate this relationship, we performed multiple sequence alignment (MSA) of vomeronasal type 1 receptor (V1R) protein sequences followed by phylogenetic tree analysis (Fig. 4A). Among V1Rs expressed in functionally similar types, VMN1R89, VMN1R86, and VMN1R85 were close to each other in the sense of putative evolutionary distance. Two of the functionally similar cell types, VMN1R78 and VMN1R237, belonged to different branches. In this representation, VMN1R58, which was the chosen outlier in terms of function, was not notably more divergent from VMN1R89, VMN1R86, and VMN1R85 than the more functionally similar VMN1R78 and VMN1R237. Consequently, this tree structure representation did not suggest a particularly close correspondence between function and primary sequence.

However, when the pairwise distances acquired from the MSA analysis were examined carefully, we noticed that, despite their apparent evolutionary distance, VMN1R78 and VMN1R237 were consistently among the VR genes most closely related to VMN1R89, VMN1R86, and VMN1R85 (Fig. 4, B and C). Among 204 V1R sequences examined, the median neighbor rank among these VRs was 11, implying that these VRs are typically among the top-few-percent–closest matches to one another. We therefore considered the possibility that phylogeny, in attempting to construct plausible ancestry, inaccurately models relatedness between sequences that are not nearest neighbors. To more comprehensively evaluate and visualize all pairwise distances, we performed a classical multidimensional scaling analysis to project the high-dimensional sequence-based distance relationships into two dimensions (Fig. 4D). This resulted in a

different picture of the gene family, whose dominant feature was the presence of seven to nine apparent clusters. The functionally similar receptors, including VMN1R78 and VMN1R237, were near neighbors, whereas the functionally dissimilar receptor VMN1R58 was positioned in a distant cluster. We conclude that relationships among the primary sequences of these VRs are strongly correlated with their degree of functional similarity.

Our results demonstrate the utility of PhOTseq, an all-optical solution to the problem of cell type selection and labeling. It provides an opportunity to study the relationships between genes and physiological function even in extremely rare cell types that can be defined only through extensive functional characterization. PhOTseq will be applicable when individual neurons respond combinatorially to various conditions, including sensory stimulation, emotional status, and behavior. We further anticipate that in vivo ectopic expression will be widely applicable for characterizing VR-ligand pairings. In addition, our data show unexpected receptor coexpression, a specific and systematic departure from the "one neuron, one receptor rule" (*21*), whose potential roles will need to be evaluated in many aspects of olfactory function. Lastly, our study provides biological insight on chemosensation by comprehensively mapping receptor-ligand pairings for chosen subsets of the vomeronasal sensory population and suggests that saturation analyses can reveal sequence-function coupling unanticipated by single-ligand studies (*24*, *25*).

## REFERENCES AND NOTES


1. H. Markram *et al*., *Nat. Rev. Neurosci.* **5**, 793–807 (2004).
2. M. Nassar *et al*., *Front. Neural Circuits* **9**, 20 (2015).
3. P. S. Xu, D. Lee, T. E. Holy, *Neuron* **91**, 878–892 (2016).
4. H. Okuno, *Neurosci. Res.* **69**, 175–186 (2011).
5. S. Haga-Yamanaka *et al*., *eLife* **3**, e03025 (2014).
6. J. Fuzik *et al*., *Nat. Biotechnol.* **34**, 175–183 (2016).
7. S. Qiu *et al*., *Front. Genet.* **3**, 124 (2012).
8. C. R. Cadwell *et al*., *Nat. Biotechnol.* **34**, 199–203 (2016).
9. B. F. Fosque *et al*., *Science* **347**, 755–760 (2015).
10. J. Akerboom *et al*., *J. Neurosci.* **32**, 13819–13840 (2012).
11. T.-W. Chen *et al*., *Nature* **499**, 295–300 (2013).
12. F. V. Subach *et al*., *Nat. Methods* **6**, 153–159 (2009).
13. M. D. Ryan, A. M. Q. King, G. P. Thomas, *J. Gen. Virol.* **72**, 2727–2732 (1991).
14. K. C. P. Wilson, G. Raisman, *Brain Res.* **185**, 103–113 (1980).
15. T. Bozza, P. Feinstein, C. Zheng, P. Mombaerts, *J. Neurosci.* **22**, 3033–3043 (2002).
16. P. Feinstein, T. Bozza, I. Rodriguez, A. Vassalli, P. Mombaerts, *Cell* **117**, 833–846 (2004).
17. E. A. Hallem, M. G. Ho, J. R. Carlson, *Cell* **117**, 965–979 (2004).
18. X. Grosmaitre, A. Vassalli, P. Mombaerts, G. M. Shepherd, M. Ma, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1970–1975 (2006).
19. Y. Isogai *et al*., *Nature* **478**, 241–245 (2011).
20. W. M. Wong *et al*., *bioRxiv* 766592 [Preprint]. 12 September 2019. https://doi.org/10.1101/766592.
21. P. Mombaerts, *Nat. Rev. Neurosci.* **5**, 263–278 (2004).
22. J. Loconto *et al*., *Cell* **112**, 607–618 (2003).
23. B. Stein, M. T. Alonso, F. Zufall, T. Leinders-Zufall, P. Chamero, *PLOS ONE* **11**, e0156092 (2016).
24. Y. Jiang *et al*., *Nat. Neurosci.* **18**, 1446–1454 (2015).
25. B. von der Weid *et al*., *Nat. Neurosci.* **18**, 1455–1463 (2015).


## ACKNOWLEDGMENTS

We thank R. Yu for sharing the TetO vector and OMP-IRES-tTA mouse, P. H. Taghert for the HEK cell line, H. Schoknecht for mouse care, and D. W. Kim for developing imaging software. We also thank T. Barnes, J. Chen, J. Corbo, X. Fu, C. Greer, A. Livi, R. Roberts, and M. Zhang for suggestions and comments. This work was supported by the Hope Center Viral Vectors Core and the Mouse Genetics Core at the Washington University School of Medicine in St. Louis. **Funding:** This work was funded by NIH/NIDCD R01 DC005964 and DC010381. **Author contributions:** M.K. performed in vivo imaging and PA. D.L. performed all other experiments and analysis. T.E.H. supervised the study and analysis. **Competing interests:** The authors have no competing interests. **Data and materials availability:** Accession numbers are available in the supplementary materials.

## SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6471/1384/suppl/DC1
Materials and Methods
Figs. S1 to S13
References (*26*–*52*)
Movies S1 and S2
Data File S1
View/request a protocol for this paper from *Bio-protocol*.

25 April 2019; accepted 25 October 2019
10.1126/science.aax8055

**Automated Peptide Synthesizer**
Gyros Protein Technologies introduces PurePep Chorus, its next-generation automated peptide synthesis platform. The scalable instrument and software platform enables production of even the most challenging sequences at the highest-possible crude purity and yield. It has a modular design and can be purchased with—or lab-upgraded to include—two, four, or six channels, with independent induction heating and real-time UV monitoring on two or more channels, depending upon researchers' needs and budgets. PurePep also employs flexible programming and Single-Shot additions, to eliminate the need for priming while minimizing waste when using special amino acids and reagents. The instrument incorporates an intuitive software interface that simplifies the chemist's workflows and accelerates setup with preprogrammed methods, while being fully compliant with 21 CFR Part 11.
**Gyros Protein Technologies**
For info: 800-477-6834
www.gyrosproteintechnologies.com

**Automated Sequencing**
Complex library prep workflows are fully automated on the Ion Chef System to deliver reproducible data on any Ion Torrent sequencer, even from degraded samples such as FFPE DNA, or RNA samples. When paired with the Ion GeneStudio S5 System and Ion AmpliSeq panels, the Ion Chef enables a walkaway, targeted NGS workflow with just 45 min of hands-on time, so that you can focus on finding meaningful answers to your research questions. The Ion Chef simplifies the Ion Torrent NGS workflow by integrating automated template preparation and chip loading in a single instrument for maximum efficiency and throughput. Simply select the run parameters using the intuitive touchscreen interface, load the library samples, reagents, and consumables, and go.
**Thermo Fisher Scientific**
For info: 800-955-6288
www.thermofisher.com

**High-Volume Nucleic Acid Extraction**
The ID NIMBUS Presto assay ready workstation from Hamilton Company offers fast, hands-free nucleic-acid extraction of high-volume evidentiary samples, such as calcified tissues and other challenging samples. Up to 24 or 96 samples in volumes ranging from 50 µL–5,000 µL may be processed simultaneously in microplate or deep-well plate format for use in sensitive downstream analyses. ID NIMBUS Presto includes a deck-mounted KingFisher Presto purification system from Thermo Fisher Scientific, and uses magnetic particle-based technology to facilitate rapid, high-quality purification without the risks of manual variability and errors. Barcode scanning automatically creates a chain of custody to independently verify each sample and correlate it with the case file and assigned analyst. Air-displacement pipetting technology bypasses the risk of liquid-induced contamination; and for superior pipetting accuracy, Hamilton's Compressed O-Ring Expansion (CO-RE) tips are recommended for use with ID NIMBUS Presto. In addition to liquid handling, on-deck gripper paddles use CO-RE technology to transport samples without the need for manual interactions. A user-friendly interface guides analysts through assay setup and displays pertinent information regarding each processing step. Additionally, all robotic actions and sample locations may be downloaded for inclusion in case-file documentation.
**Hamilton Company**
For info: 800-648-5950
www.hamiltoncompany.com/robotics

**Nucleic Acid Purification System**
The Ionic Purification System from Purigen Biosystems is a small benchtop system that uses the company's proprietary isotachophoresis (ITP) technology to extract, purify, and concentrate nucleic acids from biological samples in 1 h with less than 3 min of hands-on time per sample. It enables the automated extraction of nucleic acids with dramatically increased yields and improved purity from a wide range of sample types, including cultured or sorted cells and FFPE tissues. Biological samples are gently lysed, then loaded into the Ionic Fluidics Chip. The Ionic system then applies an electric field to the chip, and the nucleic acids are isolated in their natural, native form using our ITP technology. The nucleic acids are not denatured or dehydrated, and there is no binding or stripping from fixed surfaces. The process minimizes fragmentation and eliminates bead or buffer contamination. The extracted nucleic acids are pure, abundant, and ready for analysis by any downstream technique, including NGS or PCR.
**Purigen Biosystems**
For info: 925-264-1364
www.purigenbio.com/products/ionic-system

**Robotic Liquid-Handling Instrument**
Tecan's Cavro Magni Flex OEM robotic liquid-handling instrument is a modular, scalable solution intended for low- to medium-throughput workflows, offering the flexibility and functionality to quickly and easily create innovative automation systems for virtually any life sciences application, from immunoassay processing to molecular diagnostics. A variety of carriers, cutouts, and custom grids are available to create any number of layouts. The system is available with a choice of two, four, or eight independent pipetting channels. The platform's liquid-handling arm offers variable tip spacing capabilities (independent channels)—to suit different labware types and enable reformatting activities—and can act as both a pipettor and gripper, optimizing use of available space. Supplied with Tecan's MAPlinx development software kit—which includes high-level functional commands for straightforward programming, a worktable editor with virtual library and drag-and-drop features, and a 3D simulator to simplify software testing—the Cavro Magni Flex aims to accelerate instrument development and time to market.
**Tecan**
For info: +41-(0)-44-922-81-11
www.tecan.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.

# 2019 SANOFI - INSTITUT PASTEUR AWARDS

## 8TH EDITION

For the 8th edition, the Sanofi - Institut Pasteur Awards foster scientific excellence by rewarding four researchers whose work demonstrates significant breakthroughs for global health.

SANOFI - INSTITUT PASTEUR
2019 AWARDS
FOR BIOMEDICAL RESEARCH

## INTERNATIONAL SENIOR ADWARD

### Dr. Carl F. NATHAN
Microbiology & Infection

Carl F. Nathan is awarded for his major discoveries in our understanding of host-pathogen relations, such as macrophage activation by interferon gamma which strongly shaped the field of immunology and dramatically influenced treatment of infectious diseases.

## INTERNATIONAL JUNIOR ADWARDS

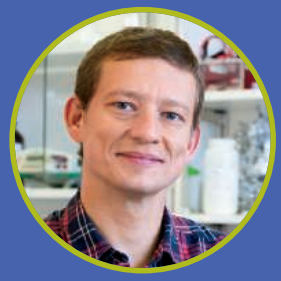
### Pr. Ido AMIT
Immunology

Ido Amit is awarded for his breakthrough single cell studies elucidating the diversity of the immune cell types which revolutionized our view in basic immunology and immunotherapy research.

### Pr. Andrea ABLASSER
Immunology

Andrea Ablasser is awarded for her major contributions to a fundamental discovery in innate immunity: the role of DNA recognition during infection which has inspired efforts to implement strategies interfering with innate immune responses to treat inflammation-related disorders.

### Dr. Marek BASLER
Microbiology & Infection

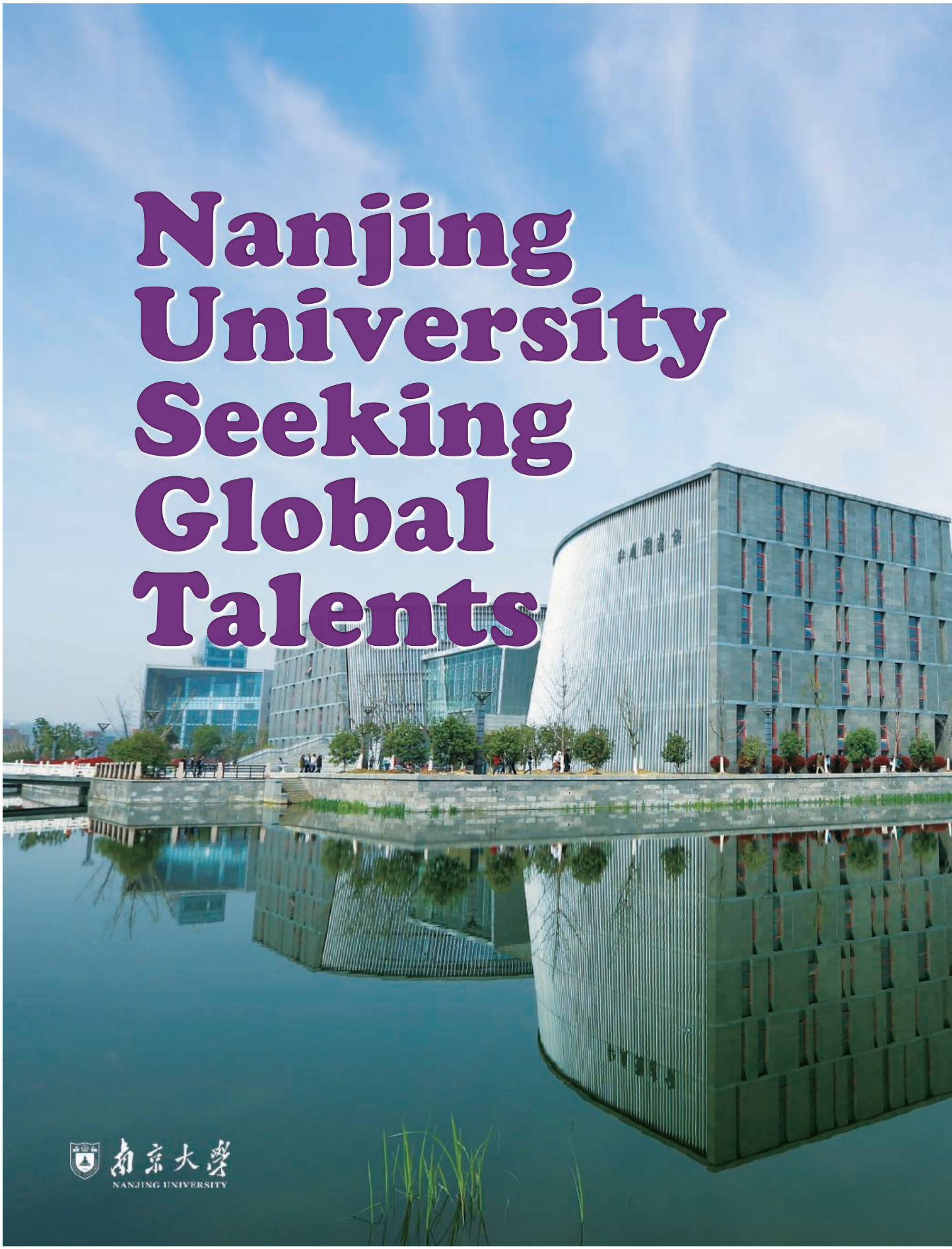


Marek Basler's research has fundamentally advanced our understanding of a very important bacterial nanomachine which delivers effectors in neighboring cells, influencing many areas of microbiology, such as pathogenesis and polymicrobial communities.

SANOFI

Institut Pasteur

FEATURED EMPLOYER
ScienceCareers online @sciencecareers.org




ADVERTISEMENT

Located in the ancient capital of Nanjing, Nanjing University and its predecessors have taken responsibility for the country and the nation during its development of over 100 years. Committed to the prosperity of the nation and the advancement of science, Nanjing University makes significant contributions to the revitalization of China.

Especially since the Reform and Opening-up, Nanjing University has seized this historic opportunity and gained vitality. It was among the first top universities to be selected for prioritized support by the Central Government of China under "211 Project" and "985 Project". NJU was selected by the General Office of the State Council as one of the first group of National Demonstration Bases for Innovation and Entrepreneurship. NJU has been selected as among Category A universities under National Double First Class Initiative, while 15 subject areas of NJU are on the same list. As a member of the C9 League (an alliance of nine elite universities in China), NJU is known for its outstanding faculty, coordinated disciplines, and excellent research, and is recognized as a leading university with academic strength and social prestige. NJU's mission is to cultivate talents, seek truth, serve society, and inherit and transmit Chinese culture. Its aim is to become a world-class research university of Chinese characteristics and NJU style, and to make contributions to the Chinese nation and the international community.

## Distinctive arrangement of disciplines

NJU has built up a well-coordinated development pattern of disciplines by maintaining its strength in foundational disciplines and seeking further improvement in newly-emerging and interdisciplinary areas. The development pattern has given full consideration to development of social, natural, engineering and medical sciences.

Now, NJU has 29 departments under its direct administration. A whole raft of national first-class disciplines with global influence and distinctive features have been built up, which include 8 national first-class key disciplines and 13 national second-class key disciplines, 18 Jiangsu Province First-level Key Discipline□88 undergraduate majors and 38 national doctoral degrees authorized first-level disciplines. NJU ranks the first among China Universities(7th in the world) according to the Nature Index issued in 2019. Besides, QS World University Rankings of 2019 shows that 13 disciplines of NJU rank in the top 100 disciplines of universities around the world. As shown in Essential Science Indicators (ESI) issued by Thomson Reuters, 17 disciplines of NJU rank in the top 1% in the world.( Data statistics as of November 2019).

## Powerful research platform

On the strength of its gold-medal, newly-emerging and interdisciplinary disciplines, NJU has made further headway according to trend of international academic frontiers and national strategic demands. It has established a high-standard research platform represented by a batch of national or provincial-level labs, engineering centers or key research bases. Steady program has been made in the construction of both scientific research and research forces.

Now, NJU has 1 national-level laboratory (under construction), 2 national-level coordination and innovation centers (2011 Program), 7 national key laboratories, 8 key laboratories of the Ministry of Education, 1 national-level engineering research center, 2 engineering centers of the Ministry of Education, 1 engineering and technology research center of the Ministry of Environmental Protection, and 4 humanities and social sciences research centers of the Ministry of Education.

ADVERTISEMENT

南京大學
NANJING UNIVERSITY

ADVERTISEMENT

## The first-class faculty

NJU focuses on pushing forward its development by training of talents. It has been devoted to the introduction and cultivation of talents. To provide powerful, systematic and holistic support for talent training, NJU unites professional forces during its development, cultivates young professionals during entrepreneurship and innovation practices, and shows considerable care for daily life of talents. The result is that a high-standard young faculty with global views and innovative abilities has been established.

In the faculty of NJU, there are 29 academicians of the Chinese Academy of Sciences, 4 academicians of the Chinese Academy of Engineering, 86 distinguished professors out of Chang Jiang Scholars Program, 116 awardees of National Fund for Distinguished Young Scholars, 10 national-level professors. NJU excels among Chinese universities in terms of the above numbers.

## Outstanding students

NJU has been committed to cultivating potential industrial leaders and innovative talents with international views and practical skills. For bachelor program, NJU provide students with individualized courses, independent options, and diversified possibilities. The research program "Double Three Model" has been awarded the Prize in the Seventh National Higher-Education Achievements Award. NJU spares no efforts to push forward "4-3-3 PhD Students Education Reform", planning to establish an ability-oriented interactive training mechanism. Students at NJU are characterized by active mind and strong innovation ability. They have won numerous awards in Chinese and foreign competitions. The graduation employment rate has remained high and stable for many years, and the level of employment and quality have been continuously improved. NJU has been awarded the Prize in Model University for Employment of National Graduates.

## Attractive Campus and Complete Living Facilities

Xianlin Campus of Nanjing University is well known for its notable international elements. With a floor space of about 1.2 million $m^2$ and a green area of some 2.7 million $m^2$, representing the highest standards of construction, modernization and intellectualization in China.

Xianlin Campus has convenient transportation and direct access to Metro Line 2. While providing students with adequate teaching and living facilities, this campus brings out the best of Nanjing University by its human and natural landscapes. The campus has the Jiuxianghe Wetland Park and Yangshanhu Park in its west, the science park in its north, the teaching quarters in its middle, and the Heyuan Faculty Residential Quarters (solely-owned property neighboring two school catchment areas and enjoying teaching resources of surrounding kindergartens and elementary and junior middle schools) on its south. It has been expanded into a large complex which integrates local community, teaching quarters, and science park into a single whole.

Gulou Campus sits in downtown area. As the former site of University of Nanking, it is now listed as a major historical and cultural site protected at the national level. Beidalou Building Groups are the landmark of Nanjing city, which fully displays traditional Chinese architectural styles. As a campus with a history of over a hundred years, it appears quiet and attractive throughout seasons of a years, boasting of rich traditions and spiritual legacies.

## Joining Nanjing University

Nanjing University has established a talent incentive and training system from leading talents, young and middle-aged academic leaders to outstanding young talents. The system matches up with the national talent plan and provides talents of various disciplines and at different levels with promising opportunities.

**Contact: http://zp.ehallapp.nju.edu.cn**

**Tel: 86-025-89686783 / 86-025-89680253 / 86-025-89687256**

ScienceCareers online @sciencecareers.org
OPPORTUNITIES IN CHINA
ADVERTISEMENT

# Geography of Henan University

The College of Environment and Planning (CEP) at Henan University is the successor of the Department of Geography of Zhongzhou University established in 1923. Professor Jinglan Feng, a distinguished geologist and academician of Chinese Academy of Sciences was the first dean of CEP. After years of continuous efforts, CEP has developed to become a comprehensive and interdisciplinary college featuring with Geography as main body, covering Environmental Science, Ecology, Science and Technology of Surveying and Mapping and Regional Economics.

CEP sets doctoral degrees of a first-level discipline in Geography, as well as post-doctoral research program. Besides, there are five doctoral programs of second-level disciplines, including Physical Geography, Human Geography, Cartography and Geographic Information System, Economic Geography, Remote Sensing Information Science and Technology. The college also offers two doctoral programs of second-level disciplines and post-doctoral studies in Regional Economics and Ecology. In the fourth round of national assessment of the geography discipline, CEP is evaluated B+.

CEP has established a number of research and educational platforms, including the National Experimental Teaching Demonstration Center of Environment and Planning, the Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Key Laboratory of the Ministry of Education), Research Institute of Yellow River Civilization & Sustainable Development (Key Research Base of Humanities and Social Science of the Ministry of Education), the Middle and Lower Yellow River Regions Science Data Sharing Platform of National Earth System Science Data Sharing Infrastructure, Regional Economy Research Center of Henan Province(Key Research Base of Humanities and Social Science of Henan Province), the Spatio-Temporal Big Data Industry Technology Research Institute of Henan Province, the Integrative Prevention of Air Pollution and Ecological Safety Key Laboratory of Henan Province.

CEP has a faculty of 160 members, including one academicians of the Chinese Academy of Engineering (CAE), one academician of Inter-national Eurasian Academy of Sciences (IEAS), one Yangtze River Scholars, two Distinguished Young Scholars of National Science Foundation of China (NSFC).

In order to achieve two strategic targets, focusing on natural environment, ecological process and regional high quality development in the middle and lower reaches of the Yellow River, CEP has developed five research fields as human geography, economic geography, Geographic Information System, environmental geography and ecological geography. More than 70 national scientific research projects were granted, 48 research rewards of provincial, ministerial and national level were awarded (including the first-place prize of Henan Province Science and Technology Progress of "the Research and Application of Key Technologies of Earth System Science Data Sharing project" in 2013) and more than 180 articles in SCI, SSCI, EI and other Chinese top journals were published in last three years.

We are eager for ambitious and outstanding researcher to join us from all around the world! The one hundred-year-old Henan University welcomes you! The one thousand-year-old Kaifeng welcomes you! The ten thousand-year-old Henan welcomes you! Let us hand in hand to create brilliant CEP.

**Website:** http://cep.henu.edu.cn
**Email:** psq@henu.edu.cn

ADVERTISEMENT

# Welcome to the first NanjingTech "Exploring and Wisdom" International Young Scholars Forum

**About the Forum**

The first NanjingTech "Exploration and Wisdom" International Young Scholars Forum will be held in Nanjing from December 26 to December 28, 2019.

This forum aims to gather outstanding young scholars from different academic backgrounds at home and abroad, focusing on the integration of cutting-edge-research, such as artificial intelligence and big data, together with traditional industries. The forum aspires to build a stage for talents, exploring the frontiers of science, climbing academic peaks and promoting value creation.

This forum is jointly organized by NanjingTech, Nanjing Jiangbei New Area and China Zhi Gong Party. It consists of a main forum and six sub-forums on life health, intelligent manufacturing and advanced materials, energy and environmental protection, digital finance, smart city and software information. We hereby invite outstanding young scholars at home and abroad to sign up for this event.

**Participants**

(1) Domestic invitees: national standard high level talents or corresponding standard excellent talents.

(2) Overseas invitees: With a Ph.D. from a well-known overseas university; or with a domestic doctoral degree and more than 2 years of continuous overseas research experience. For talents within the humanities and social sciences, or talents in shortage within the science and engineering, the qualification requirements are flexible.

**Application**

If you are qualified and interested, please submit your resume (see Attachment 3 for details) and fill out the "Summary of the First NanjingTech "Exploration and Wisdom" International Young Scholars Forum in 2019" (see Attachment 3 for details). Please send the above materials to: talent@njtech.edu.cn and note the subject of the email as "Name + Forum Participation + Intention College".

**Schedule**

Registration deadline: Dec. 5, 2019
Invitation deadline: Dec. 10, 2019
Forum Date: Dec. 26~28, 2019

**Expenses**

There is no charge for this forum. After you receive the official invitation, please book the tickets by yourselves (economy class air ticket or second-class high-speed rail ticket). NanjingTech will provide free meals and accommodation and offer reimbursement for round-trip transportation expenses, for which original bills are required. (The maximum reimbursement is 15,000 RMB per person for European and the American invitees,etc; 8,000 RMB per person for Asian invitees; 6,000 RMB per person for domestic invitees, including Hong Kong, Macau and Taiwan).

**About NanjingTech**

Nanjing Tech University has a history of more than one hundred years as a cradle of education. It is one of the first batch of 14 universities of the Higher Education Innovative Capacity Promotion Plan ("2011 Plan" for short). It is a key institution of higher learning in Jiangsu Province, which is piloting comprehensive reforms and strategy for developing a strong faculty of talents. Nanjing Tech University is among of the first group of institutions of higher learning approved by the Chinese Ministry of Education for the training of "Excellent Engineers" and the comprehensive reform of graduate education for professional degrees.

Nanjing Tech University is located in the dual-core Jiangbei New District and National Free Trade Zone (Nanjing District) of the National New District, and was recognized as the "The Most Beautiful Campus in Nanjing". Our university has 11 faculties, 29 colleges and more than 30,000 students, one first-rate National Key Discipline, one first-rate Jiangsu Provincial Development Base for National Key Disciplines, two Jiangsu University Development Bases for National Key Disciplines and four Jiangsu Provincial Preponderant Disciplines. It has seven postdoctoral research stations, six primary disciplinary categories that offer 38 doctoral programs in subordinate disciplines, 22 primary disciplinary categories that offer 112 Master Degree programs in subordinate disciplines and 87 undergraduate programs which comprise eight. Nanjing Tech University is ranked 56th in the Chinese mainland universities in ESI 2019(September) and Chemistry, Materials Sciences, Engineering and Biology & Biochemistry are among the top 1% in the world.

With profound cultural heritage, rich entrepreneurial atmosphere, beautiful campus environment, first-class discipline platform, competitive remuneration package and advanced intellectual property system, we sincerely welcome outstanding young talents at home and abroad to join Nanjing Tech University.

**Contacts**

NanjingTech Office of Human Resources
Contacts: Mr. Jin or Mr. Ji
Phone Number: +86-25-58139291, +86-25-58139145
E-mail: talent@njtech.edu.cn

ADVERTISEMENT

北京大學
PEKING UNIVERSITY

# Brain Science at PKU

### Seeking to promote mental health

Peking University Sixth Hospital, or Peking University Institute of Mental Health (IMH), is China's topmost psychiatric hospital. Its director, Lin Lu, an academician of the Chinese Academy of Sciences (CAS), is seeking new therapies for mental and substance use disorders, two major health concerns in China.

Professor Lu and his team have made great efforts to reveal the molecular and circuitry basis of consolidation, storage, destabilization and reconsolidation of pathological emotional memories and proposed a series of novel approaches, including procedures targeting conditioned and unconditioned stimulus induced reconsolidation, to erase memories by making them unstable. Their proposed behavioral and drug treatments have been published in leading peer-reviewed journals, including Science and JAMA Psychiatry.

With passion for science and compassion for patients, Prof. Lu's team has established a new method to eliminate adverse emotional reactions to the commonly used exposure therapy in the unconscious state of sleep, thus significantly relieving patients' distress. Also committed to unraveling the relationship between sleep and mental health, Lu's team has found that sleep disturbances can increase the risk for developing dementia, and sleep disturbances and depression often coexist in the elderly as bi-directionally predicted. Several novel fast-acting antidepressant candidates have been developed to reveal a new glutamatergic mechanism, filling the current treatment gap.

Please feel free to contact Dr. Ping Wu, **wuping@bjmu.edu.cn.**

### Exploring Brain Extracelluar Space

Professor Hongbin Han, the director of a key laboratory at Peking University (PKU),is leading a team working on brain Extracellular Space (ECS), which provides an immediate living environment for neural cells and accounts for approximately 20% of the total volume of a living brain. So far he has developed two measuring methods which can provide biophysical parameters of ECS and image the interstitial fluid (ISF) drainage in the whole brain.

By using the self-developed measuring system, Dr. Han and his team have made a series of discoveries, i.e. brain ISF drains in a compartmentalized ECS system and the diffusion in different ECS divisions are disparate. After verifying the nature of the barrier between divisions, Dr. Han has proposed the hypothesis of "compartmentalized local homeostasis", namely, the brain is protected not only by Blood Brain Barrier (BBB) which avoids exogenous damage through the vascular system, but also by an internal drainage barrier to avoid potentially harmful interference from other divisions. Based on these new findings, a new drug delivery method via ECS has been established, by which the drug arrives directly to the space around neural cells, and overwhelms the impendence from BBB, thus solving the obstacles of low efficiency in traditional drug administration. At present, the new method and findings have been applied in many frontier fields of neuroscience, such as new drug development, aging, aerospace medicine, and artificial neural network modeling.

Please feel free to contact Professor Hongbin Han, **hanhongbin@bjmu.edu.cn**

### Promoting Clinical Translation of Neuroscience

The Neuroscience Research Institute of Peking University, as the core research center for neuroscience at Peking University, is a key laboratory for Neuroscience of the Ministry of Education and the National Health Commission, China. Its history dates back to 1965, when Professor Han Ji-Sheng, now a member of Chinese Academy of Sciences (CAS) and internationally recognized for his research on acupuncture analgesia, began his pioneering work.

At present, the Institute is devoted to bench and bedside studies in both traditional Chinese medicine and modern medicines, with special focus on translation neuroscience. Apart from the neurobiology of acupuncture, the Institute's other research interests include pain and analgesia; drug addiction and its treatment; and neurodegeneration, injury and repair. Publications have appeared in top journals, such as Nature Neuroscience, Nature Communications, and the Journal of Neuroscience. Its basic research outcomes have been translated into clinical devices, such as Han's Acupoint Nerve Stimulator, already widely applied to treat various disorders, such as pain, assisted reproduction, and autism.

The Institute strives for excellence in both research and teaching, and aims to become a nationally leading and internationally recognized academic community of neuroscience. To accomplish this goal, we will continue to embrace advanced techniques, multidisciplinary approaches and collaborative spirit.

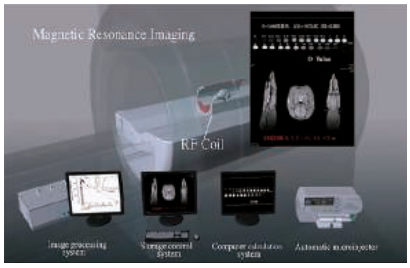
Please feel free to contact Ming Yi, Neuroscience Research Institute, Peking University, **mingyi@hsc.pku.edu.cn**

ADVERTISEMENT

北京大學
PEKING UNIVERSITY

# Exploration of Brain Extracellular Space in China

Professor Hongbin Han is the director of a key laboratory at Peking University (PKU) who is working on brain Extracellular Space (ECS), an incompletely understood and explored area which provides an immediate living environment for neural circuit and accounts for approximately 15%-20% of the total volume in a living brain. Twenty-five years ago, as an interventional radiologist, Han engaged himself in the early diagnosis and treatment of cerebral ischemia. While studying the permeability of blood-brain barrier (BBB) by using dynamic contrast enhanced MRI, he, by accident, acquired a parameter of the space outside the cellular and vascular compartments. It was the first time that the ECS parameter of the human brain was detected and demonstrated by MRI. Since then, for 15 years, he and his team have been working on developing new ECS measuring methods which can visualize the dynamic drainage process of labeled brain Interstitial Fluid (ISF) and obtain ECS parameters in the deep brain. There are two methods, RTI-TMA+ and IOI, which can help acquire the biophysical parameters of brain ECS one or two-dimensionally, while the ECS in the deep brain cannot be detected and analyzed. Tracer-based MRI, the only one based on anisotropic modeling, can acquire and calculate the molecular diffusion



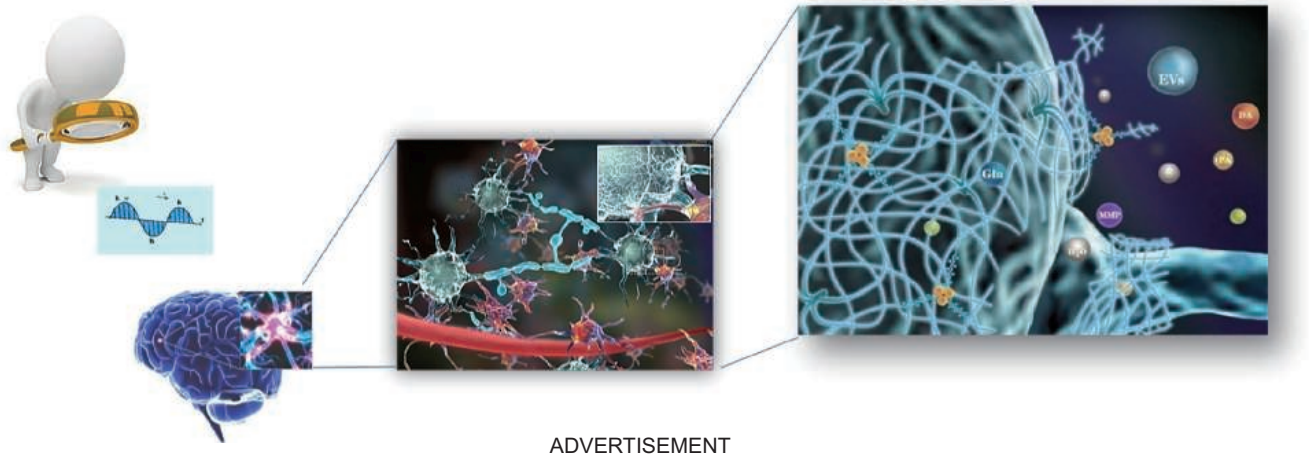

parameters on a large scale and visualize the drainage route of the labeled ISF in the whole brain.

By using their self-developed measuring system, Prof. Han and his team have made a series of discoveries about ECS and ISF. They have found that the brain ISF drains in a compartmentalized ECS system and the biophysical characteristics in separated ECS divisions are different: the traced ISF from the caudate nucleus drains to the ipsilateral cortex along myelin fiber tracts, while in the opposite direction, its movement to the adjacent thalamus is completely impeded by a barrier structure which has been identified as the converged, compact myelin fascicle. After verifying the nature of the barrier structure between different drainage divisions, Prof. Han proposed the hypothesis of "compartmentalized homeostasis" and thus demonstrated that the brain is protected not only by BBB which avoids exogenous damage through the vascular system, but also by an internal ISF drainage barrier to avoid harmful interference from other ECS divisions.

With the new findings and the proposed hypothesis, an innovative therapeutic method for encephalopathy with local drug delivery via the brain ECS has been established. By using this new administration method, the drug is delivered directly to the space around neurons or target regions, overwhelming the impendence from BBB, thus removing the obstacles of low efficiency in traditional drug administration and re-adopting Citicoline, a neuroprotective drug declared ineffective by Nature Reviews Neurology in 2012. As the tracer-based MRI method can accurately calculate and predict the drug transportation and distribution routes in ECS, the location of drug administration and the concentration dose can be precisely designed according to the characteristics of the lesions. The results have shown that the treatment effect, with only 1/800 of the standard dosage, is 6 times better than the traditional method. Furthermore, the efficiency in administering the herbal medicine "Acteoside" via ECS at PKU School of Pharmaceutical Sciences, has also been enhanced. The research was approved as a US patented invention in 2015. Recently, a non-invasive brain ECS imaging technique has been established using MRI, which can scan ECS structure in the human brain.

Today, these new measuring techniques and technologies have been applied in many frontier fields, such as neuroscience, new drug development, development and aging, aerospace medicine, clinical encephalopathy treatment and artificial neural network modeling. Neurologists used to focus only on brain cells and neural circuit, and few achievements have been made in treating AD and stroke patients via the two-compartment model. With the two-compartment model upgraded to the three-compartment model, more breakthroughs in neuroscience are expected to be made in the future.

Please feel free to contact Professor Hongbin Han, **hanhongbin@bjmu.edu.cn**




ADVERTISEMENT

# BNU Seeking Top Talents in Systems Science

**Feel free to contact us:**
Website: http://sss.bnu.edu.cn
Email: sss@bnu.edu.cn
Address: School of Systems Science, Beijing Normal University, No. 19, XinJieKouWai St., Haidian District, Beijing, P.R.China 100875

Beijing Normal University (BNU), as a venerable and dynamic institution, has a long-held tradition of partnering with leading academic institutions from around the world. Noted for its culturally rich and diverse environment, BNU endeavors to become a leader of higher education in China, Asia and even the world.

The School of Systems Science (SSS) at BNU advocates interdisciplinary research in systems science through collaboration with other disciplines in natural, social, technical and economic sciences inside and outside BNU. Having such a uniquely collaborative environment together with outstanding research facilities, SSS aspires to excel internationally in the field of Systems Science.

In 2018, SSS will establish an International Science Center for Complex Systems on the Zhuhai campus of BNU in southern China. This center aims at the frontiers of scientific research and technical innovation, specifically in the formation mechanism of human decision making behavior and its neural mechanism, together with data analysis of human behavior and artificial intelligence (including swarm intelligence). By mobilizing its top experts to lead key research projects in cooperation with talents around the world, this science center is expected to become an international platform for systems science development as well as an interface between academia, industry and governmental authorities.

In order to carry out cutting-edge research in systems science, SSS intends to expand its research team. Its current research portfolio includes, but is not limited to the following fields:

i. The fundamental theories of complex system.

ii. Social and economic system.

iii. The life ecosystem and the self-organizing behavior of brain and cognition.

iv. Multi-agent system and evolutionary algorithm.

v. Information technology of artificial intelligence systems.

vi. The science of science.

*I Think the Next Century Will Be the Century of Complexity.*

***––Stephen Hawking, 2000.***

SSS sincerely welcomes famous scholars, academic leaders, promising young scholars and postdocs in interdisciplinary areas of systems science. The applicant is expected to have a doctoral degree in relevant orientation, a solid research foundation and outstanding research findings in the above-mentioned fields, with great potential for being an excellent teacher and tutor, at the same time an excellent team player in interdisciplinary collaboration.

BNU will place each successful applicant in a proper position according to his/her academic background, research interests and teaching plan, and provide him/her with competitive salary, sufficient start-up fund, necessary laboratory and office space, and other reasonable supports. During the employment, BNU will evaluate each faculty member's capacity and potential and provide the best opportunities for vocational development in line with the international conventions and common practices in academic circles.

Over the years, SSS has made great contributions to the sustainable development of systems science in and outside China. Faced with the complex social and technological challenges, SSS will resume its effort to solve scientific problems in nature and society. By integrating many different yet relevant disciplines and building the international advanced research center of complex systems, SSS strives to cultivate high-level interdisciplinary talents so as to generate new knowledge, new minds, and new technologies that will promote the creation of a shared future for mankind. In time, SSS will become a crucial education and innovation incubator for big data, artificial intelligence, future brain and intellectual education.

SSS cordially welcomes job applicants and visiting scholars with expertise in systems science and related areas. The school also welcomes research and education collaboration from China and the rest of world.

深圳大学
SHENZHEN UNIVERSITY

# Shenzhen University
## High-level Talents from Overseas

Shenzhen University (SZU) was founded as a public university in 1983 located at the Nanshan Peninsula. Facing Hong Kong across the sea, it has undergone rapid growth and expansion in the past 36 years just like the city of Shenzhen, China's most successful Special Economic Zone, and has become a comprehensive university with complete disciplines and facilities, and a strong faculty team. Acclaimed as one of "the most beautiful universities" in China, SZU is committed to becoming a university that manifests the development of Shenzhen, a window of China's reform and opening up and a pilot zone of higher education. In the 2019 World University Ranking Analysis Report released by CUAA (Chinese University Alumni Association), SZU ranks 53rd nationwide, 24th among China's comprehensive universities, and 579th in the world. (CUAA's report is based on the top 10 university rankings including the 2018-2019 QS World University Rankings and US News rankings).

### Disciplines

Consisting of 25 schools and 2 affiliated hospitals, SZU is a multi-disciplinary university covers 11 disciplines including philosophy, literature, economics, law, education, science, engineering, management, medicine, history, and art. Providing 96 undergraduate programs, it now boasts 5 state-level specialty programs, 14 provincial-level specialty programs, and 15 provincial-level key disciplines. It also has 10 first grade doctoral degree conferring disciplines, 9 post-doctoral mobile research stations, and 1 post-doctoral research center.

SZU enjoys rich teaching resources and scientific research facilities. It has established several national engineering laboratories, national engineering technology research centers, key laboratories under the Ministry of Education, international cooperation bases under the Ministry of Science and Technology and provincial key laboratories.

Seven disciplines of the university have entered the top 1% of ESI: Engineering, Clinical Medicine, Material Science, Biology and Biochemistry, Computer Science, Chemistry and Physics among which Engineering, Computer Science and Materials Science have entered the world's top 0.5%.

### Scientific Research

SZU has been deepening the reform of the scientific research system, and the research projects and funding have grown significantly. In 2018, the total research funding exceeded 1.1 billion yuan with 302 new NSFC (National Natural Science Foundation of China) projects and 33 new NSSFC (National Social Science Foundation of China) projects, taking the lead in the country and ranking first in Guangdong. Our faculty published a total of 3,293 papers in SCI indexed journals and 238 papers in SSCI indexed journals, receiving 2,846 scientific research awards.

### Faculty Team

SZU boasts many eminent scholars with a total of 2,387 full-time teachers, 50% of whom are senior professionals.

With the strong backing of Shenzhen municipality, and on the strength of the Guangdong-Hong Kong-Macao Greater Bay Area, SZU will continue to strive to become a global institution, foster high-quality innovative and entrepreneurial talents in the spirit of reform and innovation, strive for cultural leadership to realize innovative and qualitative development, and build the high-level Special Economic Zone University for the people.

### Talent Recruitment Plan in 2019

SZU is seeking applicants with a strong background in all disciplines. We also appreciate the interest of candidates in employment at our growing institution.

**I. Qualifications (meet one of the following conditions):**

1. Applicants must be recognized as outstanding and high-level talents in the field of disciplines at home and abroad;
2. Applicants must have more than 3 years research experience and have formal teaching or research positions in well-known universities and other institutions.

**II. Remuneration**

1. Remuneration: SZU provides competitive remuneration. The successful candidate with outstanding achievements in scientific research will be granted with an annual salary (pretax) up to 1.7 million yuan (about USD 245,000);
2. Research start-up funds: SZU provides funding through various channels. Funding for natural sciences can be up to 8 million yuan, for humanities and social sciences, up to 500,000 yuan. Start-up funds for academicians and outstanding talents with considerable contributions can be determined by negotiations;
3. Competitive benefit package for high level talents in Shenzhen. (Negotiable)
4. Team building: SZU provides preferential policy for the successful candidate to recruit postgraduates, post-doctors and full-time researchers;
5. SZU will assist in the settlement of spouses and children of the successful candidate, and spousal hires may be arranged;
6. Education support: The SZU Affiliated Education Group has kindergarten, primary school and secondary school to provide high-quality education services for the children of faculty and staff;
7. Health care: The Shenzhen University General Hospital (First-Class Hospital at Grade III) provides quality medical services for faculty and staff.

### Contact

For more information, please visit us at www.szu.edu.cn
For further information regarding application procedures, please contact
Ren Qiang at szuad@szu.edu.cn or 0086-755-26535295,
Liang Fuxin at szurcb@szu.edu.cn or 0086-755-26918415.

# Academic Positions Available at Anhui University of Technology, Ma'anshan, China

Anhui University of Technology (AHUT) was founded in 1958. It is a multidisciplinary university with engineering as its focus, while maintaining balanced programs in science, management, humanities, economics, law and arts. The university is located in Ma'anshan, Anhui Province, a city with titles of "National Civilized City", "National Garden City", "Top Quality Environment City", "National Tourism City", and with 40 kilometers to Nanjing, the "Capital City for Six Dynasties in Ancient China". AHUT is a key institution of higher learning in Anhui Province, one of the 100 key universities in central and western China receiving privileged support from the Chinese Ministry of Education (MOE), and also a higher institution listed by MOE for implementing "Outstanding Engineering Education Plan". Its undergraduate teaching is rated "Excellence" in MOE's national teaching-quality assessment.

**Recruitment Requirements**

**A. Top-level Excellent Talents**

***1.Qualifications Required***

1) First-level excellence

Meet the excellence or equivalent level of Academicians of Chinese Academy of Sciences or Chinese Academy of Engineering.

2) Second-level excellence

Meet the excellence or equivalent level of Changjiang Scholarship Specialty Professors or the winner of National Outstanding Youth funds.

3) Third-level excellence

Meet the excellence or equivalent level of listed Candidates of National Ten -Million Talents Project, National candidates of the New Century Talents Program, or listed candidates of other national key talent platforms.

***2.Payment and Welfare***

Negotiable.

**B. Young Excellent Talents**

***1. Requirements***

1) Doctoral degree obtained from well-known universities (institutions)

2) Research experiences forover two years with great potential.

***2.Payment and Welfare***

1) Payment of Professor & Associate Professor;

2) Setting-in allowance: 100,000 RMB;

3) Research start-up fund: 100,000-400,000 RMB;

4) Housing subsidy : 200,000-500,000RMB;

5)Transitional housing with over 90 square meters

6) Spouse's job placement.

**Contact information**

Contact person: Mr. Shi

Tel: 86-555-2311647

E-mail: ahgydxzp@163.com

Address: NO.59 HuDong Road, Ma'anshan City, Anhui Province

# Changsha University
# Dedicated for Regional First-Class University with Global Intelligence

## Faculty Opportunities in Changsha University

Located in the historical and cultural city of Changsha, capital of Hunan Province, Changsha University (CCSU) is proudly designated as application-oriented university in Industry-University Cooperation Program of National "Thirteenth Five-Year Plan". As one of the few newly authorized universities in Hunan Province to establish postgraduate programs and the only municipal-level university in Changsha City, CCSU enjoys both provincial and municipal funding and is principally directed by municipal administration. Boasting 133 hectares garden-like campus, CCSU runs 16 schools (departments) with 44 undergraduate programs, and attracts 1,029 faulty members and over 14,300 full-time students. With decades of dedication, a multidisciplinary development vision in CCSU has been generally realized, featuring a diverse yet integrated structure including science, engineering, law, management, humanities and arts, which is prominently underpinned by engineering application programs with strong support from programs in cultural creativity and modern service fields. To further pursue excellence in both research and innovation, we earnestly seek for qualified candidates worldwide, offering favorable environment and facilities, internationally competitive salaries and benefits.

## Recruitment Disciplines

Civil Engineering, Transport Engineering, Management Science and Engineering, Mechanical Engineering, Power Engineering and Thermophysics Engineering, Material Science and Engineering, Computer Science and Technology, Software Engineering, Information and Communications Engineering, Control Science and Engineering, Mining Engineering, Electrical Engineering, Mathematics, Electronic Science and Technology, Optical Engineering, Instrument Science and Technology, Physics, Cyberspace Security, Chemistry, Biology, Pharmacy, Fisheries Science, Animal Husbandry, Ecology, Drama and Film Studies, Chinese Language and Literature, Journalism and Communication, Art, Law, Foreign Language and Literature, Business Management, Geography, Applied Economics, Public Management, Educational Economics and Management, Architecture, Music and Dance, Marxist Theory, Politics, Philosophy, Historiography, Physical Education, Psychology .

## Requirements, Remuneration and Benefits

| Positions | Requirements and Qualifications | Remuneration and Benefits |
|---|---|---|
| Leading Scholars | Qualified to be listed in "Overseas high-level talents introduction program", China National Funds for Distinguished Young Talents; professors or scholars with the equivalent academic achievements from elite international universities or institutes; senior technical or managerial specialists from world-renowned enterprises or institutions. | 1. Research startup fund: RMB 5,000,000 or above for natural sciences; RMB 1,500,000 for social sciences;<br>2. Housing allowance: RMB 2,000,000.<br>3. Annual salary (pre-tax): RMB 1,150,000 or above. |
| Academic Leaders | Qualified to be listed in "Overseas high-level talents introduction program" for young professionals, National Science Foundation for Outstanding Youth Program; associate professors or scholars with the equivalent or above academic achievements from elite international universities or institutes; senior technical or managerial specialists from world-renowned enterprises or institutions. | 1. Research startup fund: RMB 2,000,000 or above for natural sciences; RMB 800,000 for social sciences;<br>2. Housing allowance: RMB 1,200,000.<br>3. Annual salary (pre-tax): RMB 700,000 or above. |
| Academic Backbones | Elite Chinese young scholars; assistant professors or scholars with the equivalent or above academic achievements from elite international universities or institutes; intermediate technical or managerial specialists from world-renowned enterprises or institutions. | 1. Research startup fund: RMB 1,000,000 or above for natural sciences; RMB 500,000 for social sciences;<br>2. Housing allowance: RMB 800,000.<br>3. Annual salary (pre-tax): RMB 500,000 or above |
| Excellent Doctors and Post Doctoral Fellows | 1. Post-doctor or Ph.D holder, candidates with consecutively 3+ years of working experience in industries or enterprises related to their research fields .<br>2. Comparatively strong and promising research capacity. | 1. Research startup fund: RMB 50,000-100,000 for natural sciences; RMB 30,000-50,000 for social sciences;<br>2. Housing allowance: RMB 300,000-600,000.<br>3. Annual salary (pre-tax): RMB 150,000 or above. |

1. Candidates are required to be able to teach in correct and fluent Chinese;
2. Individual treatment for candidate with outstanding achievements is negotiable;
3. Special treatment for academic team recruitment is negotiable.

Contact person: Mr Nan Mr Li
Telephone: 0731-8426 1360 Mailbox: ccsuhr@163.com Website: www.ccsu.cn
Address: No.98 Hongshan Road, Kaifu District, Changsha City, Hunan Province

Scan the code for application

ScienceCareers online @sciencecareers.org
OPPORTUNITIES IN CHINA

ScienceCareers online @sciencecareers.org
OPPORTUNITIES IN CHINA

流程工业综合自动化
国家重点实验室
STATE KEY LABORATORY OF SYNTHETICAL AUTOMATION FOR PROCESS INDUSTRIES

Please feel free to contact Dr. Jialu Fan or Prof. Jinliang Ding, State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University:
jlfan@mail.neu.edu.cn
jlding@mail.neu.edu.cn

# Automation Science and Technology at NEU

**Seeking to promote automation and intelligence of process industry**

The State Key Laboratory of Synthetical Automation for Process Industries (Northeastern University) is China's topmost lab in information science area. It's director, Tianyou Chai, an academician of the Chinese Academy of Engineering (CAE), IEEE Fellow and IFAC Fellow, is seeking new theory and technology for Industrial AI, main concerns of smart manufactory.

Professor Chai and his team have made great efforts to develop pioneering automation science and technology to address the challenges in a systematic manner, so as to accentuate benefits to the industry. These show novel methods to confront various industry-specific important complexities such as different time scales, extreme process and operational nonlinearities, multiple and strongly coupled operational indices, frequently changing process conditions and widely varying raw material composition. His works emphasize the modeling and feedback control challenges due to lack of methods for on-line quantification of production indices such as energy and material consumptions, product quality and production efficiency; and strongly coupled and often contradictory performance indices involving difficult design and control compromises. Moreover, these discuss difficulties in using existing control methods in this industry due to wide fluctuations in production/process variables, which involve frequent adjustments of multiple set-points. Consequently, the control targets could not be accurately tracked in a timely manner, which would lead to high energy and resource consumptions, and low product quality.

Professor Chai and his team have developed data-driven dual closed-loop intelligent optimal operational control methods. The proposed methods employ a two-layer structure: a control loop setting layer; and a closed-loop control layer. Owing to vast and nonlinear variations in the compositions of raw materials, the control loop setting layer is to generate the time- and process-varying set-points. The closed-loop control layer is to achieve set-point tracking considering virtual unmodeled dynamics compensation.

The proposed methods employ detection of potential faults caused by inappropriately selected set-points, and a self-adjusting strategy that adjusts the set-points to ensure minimal unscheduled process interruptions. Subsequently, they proposed an intelligent closed-loop optimal control method composed of the control loop pre-set model, performance indices forecasting model, feedforward and feedback compensators, fault condition diagnosis and self-adjusting intelligent controllers.

The technology developments have contributed enormously to economic developments in China and a few other countries. These include the largest hematite and magnesia production plants. These have also been implemented for alumina, nickel and cobalt production plants in Vietnam and Papua New Guinea, which have provided significant economic benefits. The technology has helped industry to improve metal recovery rate by 2.01%, grade concentration by 0.57% and energy consumption reduction by 20%, which amounted to increase in operating profit of 6.5 million dollars for the hematite mineral processing enterprise. The applications in the magnesia production enterprise also showed significant economic benefits. These included 5.12% reduction in energy consumption per ton of the material, 2.5% increase in the rate of finished magnesia and 2.7% reduction in electrode dissipation, which culminated in annual profit increase of 5.5million dollars.

**Laboratory Introduction**

The State Key Laboratory of Synthetical Automation for Process Industries (SAPI) pursues fundamental research and advanced cutting-edge technology development to meet the requirements on high automation performance, intelligent and optimal manufacturing in process industries. Technically, SAPI focuses on the development of data-driven control theory and technology, including intelligent data modelling, advanced control and smart decision-making techniques for complex process industries.

SAPI has very strong leadership in academia and has attracted many outstanding academics around the world, including 2 CAE Academicians (one is elected in 2019), 9 Distinguished Young Scholars of Natural Science Foundation of China and 9 Chair Professors of Cheung Kong Scholars Program, 4 NSFC innovative research groups, and 2 Discipline Basements of Innovation and Talent Recruitment (111 Plan).

SAPI has established experimental platforms for teaching and research on the following specific topics such as intelligent data modelling, control and decision-making; smart control system design for process industry, big data processing for industrial monitoring systems, networked control systems and cloud computing; and integrated automation system design.

Along with the associated National Engineering Research Center, SAPI continuously renovates itself and opens a door for rapidly transferring fundamental research outcomes to advanced technologies with applications. Over the past years, SAPI has devoted to developing a theoretical framework for operational optimization and control, which is well recognized internationally. Its technical contributions have been reflected by many best paper awards of top journals and prestigious international conferences. On September 11th, 2013, Vice Premier of China, Yandong Liu, visited SAPI and made the following comments: In regard to the great demand for production automation, the lab has deeply integrated industrialization with information technology, academic study with talent training, and scientific research with industrial applications, and has achieved great innovative accomplishments with great reputation and recognition by international experts. These contributions substantially support the national goal of industrial production automation, which could play an important role in implementation of the 'Four Modernizations' of China.

SAPI carries out extensive and comprehensive international collaboration, having a great impact on the global research and engineering community. Three former presidents of IEEE Control System Society, Prof. Roberto Tempo, Prof. Maria Elena Valcher and Prof. Francesco Bullo, visited SAPI in 2010, 2015 and 2018, respectively, and they all give high appraisal. For example, Prof. Bullo made the comments: "I am especially thankful to Prof. Tianyou Chai for hosting me at Northeastern University in Shenyang, China. His highly regarded State Key Laboratory is a truly unique and impressive center focused on control science and technology transition in the context of automation and greening of complex industrial processes."

Research fellows and PhD graduates in related fields, such as information science, computer, artificial intelligence or mathematics, are warmly welcome to join us. Our mission is to promote industrial automation to industrial artificial intelligence. We will provide competitive salary and start-up funding support, excellent facilities and environment.

ADVERTISEMENT

# 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**
Notheastern University
Xuzhou Medical University
Shanghai University of Electric Power
The First Affiliated Hospital of Jinan University

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, Nanchang University, Xi'an Technological University, Inner Mongolia University of Technology, Inner Mongolia University of Finance and Economics, Inner Mongolia University, Hunan University of Technology and Business, China Three Gorges University, Hubei University of Technology, Wuhan Polytechnic University, Jiangsu Normal University, Northeastern University, Hangzhou Normal 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**

OPPORTUNITIES IN CHINA
Science Careers online @sciencecareers.org

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
贵州医科大学
GUIZHOU MEDICAL UNIVERSITY
南京工业大学
NANJING TECH UNIVERSITY
湖南大学
HUNAN UNIVERSITY
上海外国语大学
SHANGHAI INTERNATIONAL STUDIES UNIVERSITY
天津师范大学
Tianjin Normal University

**Candidates to Research Associates must have a PhD (or equivalent degree). There is no nationality restriction.**

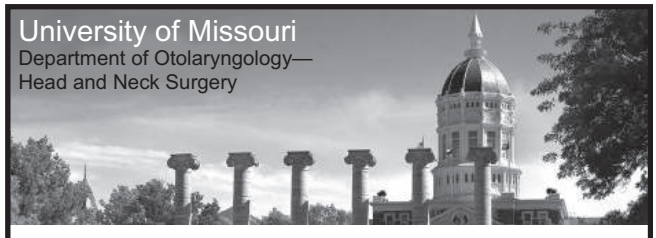
**Inserm is the only French public research institute to focus entirely on human health.**
The Institute brings together 15,000 researchers, engineers, technicians, and administrative staff around one common goal: to improve the health of all by advancing knowledge of life and disease, innovation in treatment, and public health research.
Through its diversity of approaches, Inserm provides a unique environment for researchers.
In order to lead its research policy as effectively as possible, Inserm owns over 350 research structures spread across France and abroad.

**Application modalities: visit ourwebsite:** https://eva3-accueil.inserm.fr
**Application deadline:** January 14th, 2020 - 5.00.pm (GMT+1)

Inserm
La science pour la santé
From science to health

ICA/JM






The University of Missouri (MU) invites aplications for a faculty position in the Department of Otolaryngology-Head and Neck Surgery. Candidates should have an earned doctorate and the ability to lead a federally-funded research group in an area that would complement existing strengths in medical biochemistry. Applications at the Assistant Professor level are encouraged, but candidates at a higher level will be considered if there is evidence of continuous extramural research funding, high impact publications and excellence in teaching and service expected of tenured faculty at a research-intensive university.

For additional information about the position, please contact:
*Robert P. Zitsch III, M.D.*
*William E. Davis Professor and Chair*
*Department of Otolaryngology—Head and Neck Surgery*
*University of Missouri—School of Medicine*
*One Hospital Dr MA314 DC027.00*
*Columbia, MO 65212*
*zitschr@health.missouri.edu*

To apply for a position, please visit the MU website at:
**hrs.missouri.edu/find-a-job/academic/**

*The University of Missouri is an Equal Opportunity/Access/Affirmative Action/Pro Disabled & Veteran Employer. We are fully committed to achieving the goal of a diverse and inclusive academic community of faculty, staff and students. We seek individuals who are committed to this goal and our core campus values of respect, responsibility, discovery and excellence.*



VIMS Virginia Institute of Marine Science | William & Mary

## Endowed Chair in Shellfish Aquaculture and Program Coordinator Associate/Full Professor

The Virginia Institute of Marine Science (VIMS)/School of Marine Science of William & Mary invites applications for a full-time tenure-eligible Associate/Full Professor. The position will begin in Fall 2020.

**Qualifications:** The successful candidate will hold an earned Doctorate (Ph.D.) or an equivalent degree in Marine or Biological Sciences or a related discipline. Candidates must have research interests and demonstrated experience in molluscan shellfish aquaculture. Individual interests could include, but are not limited to, commercial aquaculture methods, larval biology, aquaculture nutrition, aquaculture-environment interactions, aquaculture-society interactions, aquaculture health, genetics and breeding, and physiology. Candidates must have a strong scholarship record commensurate with experience, demonstrated potential to establish an active research program, demonstrated potential for excellence in teaching, and strong communication skills. We are seeking candidates who share VIMS' commitment to the principle that diversity and inclusion are critical to maintaining excellence.

**Responsibilities:** VIMS is seeking to fill an endowed faculty position, the Marshall Acuff Professorship, within its Department of Fisheries Science with a nationally and internationally recognized scholar whose expertise will complement our existing strengths in shellfish aquaculture and provide strategic leadership in coordinating our institute-wide Shellfish Aquaculture Program. The successful candidate will be expected to develop an extramurally-funded research program, contribute to the development of an aquaculture curriculum, participate in advisory service and outreach related to aquaculture, and provide leadership in driving the continued development and sustainability of shellfish aquaculture in the region.

**About the Virginia Institute of Marine Science:** Chartered in 1940, the Virginia Institute of Marine Science is currently among the largest marine research and education centers in the United States. VIMS has a three-part mission to conduct interdisciplinary research in coastal ocean and estuarine science, educate students and citizens, and provide advisory service to policy makers, industry, and the public. The School of Marine Science at VIMS is the graduate school in marine science for William & Mary. VIMS currently employs 55 full-time faculty members and 256 staff, and has 90 graduate students in Master's and Doctoral programs. There are four academic departments at VIMS: Aquatic Health Sciences, Biological Sciences, Fisheries Science, and Physical Sciences. Further information on VIMS and the School of Marine Science may be accessed at: **http://www.vims.edu**.

**Application materials for the position should include:** 1) a cover letter describing professional education, experience, and suitability for the position; 2) a full curriculum vitae; 3) a statement on research interests and teaching philosophy, including a specific statement on how the applicant is interested in and fully committed to diversity and inclusion (3 pages maximum); and 4) the names and titles, institutional addresses, email addresses, and telephone numbers of five references.

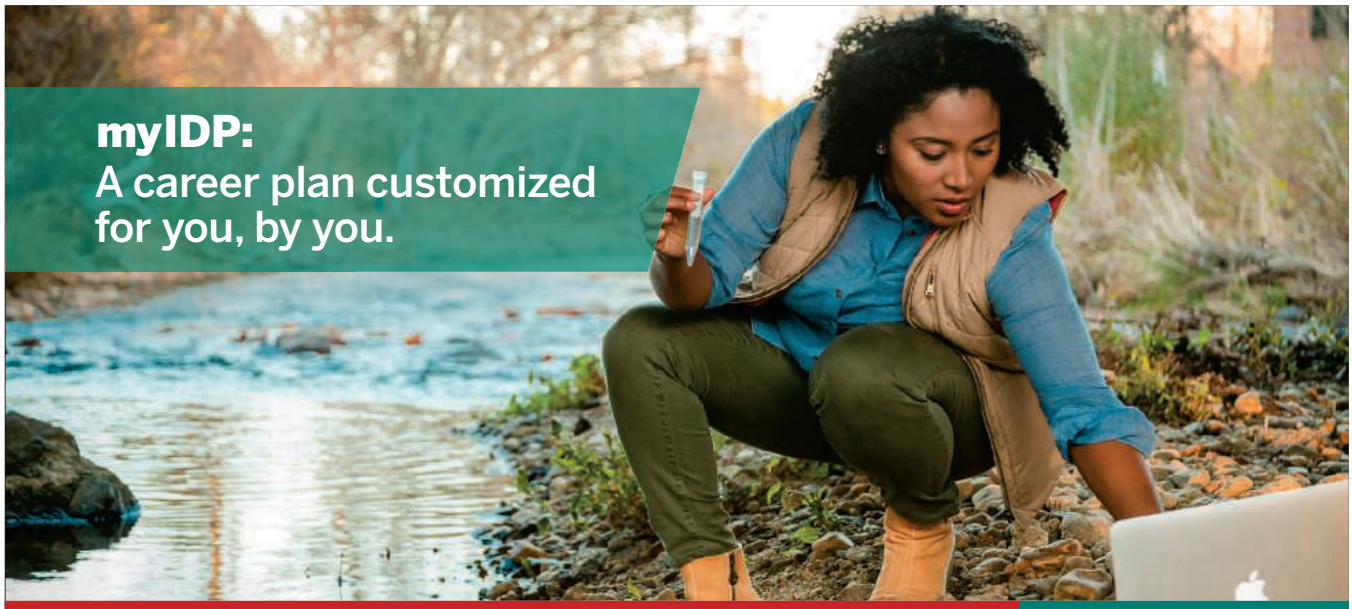
**Application materials should be addressed to:** Search Committee Chair, Shellfish Aquaculture Program Coordinator, and will be accepted through our Online Application System at **https://jobs.wm.edu/postings/37676**

Application materials are due February 2, 2020 for full consideration. Applications received after that due date will be considered if needed.

*William & Mary values diversity and invites applications from underrepresented groups who will enrich the research, teaching and service missions of the university. The university is an Equal Opportunity/Affirmative Action employer and encourages applications from women, minorities, protected veterans, and individuals with disabilities. William & Mary conducts background checks on applicants for employment.*


Science Careers online @sciencecareers.org

ScienceCareers online @sciencecareers.org




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:

FASEB Federation of American Societies for Experimental Biology
UMASS University of Massachusetts Medical School
UCSF University of California San Francisco
UIC
BURROUGHS WELLCOME FUND

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

## Kearns Metcalf and Flint Endowed Chair in Insect Toxicology
## Department of Entomology, School of Integrative Biology

The School of Integrative Biology and the Department of Entomology at the University of Illinois at Urbana-Champaign invite applications from outstanding scholars with a stellar record of accomplishments and an international reputation for excellence to hold the Kearns Metcalf and Flint Endowed Chair in Insect Toxicology, a full-time, 9-month faculty position at the level of full professor with indefinite tenure. Salary is commensurate with qualifications. Toxicology is defined broadly here to encompass the study of any chemical substance with detrimental effects on arthropods and thus candidates working in the areas of chemical ecology, venom biology, insect pathology, plant-insect interactions, chemical signaling, and allied disciplines are welcome to apply, as are candidates working on pesticide mode of action, detoxification, environmental persistence, and resistance evolution. The successful candidate will be expected to develop an externally funded research program, teach at the undergraduate and graduate levels, and collaborate with other faculty to develop research initiatives. The ideal candidate will work at the interface of some combination of toxicology, physiology, molecular genetics, and genomics. A PhD in biology, entomology, plant biology, physiology, molecular genetics, genomics, or other relevant areas is required. The proposed start date is August 16, 2020.

The Illinois College of Liberal Arts and Sciences is a world leader in research, teaching, and public engagement. Faculty in the College create knowledge, address critical societal needs through the transfer and application of knowledge, and prepare students for lives of impact in the state, nation, and globally. To meet these objectives, the College embraces and values diversity and difference through hiring faculty candidates who can contribute through their research, teaching, and/or service to the diversity and excellence of the Illinois community. The University of Illinois is an Equal Opportunity, Affirmative Action employer. Minorities, women, veterans and individuals with disabilities are encouraged to apply. For more information, visit http://go.illinois.edu/EEO. The Department of Entomology has a longstanding reputation for its paradigm-changing research in entomology and its distinguished graduates who have provided leadership in research, teaching, and professional service to the discipline of entomology. Our emphasis is on basic biology, with the goal of discovering new knowledge about the biology of insects as models for understanding basic principles and as organisms whose distinctive attributes have contributed to unparalleled diversity and economic impact. Within entomology, core strengths are in ecology, systematics and evolution, physiology and behavior, genomics, and integrative pest management, with vector-borne disease, invasion ecology, chemical ecology, and pollinator health as major areas of interest. Complementing line faculty are more than a dozen campus affiliates from four Colleges and the Illinois Natural History Survey.

To ensure full consideration please create your candidate profile through **https://go.illinois.edu/KMFChair** and upload your application materials: application letter addressed to search committee chair Dr. Gene Robinson, curriculum vitae, statement of research interests, statement of teaching philosophy and experience, and contact information for three or more professional references by the closing date of **January 20, 2020**. Letters of recommendation may be requested electronically from referees at a later date. Only applications submitted through the University of Illinois Job Board will be considered. For further information, contact Kearns Metcalf and Flint Endowed Chair in Insect Toxicology Search Committee sib@life.illinois.edu or call 217 333-3488. Applicants may be interviewed before the closing date; however, no hiring decision will be made until after that date. The University of Illinois conducts criminal background checks on all job candidates upon acceptance of a contingent offer.


ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

Confused about
your next career move?

Download Free Career
Advice Booklets!

ScienceCareers.org/booklets

## THE UNIVERSITY OF CHICAGO

### Clinical Associate-Gastroenterology #MED122
### The University of Chicago: Biological Sciences Division: Department of Medicine

**Position Description**

The University of Chicago's Department of Medicine, Section of Gastroenterology, Hepatology and Nutrition, seeks a full-time gastroenterologist for appointment as a Clinical Associate for a renewable two-year term. Clinical responsibilities will include seeing outpatients in general GI clinics, and performing outpatient endoscopy, primarily at satellite clinics in Chicagoland area locations. This position does not require teaching or scholarly activity. Compensation (including a generous package of fringe benefits) is dependent upon qualifications and experience.

Prior to the start of employment, qualified applicants must: (1) have an medical doctorate or equivalent, (2) hold or be eligible for medical licensure in the State of Illinois, and (3) be BC/BE in Gastroenterology.

To be considered, those interested must apply through The University of Chicago's Academic Recruitment job board, which uses Interfolio to accept applications: **http://apply.interfolio.com/70000**. Applicants must upload: a CV. Review of applications ends when the position is filled.

For instructions on the Interfolio application process, please visit http://tiny.cc/InterfolioHelp.

**Equal Employment Opportunity Statement**

*The University of Chicago is an Affirmative Action/Equal Opportunity/Disabled/Veterans Employer and does not discriminate on the basis of race, color, religion, sex, sexual orientation, gender identity, national or ethnic origin, age, status as an individual with a disability, protected veteran status, genetic information, or other protected classes under the law. For additional information please see the University's Notice of Nondiscrimination.*

Job seekers in need of a reasonable accommodation to complete the application process should call 773-702-1032 or email equalopportunity@uchicago.edu with their request.

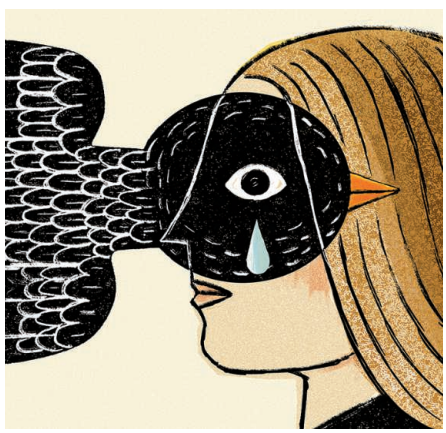
*By* **Eva Kinnebrew**

# Turning a blind eye

This summer, I killed a bird while on a research trip. The project leader and I were sitting under a tree, protected from the hot sun, while we placed leg bands on birds we had just caught. I hadn't handled a bird in a year, and I was nervous that it would escape. I held the bird tightly as it struggled in my hand. Then, as I readjusted my grip, I saw it was lying limp. "Is he OK?" I asked the project leader in a shaky voice. I handed her the bird and she looked down uncomfortably, setting it on the ground. We both stared at it, wishing the wind blowing its chest feathers was actually its heartbeat returning.

I'm not an ornithologist; I'm a Ph.D. student who studies soil biology. But I had worked on birds as an undergraduate student, and I was assisting the project leader, a close friend, with her research. I make the cross-country trip each year to help her and welcome it as a fun change of pace from my usual fieldwork, which involves digging up soil samples from agricultural fields.

This year, though, the experience left me in turmoil. It's not uncommon for animals to die when biologists catch them—and some scientists consider this an unavoidable cost of doing research—but the knowledge that I had taken a life nagged at me. I felt ashamed of my mistake and agonized over the bird for days.



**"As researchers, we must be mindful of the ethical dimensions of our work."**

After I traveled back to my university and reoriented myself to my dissertation research, I realized that I had more on my conscience than the accidental killing of one bird. I regularly capture and preserve huge numbers of soil invertebrates (mites, beetles, and other critters) for my research; this summer alone I killed more than 10,000. I had turned a blind eye to this invertebrate massacre. But as I resumed my normal work, sorting and categorizing hundreds of vials of dead invertebrates, the conflict between my methods and my values increasingly weighed on me.

I face similar struggles in my personal life. Before grad school, I was dedicated to walking or biking around town and buying food from local and organic sources. In grad school, where I'm pressed for time and money, I drive more often and have become increasingly relaxed about my consumer choices. Conventionally farmed bananas shipped thousands of kilometers to Vermont are often the easiest and cheapest snack option I can find, even though eating them while I research sustainable agriculture feels enormously hypocritical.

Scientific research can make a real difference in the world. But I've come to realize that the life of a scientist comes with ethical trade-offs, including ones that aren't easy to stomach—or avoid. Some scientists, such as myself, may debate the ethics of methods that could harm or kill animals. Others may struggle with whether it's the right choice to fly around the world to meetings about climate change.

I entered my Ph.D. program because I care deeply about environmental conservation, and I want to stay grounded and true to those motivations. But I am not perfect. I've made compromises between getting research done and following my personal values. Although these compromises can be necessary and practical, I sometimes look back and wonder whether I cut the wrong corners. The unfortunate reality is that in grad school, moments of reflection are sometimes difficult to come by. Juggling classes and research, jumping through various Ph.D. milestones, and learning technical and professional skills—on top of worrying whether my research is novel, relevant, and impactful—leave me little time and emotional space for much else.

Going forward, though, I want to be more deliberate about these ethical considerations. I plan to ask myself whether my methods are truly necessary and to think more carefully about ways to minimize harm or risk.

The death of this bird hit me as a wake-up call, forcing me to reflect and recalibrate. In grad school, it's easy to be consumed by pressures and demands. But it's important to have moments where we can step back and reconnect with our values. As researchers, we must be mindful of the ethical dimensions of our work. ■

Eva Kinnebrew is a Ph.D. student at the University of Vermont in Burlington. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER



Published by AAAS