

Disentangling overlapping  
memories pp. 1182 & 1227

Larger brains are built differently  
than smaller brains pp. 1254 & 1222

A neural network takes  
in a scene pp. 1168 & 1204

# Science

\$15  
15 JUNE 2018  
sciencemag.org

AAAS

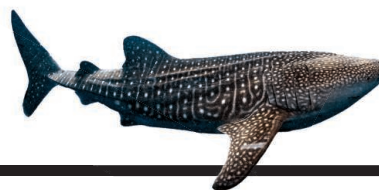
## FAR FROM OVER

Three places where  
"ending AIDS" is a  
distant hope p. 1162





# CONTENTS



1180

Finding rare species

15 JUNE 2018 • VOLUME 360 • ISSUE 6394

## HIV/AIDS: Far from over

### FEATURES

#### 1162 INTRODUCTION

Graphics highlight trouble spots  
By J. Cohen and J. You

#### 1164 THE MOTHER OF ALL CHALLENGES

Nigeria struggles to slow the spread of HIV to babies  
By J. Cohen

1167 Building TRUST in an LGBTQ-hostile country

1168 Babies who dodge HIV may not be unscathed

#### 1170 DARK NIGHTS, BRIGHT STARS

Russia's explosion of HIV reflects stigma and neglect, but it spotlights courageous actions of a few  
By J. Cohen

1171 Status symbol

1172 Poster couple

1174 The pill exchange

1175 The loyal opposition

#### 1176 THE SUNSHINE STATE'S DARK CLOUD

New efforts aim to curb Florida's startlingly high new infection rate  
By J. Cohen

► PODCAST; EDITORIAL P. 1153, VIDEO

### ON THE COVER



Yusuf Adamu, a 12-year-old boy with an advanced case of AIDS, gets an x-ray at Asokoro District Hospital in Abuja, Nigeria. Although the world is getting closer to its goal of "ending AIDS," HIV still wreaks havoc in some places. Powerful antiretrovirals can thwart transmission and extend life span, but many infected people do not have the means to obtain these drugs, owing to poverty and other socioeconomic factors. See p. 1162. Photo: Misha Friedman

#### 1158 QUANTUM PHYSICS COULD GET BIG BOOST FROM U.S. CONGRESS

Pending legislation seeks to counter foreign investment in quantum computing and related technologies  
By G. Popkin

#### 1159 REPORT DETAILS PERSISTENT HOSTILITY TO WOMEN IN SCIENCE

National Academies study urges cultural change to curb multiple kinds of sexual harassment  
By M. Wadman

#### 1160 SEAFLOOR FIBER OPTIC CABLES CAN LISTEN FOR EARTHQUAKES

Method offers a low-cost way to fill seismic detection gaps  
By E. Hand

#### 1161 STRICTER CHINESE STUDENT VISAS RAISE ALARM

Entry permission cut to 1 year for graduate students in some STEM fields  
By J. Mervis

## INSIGHTS

### PERSPECTIVES

#### 1180 A TOOL FOR FINDING RARE MARINE SPECIES

Environmental DNA analysis shows promise for studying rare and elusive marine species  
By E. K. Pikitch

#### 1182 CRYSTALLIZING A MEMORY

Researchers identify and modulate synaptic correlates of a memory engram  
By S. Ramirez

► REPORT P. 1227

#### 1184 SCALING OF HUMAN BRAIN SIZE

Higher cognitive regions are preferentially expanded in individuals with larger brains  
By D. C. Van Essen

► REPORT P. 1222

#### 1185 ANIMALS FEEL SAFER FROM HUMANS IN THE DARK

Mammals shift their activities to twilight and night hours in response to human disturbance  
By A. Benítez-López

► REPORT P. 1232

#### 1186 FACING YOUR FEARS

Activation of remote fear memory promotes fear attenuation  
By P. W. Frankland and S. A. Josselyn

► REPORT P. 1239

#### 1188 UNDERSTANDING SPATIAL ENVIRONMENTS FROM IMAGES

An algorithm that is trained to predict views of a spatial environment also infers its 3D structure  
By M. Zwicker

► REPORT P. 1204

### POLICY FORUM

#### 1189 PREPARING OCEAN GOVERNANCE FOR SPECIES ON THE MOVE

Policy must anticipate conflict over geographic shifts  
By M. L. Pinsky et al.

## NEWS

### IN BRIEF

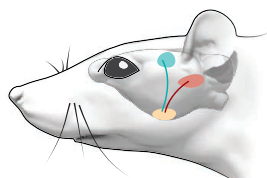
1154 News at a glance

### IN DEPTH

#### 1157 SEAWeed MASSES ASSAULT CARIBBEAN ISLANDS

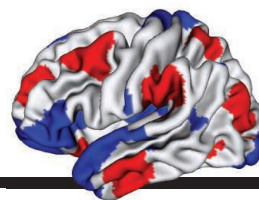
Scientists scramble to explain unusual bloom of *Sargassum* weed  
By K. Langin





1182 &amp; 1227

Synapse-specific memory



1184 &amp; 1222

Brain shape and size

## BOOKS ET AL.

## 1192 ASPERGER'S CHILLING COMPLICITY

The physician behind the eponymous syndrome was an active participant in Nazi eugenics *By M. Wadman*

## 1193 PLATFORMS OF POWER

A Microsoft researcher confronts how companies shape what we see and say online *By A. L. Hoffmann*

## LETTERS

## 1195 CONSERVATION ACCORD: LET COUNTRIES GOVERN

*By M. Chen et al.*

## 1195 CONSERVATION ACCORD: CASH IS NOT ENOUGH

*By J. Ghazoul*

## 1195 CONSERVATION ACCORD: CORPORATE INCENTIVES

*By P. F. E. Addison and J. W. Bull*

## 1196 RESPONSE

*By E. B. Barbier et al.*

## 1197 TECHNICAL COMMENT ABSTRACTS

## 1197 ERRATA

## RESEARCH

## IN BRIEF

1198 From *Science* and other journals

## RESEARCH ARTICLES

## 1201 IMMUNOLOGY

Induction of CD4<sup>+</sup> T cell memory by local cellular collectivity *M. Polonsky et al.*

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

[dx.doi.org/10.1126/science.aaj1853](https://doi.org/10.1126/science.aaj1853)

## 1202 CANCER

Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases *A. Pommier et al.*

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

[dx.doi.org/10.1126/science.aao4908](https://doi.org/10.1126/science.aao4908)

## 1203 NANOTECHNOLOGY

Improving mechanical sensor performance through larger damping *S. K. Roy et al.*

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

[dx.doi.org/10.1126/science.aar5220](https://doi.org/10.1126/science.aar5220)

## 1204 MACHINE LEARNING

Neural scene representation and rendering *S. M. Ali Eslami et al.*

► PERSPECTIVE P. 1188

## REPORTS

## 1210 PHOTOSYNTHESIS

Photochemistry beyond the red limit in chlorophyll f-containing photosystems *D. J. Nürnberg et al.*

## MAGNETISM

1214 Giant tunneling magnetoresistance in spin-filter van der Waals heterostructures *T. Song et al.*

1218 Probing magnetism in 2D van der Waals crystalline insulators via electron tunneling *D. R. Klein et al.*

## 1222 NEUROSCIENCE

Normative brain size variation and brain shape diversity in humans *P. K. Reardon et al.*

► PERSPECTIVE P. 1184; PODCAST

## 1227 NEUROSCIENCE

Synapse-specific representation of the identity of overlapping memory engrams *K. Abdou et al.*

► PERSPECTIVE P. 1182

## 1232 HUMAN IMPACTS

The influence of human disturbance on wildlife nocturnality *K. M. Gaynor et al.*

► PERSPECTIVE P. 1185

## 1235 PLANT SCIENCE

Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle *L. Caputi et al.*

## 1239 NEUROSCIENCE

Reactivation of recall-induced neurons contributes to remote fear memory attenuation *O. Khalaf et al.*

► PERSPECTIVE P. 1186

## 1242 STRUCTURAL BIOLOGY

Near-atomic model of microtubule-tau interactions *E. H. Kellogg et al.*

## 1246 CYTOMETRY

Ghost cytometry *S. Ota et al.*



## DEPARTMENTS

## 1153 EDITORIAL

HIV—No time for complacency *By Quarraisha Abdool Karim and Salim S. Abdool Karim*

► HIV/AIDS: FAR FROM OVER SECTION P. 1162

## 1262 WORKING LIFE

Hitting the wall *By Sharon Ramos Goyette*

|                       |      |
|-----------------------|------|
| Science Staff .....   | 1150 |
| New Products .....    | 1255 |
| Science Careers ..... | 1256 |

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2018 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): \$165 (\$74 allocated to subscription). Domestic institutional subscription (51 issues): \$1808; Foreign postage extra: Mexico, Caribbean (surface mail) \$55; other countries (air assist delivery): \$89. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request. GST #125488122. Publications Mail Agreement Number 1069624. Printed in the U.S.A. Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090-6178. Single-copy sales: \$15 each plus shipping and handling; bulk rate on request. Authorization to reproduce material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that \$35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.



**Editor-in-Chief** Jeremy Berg

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

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

## Research and Insights

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

## News

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

**Executive Publisher** Rush D. Holt

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

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

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

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

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

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

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

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

**SENIOR EDITOR, CUSTOM PUBLISHING** Sean Sanders: 202-326-6430 **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst: 202-326-6463 **ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT** Justin Sawyers: 202-326-7061 [science\\_advertising@aaas.org](mailto:science_advertising@aaas.org) **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **SR. PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle **SR. TRAFFIC ASSOCIATE** Christine Hall **DIRECTOR OF BUSINESS DEVELOPMENT AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu: +86-131 6136 3212, [xchu@aaas.org](mailto:xchu@aaas.org) **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu + 81532-81-5142 [asandhu@aaas.org](mailto:asandhu@aaas.org) **EAST COAST/Y.E. CANADA** Laurie Faraday: 508-747-9395, FAX 617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FAX 415-520-6940 **MIDWEST** Jeffrey Dembski: 847-498-4520 x3005, Steven Loerch: 847-498-4520 x3006 **UK EUROPE/ASIA** Roger Goncalves: TEL/FAX +41 43 243 1358 **JAPAN** Kaoru Sasaki (Tokyo): + 81 (3) 6459 4174 [ksasaki@aaas.org](mailto:ksasaki@aaas.org)

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

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

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

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

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

**INFORMATION FOR AUTHORS** See [www.sciencemag.org/authors/science-information-authors](http://www.sciencemag.org/authors/science-information-authors)

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

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



## EDITORIAL

# HIV—No time for complacency

**T**oday, the global HIV epidemic is widely viewed as triumph over tragedy. This stands in stark contrast to the first two decades of the epidemic, when AIDS was synonymous with suffering and death. But have we turned the tide on HIV sufficiently to warrant directing our attention and investments elsewhere?

Highly effective treatment, through an array of about 30 new antiretroviral drugs, has given people living with HIV an almost normal life span and instilled hope in high-burden communities. Activism played a key role in ensuring global access to treatment. Simplified, low-cost, one-pill-a-day, three-drug combination regimens with minimal side effects are now widely available. The Global Fund to Fight AIDS, Tuberculosis, and Malaria and the U.S. President's Emergency Plan for AIDS Relief support global access to life-saving treatment in poor countries and helped increase treatment coverage from about 2 million people in 2005 to about 21 million in 2017. HIV transmission from mothers to their babies has been reduced to such low levels that plans for its global elimination are well under way.

The AIDS response has now become a victim of these successes: As it eases the pain and suffering from AIDS, it creates the impression that the epidemic is no longer important or urgent. Commitment to HIV is slowly dissipating as the world's attention shifts elsewhere. Complacency is setting in.

However, nearly 5000 new cases of HIV infection occur each day, defying any claim of a conquered epidemic. The estimated 36.7 million people living with HIV, 1 million AIDS-related deaths, and 1.8 million new infections in 2016 remind us that HIV remains a serious global health challenge. Millions need support for life-long treatment, and millions more still need to start antiretroviral treatment, many of whom do not even know their HIV status. People living with HIV have more than a virus to contend with; they must cope with the stigma

and discrimination that adversely affect their quality of life and undermine their human rights.

A further crucial challenge looms large: how to slow the spread of HIV. The steady decline in the number of new infections each year since the mid-1990s has almost stalled, with little change in the past 5 years. HIV continues to spread at unacceptable levels in several countries, especially in marginalized groups, such as men who have sex with men, sex workers, people who inject drugs,

and transgender individuals. Of particular concern is the state of the HIV epidemic in sub-Saharan Africa, where young women aged 15 to 24 years have the highest rates of new infections globally. Their sociobehavioral and biological risks—including age-disparate sexual coupling patterns between teenage girls and men in their 30s, limited ability to negotiate safer sex, genital inflammation, and vaginal dysbiosis—are proving difficult to mitigate. Current HIV prevention technologies, such as condoms and pre-exposure prophylaxis, have had limited impact in young women in Africa, mainly due to their limited access, low uptake, and poor adherence.

There is no room for complacency when so much more remains to be done

for HIV prevention and treatment. The task of breaking down barriers and building bridges needs greater commitment and impetus. Now is not the time to take the foot off the pedal of the global AIDS response if the United Nations Sustainable Development Goal target of a 90% reduction in HIV incidence by 2030 is to be met. The social mobilization and activism that characterized the early HIV response are even more critical now. More effort and investment are needed to support ongoing treatment, increase testing and treatment, deliver effective prevention programs to those most at risk, and expand the available prevention strategies, ultimately with a vaccine and a cure.

—Quarraisha Abdool Karim and Salim S. Abdool Karim



***“Commitment to HIV is slowly dissipating as the world's attention shifts elsewhere.”***



*Quarraisha Abdool Karim (top) and Salim S. Abdool Karim (bottom) are professors at the Mailman School of Public Health at Columbia University, New York, NY, USA, and are the associate scientific director and director, respectively, of the Centre for the AIDS Programme of Research in South Africa (CAPRISA) at the University of KwaZulu-Natal, Durban, South Africa. [quarraisha.abdoolkarim@caprisa.org](mailto:quarraisha.abdoolkarim@caprisa.org); [salim.abdoolkarim@caprisa.org](mailto:salim.abdoolkarim@caprisa.org)*



# NEWS

“With each month that passes, the challenge of energy transition becomes more pressing.”

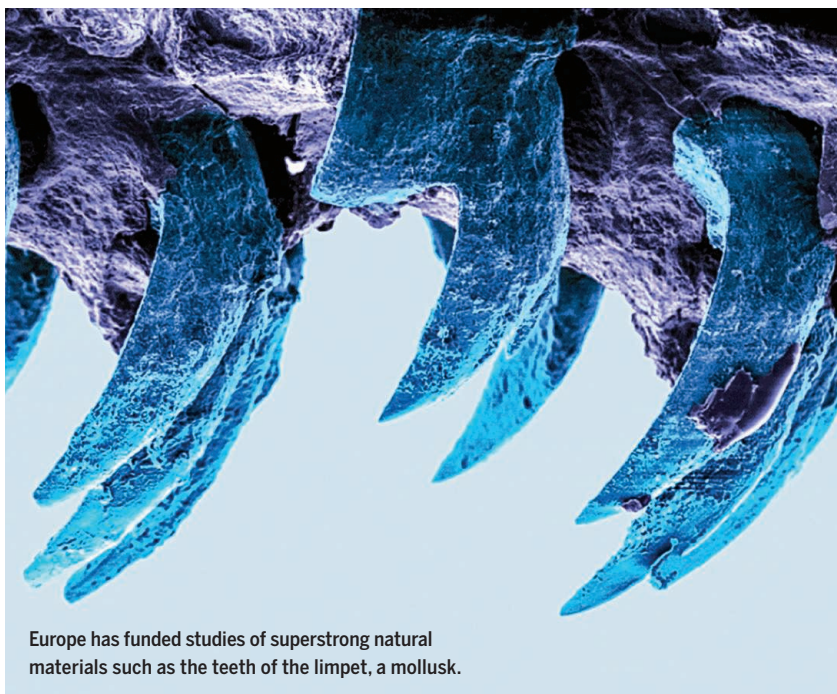
**Pope Francis**, meeting with oil industry executives to encourage reductions in the carbon emissions that cause climate change.

## IN BRIEF

Edited by **Jeffrey Brainard**

### RESEARCH FUNDING

## Fight begins over EU science budget



Europe has funded studies of superstrong natural materials such as the teeth of the limpet, a mollusk.

**T**he European Commission last week proposed a 7-year research budget that universities say favors innovation funding and doesn't include enough to stimulate basic research. The Horizon Europe plan for 2021 to 2027 would provide €94.1 billion—a 22% increase from the current 7-year program, Horizon 2020—of which the commission would allocate €16.6 billion to the European Research Council (ERC), an agency that provides basic research grants. ERC's share of the whole program's budget would remain at about 17%, however; at the same time, the commission wants to spend €10.5 billion—about 11% of the 7-year budget—on the European Innovation Council, a new agency that will fund entrepreneurs to develop new technologies (*Science*, 1 June, p. 952). Universities hoped basic science would benefit more from the overall budget hike. “We will fight for a better distribution of the budget,” said Kurt Deketelaere, secretary-general of the League of European Research Universities in Leuven, Belgium.

## Salk scientist quits under cloud

**WORKPLACE** | The prominent cancer biologist Inder Verma, who allegedly sexually harassed eight women over a 40-year period, resigned from the Salk Institute for Biological Studies in San Diego, California, on 6 June. The institute's board of trustees unanimously accepted his resignation on 11 June. Board Chairman Dan Lewis and Salk President Rusty Gage wrote in a letter to Salk colleagues that Verma tendered his “unconditional” resignation as the board prepared to act on the findings of an investigation of Verma that Salk launched in March. In April, Salk suspended Verma after *Science* sent the institute a list of questions about allegations of sexual harassment against him. Soon after, *Science* described the allegations in detail (*Science*, 4 May, p. 480). Verma denied the allegations in a statement that read in part: “I have never inappropriately touched, nor have I made any sexually charged comments, to anyone affiliated with the Salk Institute.”

## Spain gets science minister

**SCIENCE POLICY** | After visiting space twice, the first Spanish astronaut, Pedro Duque, has a new challenge: leading Spanish science toward recovery. Last week, Spain's new center-left government created a science ministry and placed Duque at its helm, two decisions widely cheered by the country's scientific community, which has become demoralized by declining budgets and bureaucratic obstacles. In 2011, the previous center-right government relegated science to a state secretariat under the economy minister. Duque, 55, is an aeronautical engineer with industry experience, a staunch defender of science and education, and a well-known science communicator. “We're still on schedule to propel Spain to the group of leading countries in science and innovation,” Duque said on 7 June during his installation ceremony.

## Premodern skull surgery worked

**BIOANTHROPOLOGY** | Cranial surgery without modern anesthesia and antibiotics may sound like a death sentence. But the practice was common in the premodern





Trepanation treated injuries such as skull fractures.

world, and survival rates in pre-Columbian Peru were remarkably high, surpassing those of American Civil War soldiers who received cranial surgeries, scientists report. For thousands of years, healers practiced trepanation—the act of drilling, cutting, or scraping a hole in patients' skulls to relieve maladies such as pressure on the brain after a head injury. A study in the June issue of *World Neurosurgery* examined hundreds of skulls from pre-Columbian Peru and suggests techniques and survival rates improved over time, to as high as 83% from the early 1400s C.E. to the mid-1500s C.E. By comparison, about 50% of Civil War soldiers receiving head surgeries died, though their injuries may have been more severe.

## Zika's fake news epidemic

**PUBLIC HEALTH** | Unfounded rumors about the Zika virus were shared online three times as often as factual news articles during an epidemic in 2016 and early 2017, according to a study in the 7 June issue of the *American Journal of Health Education*. Using the content-tracking tool Buzzsumo, researchers at the University of South Florida in Tampa and the Poynter Institute for Media Studies in St. Petersburg, Florida, identified the 120 Zika stories most shared on social networking sites during the 12 months starting in February 2016, when the World Health Organization declared Zika a public health emergency. Rumors were defined as misleading, false, or fabricated content. The most popular stories portrayed Zika, which can damage fetuses, as a conspiracy against the public, a low-risk virus, or a disease caused by pesticides.

## NASA's new leader says he supports climate missions

After a partisan confirmation fight, Representative Jim Bridenstine, a Republican from Oklahoma, was confirmed in April as NASA's newest administrator. Last week, Bridenstine met with reporters to discuss his plans for the agency. (Read a longer version of this interview at <https://scim.ag/QABridenstine>.)

**Q: NASA's next flagship astrophysics mission, the James Webb Space Telescope, has been delayed again. The White House has proposed ending work on its successor, the Wide Field Infrared Survey Telescope (WFIRST). What's the future of NASA astrophysics?**

**A:** At this point, the science that we're going to get back from the James Webb is sufficiently important that we need to finish the project. ... There is hesitation going forward with another massive flagship mission [such as the WFIRST] when you could in many ways have smaller missions that ... can in some cases produce different types of science, but still groundbreaking science.

**Q: NASA has proposed canceling climate-focused missions under development that have been endorsed by the National Academies of Sciences, Engineering, and Medicine's decadal review of the agency's science goals. Will you defend these missions?**

**A:** The 2018 decadal for earth science specifically mentioned CLARREO [the

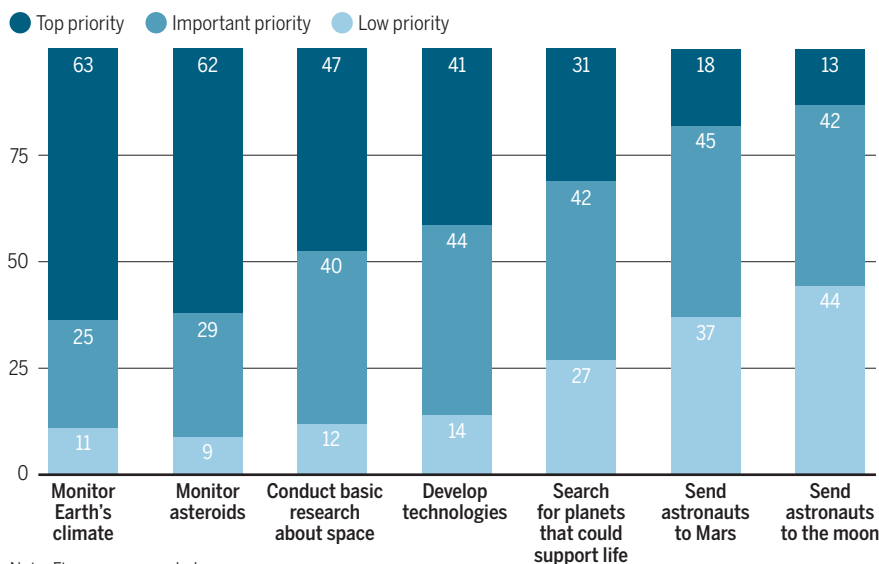
Climate Absolute Radiance and Refractivity Observatory] and PACE [the Plankton, Aerosol, Cloud, ocean Ecosystem satellite; two missions proposed for cancellation]. If we're going to follow the guidance of the decadal ... it seems to me that those projects need to be considered in the president's budget request to fund. ... I think they all [the missions] could end up in good shape.

**Q: The agency is pivoting its near-term focus to the moon, while keeping Mars as its long-term goal. What's your vision?**

**A:** It's going to be sustainable. ... We want to get robots and landers on the surface of the moon, aggressively by 2019. ... Then we want to [develop] those technologies that feed forward to larger landers. ... I've heard people say that every dollar you spend going to the moon is a dollar you're not spending going to Mars. We do not want to go to Mars and have our astronauts be marshmallows. ... We want them to survive. We want them to thrive. We want them to do science. And the way to do that is to use the moon as a proving ground.

## Americans want NASA to focus close to home

NASA's planned astronaut expeditions to the moon and Mars are big-ticket budget items. But Americans place high emphasis on practical missions close to home, such as monitoring climate and detecting asteroids that could hit Earth, according to a survey released 6 June by the Pew Research Center about the future of the U.S. space program.





## Journal confirms stats errors

**CLINICAL TRIALS** | A year after an anesthesiologist in the United Kingdom charged that roughly 2% of clinical trials published from 2000 to 2015 in eight medical journals might be inaccurate, *The New England Journal of Medicine* (NEJM) this week confirmed errors in six such papers it printed. For five, NEJM published corrections because the wrong statistical terms were used: for example, standard deviation, that measures variability across data, in place of standard error, a type of standard deviation that depends on sample size. A months-long inquiry revealed that in the sixth study, a large Spanish trial testing whether a Mediterranean diet could prevent heart disease, investigators did not assign all participants to different diets randomly. NEJM retracted that paper and published a new

version, which still supports the diet. NEJM Editor-in-Chief Jeffrey Drazen in Waltham, Massachusetts, told *Science* that the experience prompted the journal to retrain editors and further scrutinize the statistics of clinical trials before publication. A few other journals whose papers were flagged last year are also investigating.

## Summitting Everest for science

**BIOMEDICINE** | NASA's twin study, which compared changes in gene expression and function in astronaut Scott Kelly and his earthbound identical twin, just got a follow-up in one of the most forbidding environments on Earth. Mountaineers Mike Moniz and Willie Benegas, who summited Mount Everest on 20 May, collected samples of their own blood, saliva, and stool below the summit at an altitude of

7200 meters—the highest sampling done this side of space. The final blood draw is scheduled for this week in Colorado. Their DNA will be sequenced and compared with that of their twin siblings, who stayed at ground level. Geneticist Christopher Mason of Weill Cornell Medicine in New York City hopes to examine the effects of low oxygen levels, among other high-altitude stressors. He says the study isn't ideal—Moniz's twin is fraternal, for example—but could produce "the first map" of what happens to the body on a molecular level at high altitude.

## Mixed verdict on plastic bag rules

**CONSERVATION** | More than 60 countries have adopted bans or taxes on plastic bags, with mixed results, according to a report last week by the United Nations Environment Programme. The report, billed as the first comprehensive look at government actions to control the plastics pollution that is fouling oceans and other environments, found that in about 20 nations—such as Ireland, which began taxing bags in 2002—big drops in their use followed. In other countries, such as Bangladesh, which banned plastic bags the same year, little reduction was observed. The report called on governments to enforce the rules but also support affordable alternatives to plastic bags.

### BY THE NUMBERS

# 232,108

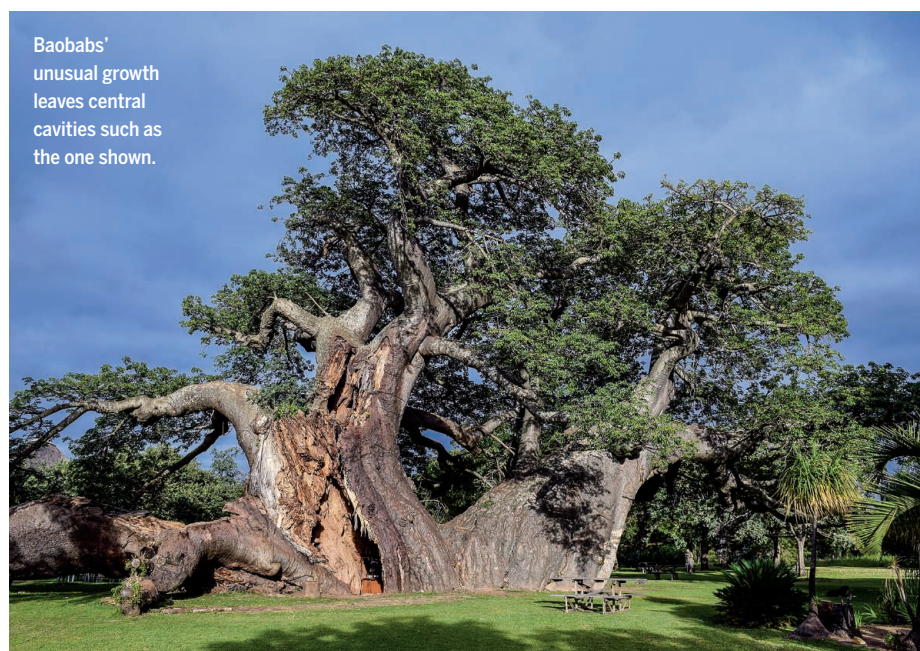
Suicides among 45- to 64-year-old Americans from 1999 to 2016, a period when suicides in that age group rose by 45%—the largest increase in any U.S. population.

# 200

## quadrillion

Calculations per second (200 petaflops) performed by Summit, the world's fastest supercomputer, unveiled last week at Oak Ridge National Laboratory in Tennessee.

**S** **SCIENCEMAG.ORG/NEWS**  
Read more news from *Science* online.



Baobabs' unusual growth leaves central cavities such as the one shown.

### CONSERVATION

## Africa's strange, old baobab trees are dying

**A**frica's baobabs (*Adansonia digitata*) are in trouble. These iconic, unusual trees lack the distinctive growth rings used to age most trees, but Adrian Patrut, a chemist at Babeş-Bolyai University in Cluj-Napoca, Romania, and colleagues developed a radiocarbon technique to date them. Some baobabs are more than 2000 years old—but the biggest, oldest trees are dying for unknown reasons, they report in this week's issue of *Nature Plants*. Eight of the 13 oldest and five of the six largest baobabs died in the past 12 years, the researchers found. The work showed that the older, bigger "trees" are formed by separate trunks that sprout in a circle over several centuries from a single root base. The trunks expand as they grow, and in many cases, eventually fuse, leaving an empty space in the middle of this circle. People have used these enclosures as prisons, churches, even bars. The researchers suspect climate change as a cause of the trees' deaths. One victim: a 501-cubic-meter baobab in South Africa, which was the largest flowering tree in the world.



## IN DEPTH



Deep drifts of *Sargassum* seaweed swept ashore on Guadeloupe in April.

## MARINE SCIENCE

# Seaweed masses assault Caribbean islands

Scientists scramble to explain unusual bloom of *Sargassum* weed

By Katie Langin

In retrospect, 2011 was just the first wave. That year, massive rafts of *Sargassum*—a brown seaweed that lives in the open ocean—washed up on beaches across the Caribbean, trapping sea turtles and filling the air with the stench of rotting eggs. “It presented immense challenges,” says Hazel Oxenford, a fisheries biologist at The University of the West Indies in Cave Hill, Barbados. Before then, beachgoers had sometimes noticed “little drift bits on the tideline,” but the 2011 deluge of seaweed was unprecedented, she says, piling up meters thick in places. “We’d never seen it before.”

Locals hoped the episode, a blow to tourism and fisheries, was a one-off. But a few years later “it came back worse,” Oxenford says. Now, the Caribbean is bracing for what could be the mother of all seaweed invasions, with satellite observations warning of record-setting *Sargassum* blooms and seaweed already swamping beaches. Last week, the Barbados government declared a national emergency. “It’s catastrophic,” says James Franks, a marine biologist at The University of Southern

Mississippi in Ocean Springs who is one of many scientists trying to explain why a part of the ocean that was once seaweed-free is now rife with *Sargassum*. “Right now there’s [another] huge mass impacting Puerto Rico, and that’s the last thing they need.”

Before 2011, open-ocean *Sargassum* was mostly found in the Sargasso Sea, a patch of the North Atlantic Ocean enclosed by ocean currents that serves as a spawning ground for eels. So when the first masses hit the Caribbean, scientists assumed they had drifted south from the Sargasso Sea. But satellite imagery and data on ocean currents told a different story.

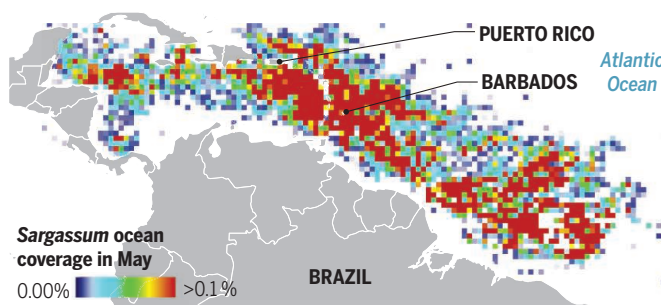
Jim Gower, a remote-sensing expert with Fisheries and Oceans Canada in Sidney, and his colleagues looked for spots on the ocean’s surface that reflected unusual amounts of near-infrared light, a part of the spectrum that plants don’t harvest. Data from May 2011 showed a huge patch of floating plants, presumably *Sargassum*, off the coast of Brazil—far to the south of its normal habitat. By September, it stretched from the Caribbean all the way to the coast of Africa, the team reported in 2013.

To confirm that the *Sargassum* fouling Caribbean beaches in 2011 came from the tropical Atlantic Ocean, east of Brazil, Franks and his colleagues traced the likely path of seaweed masses backward through time. First, they compiled records of locations where the seaweed came ashore. Then, using information about surface currents, they calculated its likely source. “Invariably, in all of those instances, it tracked back to the [tropical] region,” says Franks, who reported the findings in 2016. “None of it ever tracked northward into the Sargasso Sea.”

Since 2011, tropical *Sargassum* blooms have recurred

## Sea of trouble

A satellite map from May shows abundant *Sargassum* across a swath of the Caribbean Sea and tropical Atlantic Ocean that was largely free of floating seaweed until 2011.





nearly every year, satellite imagery showed. The new source region is encircled by currents running clockwise from South America to Africa and back again. From January to May that loop breaks down and westward flows sweep the seaweed up the Brazilian coast toward the Caribbean. "All along the way, the *Sargassum* is blooming and growing," Franks says.

Yet in satellite data prior to 2011, the region is largely free of seaweed, says Chuanmin Hu, an oceanographer at the University of South Florida in St. Petersburg and senior author of a 2016 study that examined satellite data from 2000 to 2015. That sharpens the mystery of the sudden proliferation. "Nobody has a definite answer," Hu says. Nutrient inputs from the Amazon River, which discharges into the ocean around where blooms were first spotted, may have stimulated *Sargassum* growth. But other factors, including changes in ocean currents and increased ocean fertilization from iron in airborne dust, are equally plausible. It's all "educated speculation," Hu says.

Some clues may come from the seaweed itself. Most of the *Sargassum* that's blooming off the coast of Brazil has broader

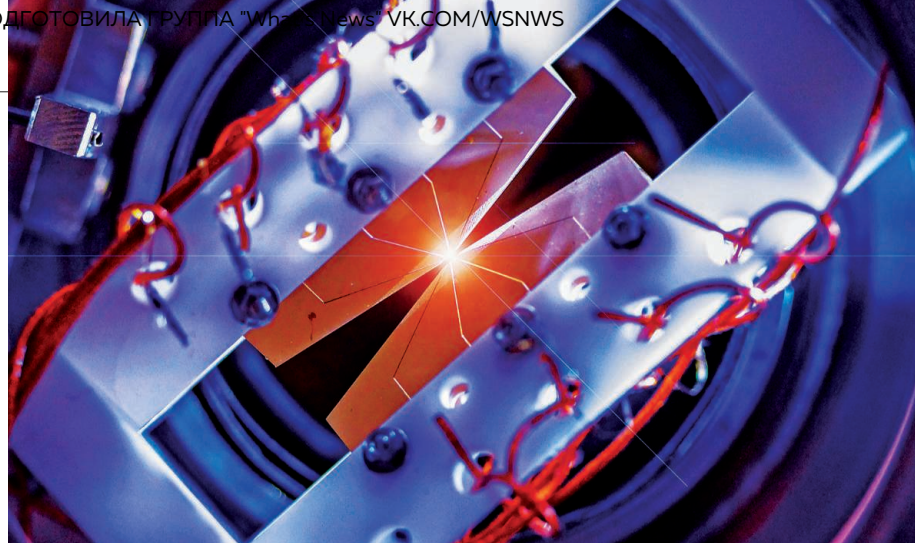
***"Right now there's [another] huge mass impacting Puerto Rico, and that's the last thing they need."***

**James Franks,**

The University of Southern Mississippi

blades than the weed in the Sargasso Sea, says Amy Siuda, a biological oceanographer at Eckerd College in St. Petersburg who has examined hundreds of samples. Siuda is now working with colleagues to determine whether the broad-blade form—which historically was rare in tropical areas—is a separate species. Figuring out the conditions that promote its growth could be key to explaining the proliferation, she says.

In the meantime, the Caribbean is struggling to cope as yearly bouts of *Sargassum* become "the new normal," says Iris Monnereau, a regional project coordinator for the Food and Agriculture Organization of the United Nations in Christ Church parish in Barbados. She's particularly worried about this year, noting that the blooms visible in satellite imagery dwarf those of previous years. "You can't solve the problem; you can't put up a wall or anything," she says. "It's difficult to go forward." ■



## POLICY

# Quantum physics could get big boost from U.S. Congress

Pending legislation seeks to counter foreign investment in quantum computing and related technologies

By **Gabriel Popkin**

**M**any members of Congress admit they find quantum physics mind-boggling, with its counterintuitive account of the subatomic world. But that isn't stopping U.S. lawmakers, as well as policymakers in President Donald Trump's administration, from backing an emerging effort to better organize and boost funding for quantum research, which could reshape computing, sensors, and communications.

In the coming weeks, the science committee of the House of Representatives is expected to introduce legislation calling for a new, 10-year-long National Quantum Initiative (NQI). The White House, for its part, is scheduled to formally launch a new panel that will guide the federal government's role in quantum science. Key science agencies are calling on Congress to accelerate spending on quantum research. And the Senate supports a boost for the field: Last week, it approved a mammoth defense policy bill that includes a provision directing the Pentagon to create a new \$20 million quantum science program.

A yearlong push by a coalition of academic researchers and technology firms helped trigger this flurry of activity. Proponents argue the United States needs a better plan for harvesting the potential fruits of quantum research—and for keeping up with global competitors. The European Union has launched a decadelong quantum research

initiative, and China is said to be investing heavily in the field. The United States is "kind of the only major country that's not doing something," says Chris Monroe, a physicist at the University of Maryland in College Park and co-founder of a startup developing quantum computers, which could outstrip conventional computers on certain problems.

Last June, a small group of academics, executives, and lobbyists that includes Monroe released a white paper calling for an NQI; they issued a blueprint for the effort in April. Meanwhile, the House science committee held a hearing on the topic last October and expects to release a bill later this month that draws extensively from the blueprint.

The legislation will authorize the Department of Energy (DOE) and the National Science Foundation (NSF) to create new research centers at universities, federal laboratories, and nonprofit research institutes, according to a committee spokesperson. These research hubs would aim to build alliances between physicists doing fundamental research, engineers who can build devices, and computer scientists developing quantum algorithms. The centers could give academics seeking to develop commercial technologies access to expertise and expensive research tools, says physicist David Awschalom of the University of Chicago in Illinois, one of the blueprint's authors. "The research needs rapidly outpace any individual lab," he says.

The proposal "sounds really promising," says Danna Freedman, a chemist at Northwestern University in Evanston, Illinois,

lons trapped between gold blades serve as information-carrying qubits in a prototype quantum computer.

who did not contribute to the proposal. But Freedman, who synthesizes materials that could be used to build new kinds of quantum computer components, says her enthusiasm "depends to what extent the government decides to prescribe the research."

The blueprint recommends that the hubs focus on three areas: developing ultraprecise quantum sensors for biomedicine, navigation, and other applications; hack-proof quantum communication; and quantum computers. The bill will likely leave it up to federal agencies, the new White House quantum panel, and an outside advisory group to determine the initiative's focus. Backers also say the effort could help advance the development of software for quantum computers—a major hurdle. Right now, just "tens or hundreds of people" can program quantum computers, says William Zeng of Rigetti Computing, a startup in Berkeley, California, seeking to build a quantum computer and offer quantum computing services. "That's not going to be able support building the full potential of the tech."

It's not yet known how much funding the House bill, which Republicans on the science panel are crafting, will recommend. The blueprint envisions channeling \$800 million over 5 years to the NQI, but even if the bill endorses that figure, congressional appropriators will have the final say. Also uncertain is whether Democrats will sign on and help ensure passage through the full House, and whether the Senate will support the idea.

In the meantime, lawmakers and the Trump administration are moving to shore up federal spending on quantum science, which analysts in 2016 estimated at about \$200 million a year. Adding to the \$20 million boost approved by the Senate (but not yet by the entire Congress), Trump's 2019 budget request would create a new \$30 million "Quantum Leap" initiative at NSF and boost DOE's quantum research programs to \$105 million.

The United States, long seen as a leader, is facing growing global competition in the quantum field, says Walter Copan, director of the National Institutes of Standards and Technology in Gaithersburg, Maryland, which has long played a role in quantum research. "It is the equivalent of a space race now," says Copan, who met last week with House science committee chair Lamar Smith (R-TX). Focusing federal resources on the field, Copan adds, "has phenomenal promise for the country—if it's done right." ■

*Gabriel Popkin is a science journalist in Mount Rainier, Maryland.*

## SCIENTIFIC COMMUNITY

# Report details persistent hostility to women in science

National Academies study urges cultural change to curb multiple kinds of sexual harassment

By Meredith Wadman

Ask someone for an example of sexual harassment and they might cite a professor's insistent requests to a grad student for sex. But such lurid incidents account for only a small portion of a serious and widespread harassment problem in science, according to a report released this week by the National Academies of Sciences, Engineering, and Medicine. Two years in the making, the report describes pervasive and damaging "gender harassment"—behaviors that belittle women and imply that they don't belong, including sexist comments and demeaning jokes. Between 17% and 50% of female science and medical students reported this kind of harassment in large surveys conducted by two major university systems across 36 campuses.

"We are trying to bring to the fore the concept of gender harassment," says anthropologist Kate Clancy of the University of Illinois in Urbana, an author of the report. "The vast majority of sexual harassment that occurs is sexist hostility and crude behavior. And the literature supports that these everyday experiences may have as bad or worse personal and professional consequences as things like unwanted sexual advances."

Decades of failure to curb sexual harassment, despite civil rights laws that make it illegal, underscore the need for a change in culture, the report says. "We have been addressing this problem for a long time. And we have not made progress," said cardiologist Paula Johnson, president of Wellesley College in Massachusetts and co-chair of the committee that wrote the report. "The legal system alone is really just not adequate for addressing the issues." The authors suggest universities take measures to publicly report the number of harassment complaints they receive and investigations

they conduct, use committee-based advising to prevent students from being in the power of a single harasser, and institute alternative, less formal ways for targets to report complaints if they don't wish to start an official investigation.

The report, *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine*, noted that many surveys fail to accurately evaluate sexual harassment. It used data from large, rigorous surveys done at the University of Texas system and the Pennsylvania State University system to describe various kinds of sexual harassment directed at students by faculty and staff.

The most common was "sexist hostility," such as demeaning jokes or comments that women are not smart enough to succeed in scientific careers, which were reported by 25% of female engineering students and 50% of female medical students in the Texas system. The incidence of female students experiencing unwanted sexual attention or sexual coercion was lower at



both systems—between 2% and 5% for unwanted attention and 1% for coercion. But the report declares that a hostile environment—even if it consists "more of putdowns than come-ons," as Johnson puts it—makes unwanted sexual attention and coercion more likely.

The report says women in science, engineering, or medicine who are harassed may abandon leadership opportunities to dodge perpetrators, leave their institutions, or leave science altogether. It also highlights the ineffectiveness of ubiquitous, online sexual harassment training and notes what is likely massive underreporting of sexual harassment by women who justifiably fear retaliation. To retain the talents of women in science, the authors write, will require true cultural change rather than "symbolic compliance" with civil rights laws. ■





## SEISMOLOGY

# Seafloor fiber optic cables can listen for earthquakes

Method offers a low-cost way to fill seismic detection gaps

By **Eric Hand**

**S**ome 70% of Earth's surface is covered by water, and yet nearly all earthquake detectors are on land. Aside from some expensive battery-powered sensors dropped to the sea floor and later retrieved, and a few arrays of near-shore detectors connected to land, seismologists have no way of monitoring the quakes that ripple through the sea floor and sometimes create tsunamis.

Now, a technique described online in *Science* this week promises to take advantage of more than 1 million kilometers of fiber optic cables that criss-cross the ocean floors and carry the world's internet and telecom traffic. By looking for tiny changes in an optical signal running along the cable, scientists can detect and potentially locate earthquakes.

The technique requires little more than lasers at each end of the cable and access to a small portion of the cable's bandwidth. Crucially, it requires no modification to the cable itself and does not interfere with its everyday use. The method "could be a game-changer," says Anne Sheehan, a seismologist at the University of Colorado in Boulder who wasn't involved in the work. "More observations from oceanic regions could fill in some pretty big gaps."

It began with an accidental discovery, says Giuseppe Marra, a metrologist at the National Physical Laboratory in Tedding-

ton, U.K., who works on the fiber optic links that connect atomic clocks in labs across Europe. He was testing a connection on a 79-kilometer buried cable that runs from Teddington to Reading, U.K., and relies on a stable, resonating loop of laser light. Vibrations near the cable—even the noise of traffic above—can bend it imperceptibly. That can shorten or lengthen the light's travel distance by less than the width of a human hair, shifting the resonating light beams slightly out of phase.

Marra was accustomed to background noise on his fiber optics. But when he reviewed data from October 2016, he saw more than the average amount of noise. It turned out to be the local effects of 5.9- and 6.5-magnitude quakes that struck central Italy late that month. "It was quite a revealing moment," Marra says. That noise, he realized, pointed to a new way to detect earthquakes.

Marra wondered whether the technique could be extended to the ocean, where the environment might be quieter. Using a 96-kilometer submarine cable connecting Malta and Sicily in Italy, he and his colleagues detected a magnitude-3.4 earthquake in the Mediterranean Sea. They couldn't pinpoint it. But by shooting lasers down a cable from both ends, he says, scientists could detect differences in the travel times of the out-of-phase signals, which would reveal just where the earthquake first caused a disruption along the cable. With three or more cables outfitted this

Seafloor cables, such as this link between the United States and Spain, can serve as seismic sensors.

way, he says, the earthquake's exact location in the crust could be triangulated.

By filling in the "seismic desert" in the ocean crust and showing where seafloor earthquakes occur and how often, the method could illuminate new fault structures and regions where tectonic plates are colliding or rifting apart, says Charlotte Rowe, a seismologist at Los Alamos National Laboratory in New Mexico. It could also help with tsunami warning systems, she says, provided the strength of the optical signal reveals an earthquake's size.

Besides mapping earthquakes, Rowe thinks the cable networks could sharpen pictures of Earth's interior. Like x-rays in a computerized tomography (CT) scan, seismic waves from big earthquakes carry clues to the density of rock they pass through. From criss-crossing waves received by multiple sensors, seismologists can construct 3D pictures of mantle convection, in which hot plumes well up and cold tectonic plates plunge toward Earth's core. Data from seafloor cables could fill in blind spots in these seismic CT scans. But Rowe says investigators will have to get better at interpreting the cable signals before using them to peer into deep Earth.

Marra says the new technique is sensitive enough to work across ocean basins thousands of kilometers wide. It requires adding a small cabinet of lasers and optical equipment that costs about \$50,000 at each end of the cable, and access to just one of the hundreds of channels in a typical cable. Renting a dedicated channel might cost about \$100,000 a year on a trans-pacific cable, and much less on one between North America and Europe, says Stephen Lentz, who works with the cable industry as director of network development for Ocean Specialists, Inc., based in Stuart, Florida. "Frankly, this is the kind of thing where the cable owner could donate the service and take the tax write-off. It costs them little or nothing to share unused wavelengths."

That's significant, says Bruce Howe, a physical oceanographer at the University of Hawaii in Honolulu, who leads a task force exploring how to stud new ocean cables with seismic, pressure, and temperature sensors, every 50 to 100 kilometers. Although the add-on sensors, at roughly \$200,000 apiece, are cheaper than operating stand-alone ocean bottom detectors, cable owners have been wary of affecting cable performance. The new technique offers a cheaper and less disruptive way to listen to the ocean floor. Howe calls the results "intriguing" and says his task force will advocate for a longer test. "It should absolutely be pursued." ■

## HIGHER EDUCATION

# Stricter Chinese student visas raise alarm

## Entry permission cut to 1 year for graduate students in some STEM fields

By Jeffrey Mervis

Reversing yet another policy of the previous administration, the U.S. Department of State this week began applying tougher restrictions on some Chinese graduate students. The new policy shortens from 5 years to 1 year the duration of visas for those planning to study aviation, robotics, and advanced manufacturing. The change is in line with a national security directive issued in December 2017, aimed at reducing “economic theft,” but university officials see it as an assault on graduate education and the free flow of scientific knowledge.

The new rule, first reported by The Associated Press on 29 May, was confirmed at a Senate hearing last week by Edward Ramotowski of the State Department, who oversees visa services for the Bureau of Consular Affairs in Washington, D.C. The policy would still allow students who receive a visa to remain in the United States for the duration of their program, but it would require those who travel abroad after 1 year to reapply.

Higher education officials say that change will make it harder for the affected Chinese students to attend international conferences, visit family back home, and work collaboratively with scientists abroad. When added to other policies by the current administration that affect non-U.S. citizens, the tighter visa policy may give talented foreign students one more reason to pursue advanced degrees in countries with lower barriers to entry.

“For decades, doing their graduate work in the U.S. was a no-brainer” for the best Chinese students, says Wojtek Chodsko-Zajko, graduate dean at the University of Illinois in Champaign. “But now, they have to decide if they really want to come here.”

The change in visa duration has caused broad anxiety among Chinese researchers. Zhang Tao, a control engineer at Tsinghua University in Beijing who works in robotics and aviation, is now worried that one of his master’s level students who would like to pursue additional training in the United

States will be shut out of any visa and that others may alter their destinations. “They are considering Japan and other countries where it might be easier to get a visa and realize their future plans,” Zhang says.

Reconciling U.S. national security with global science has always been a challenge, one that sharpened after the 9/11 terrorist attacks prompted increased vetting of all visitors to the United States. In 2004, U.S. universities and scientific organizations began expressing concern; a statement from 25 such groups warned about the “unintended consequences ... of new procedures and policies [that] have made the visa issuance process inefficient, lengthy, and opaque.” They kept up the pressure, and under then-President Barack Obama, the State Department gave consular offi-

Higher education organizations say there’s no evidence that tighter scrutiny would make the country any safer and worry that it might be applied selectively. And they say that additional roadblocks could pose difficulties for U.S. graduate programs in science, technology, engineering, and mathematics fields, which have become increasingly dependent on a steady flow of international students. After a decade of steady growth, roughly one-third of all foreign U.S. graduate students come from China.

Ramotowski did not mention the new rules in his opening statement at the Senate hearing. But he confirmed their existence when the panel’s top Democrat, Illinois Senator Richard Durbin, asked him whether it was true that, as of 11 June, “Chinese graduate students would be limited to 1-year visas if they’re studying in certain fields such as robotics, aviation, and high-tech manufacturing.”

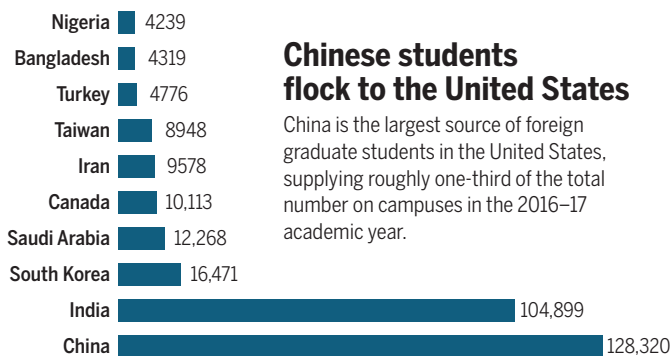
“Senator, we have issued some additional screening instructions to U.S. embassies and consulates to deal with certain individuals from China studying in certain sensitive fields,” Ramotowski replied.

He declined to answer questions from *Science* about how the new policy would be imple-

mented. In his testimony, he said that the rules “don’t in and of themselves prohibit the entry of anyone into the United States or restrict access to our country.” But university officials anticipate negative consequences.

“We might need to tell those students, if you come here, you’d better be prepared for the possibility that you won’t be going home, and you won’t be attending international conferences to present data and network,” says Chodsko-Zajko, whose university this year hosted 2600 Chinese graduate and professional students. “It’s not that you have to finish your program in a year. But ... it puts them at a competitive disadvantage against students from the U.S., or the European Union, or even elsewhere in Asia.” ■

*With reporting by Dennis Normile in Shanghai, China.*



### Chinese students flock to the United States

China is the largest source of foreign graduate students in the United States, supplying roughly one-third of the total number on campuses in the 2016–17 academic year.

cers the authority to grant 5-year visas to Chinese students, up from the 1-year limit that had long existed for that country.

But President Donald Trump has moved in the other direction since taking office, with calls for “extreme vetting” of visa applications for immigrants, students, temporary workers, and even tourists. The new student visa restrictions address China’s aggressive efforts to acquire foreign technologies, which have triggered a fierce political debate over whether universities are doing enough to help safeguard U.S. military secrets and intellectual property.

Senator John Cornyn (R-TX), who chaired a 6 June hearing on student visas, told *Science* that many university leaders “don’t prioritize national security concerns.” Scientists, he added, typically are “focused on attracting good students and just doing their research.”



## FEATURES

# FAR FROM OVER

By **Jon Cohen**; Graphics and data analysis by **Jia You**

**AT FIRST GLANCE**, Nigeria, Russia, and Florida have little in common. But each has had difficulty mounting an effective response to HIV/AIDS at a time when neighboring countries or states, buoyed by recent research advances, have made progress toward bringing their epidemics to an end.

Five of the main metrics that public health experts track to gauge progress against HIV are: How many people are living with the virus? What is the rate of new infection? What percentage of infected people are receiving anti-retroviral drugs, which both stave off disease and prevent transmission? How many infected people have progressed to AIDS and how many have died from it? And how many children are infected by their mothers?

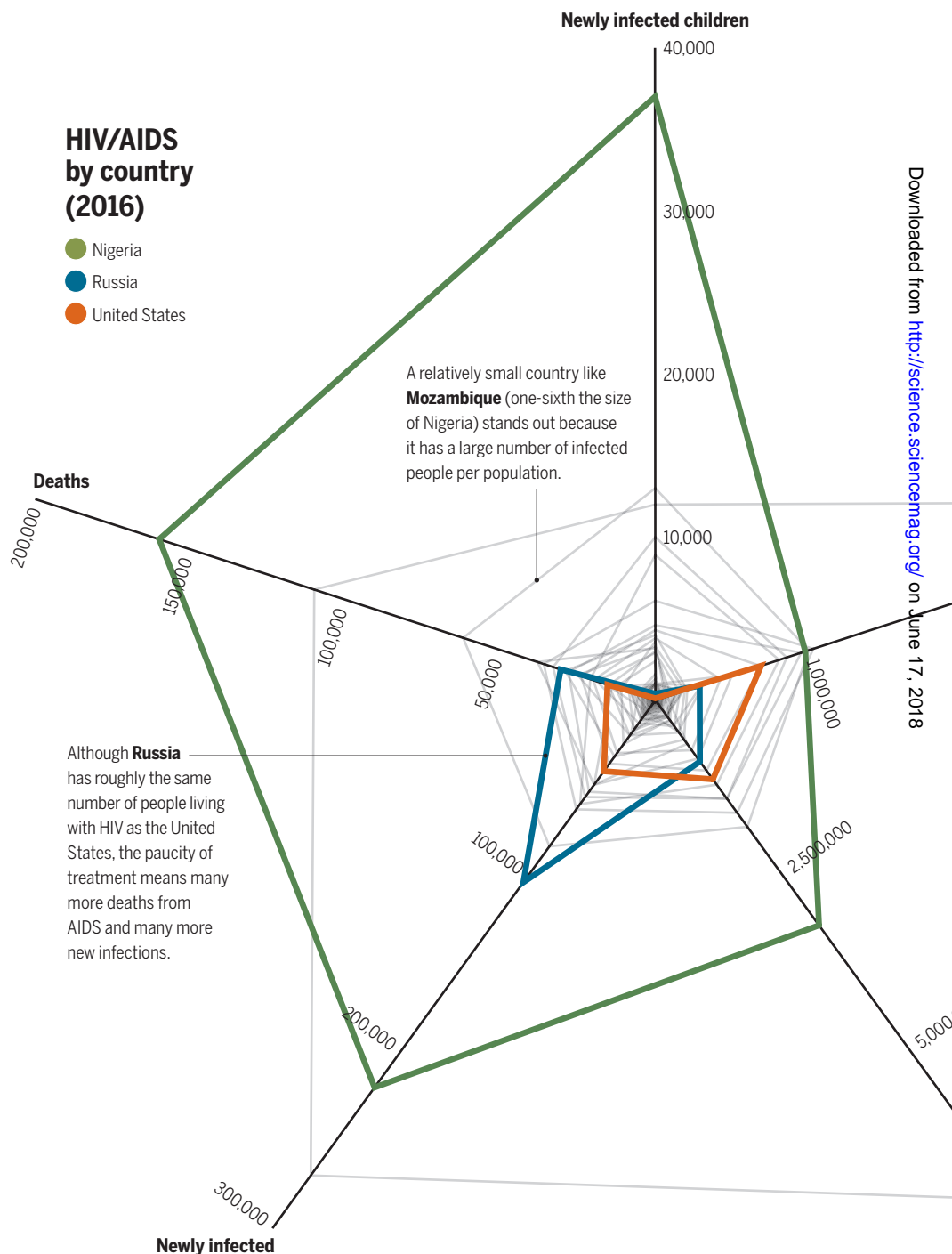
Much of the world has seen encouraging declines on many of those fronts. But Nigeria, Russia, and Florida stand out from their neighbors and, in some cases, the entire world. None of these three locales has high numbers on every one of these measures. But each ranks first—an unenviable distinction—in at least one of the five metrics assessed by total cases, rates, or proportions.

The aim of this package is not to shame Russia, Nigeria, or Florida, or, given their profound differences in population, politics, and economies, to compare them head-to-head. Rather, these stories describe the distinct challenges that have hampered each locale's response to HIV/AIDS. And they highlight people who are confronting those shortcomings and coming up with tailor-made, local solutions.

Science produced these stories in collaboration with the PBS NewsHour, which is airing a five-part series starting 11 June. Reporting for this project was supported by the Pulitzer Center.

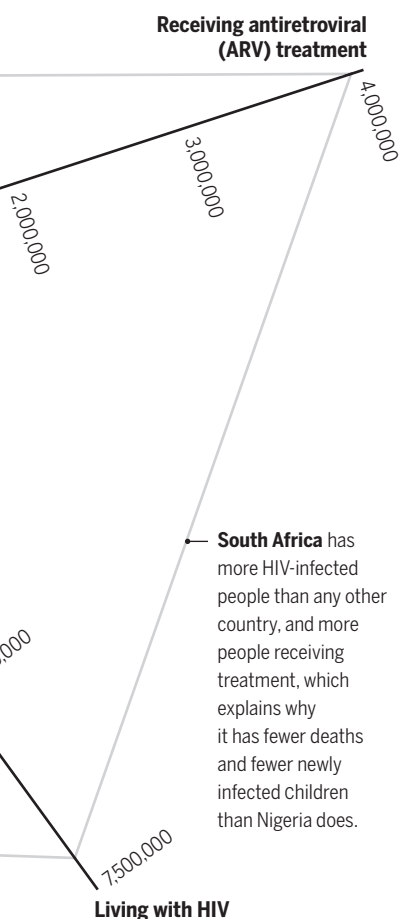
## HIV/AIDS by country (2016)

- Nigeria
- Russia
- United States



## Comparing epidemics

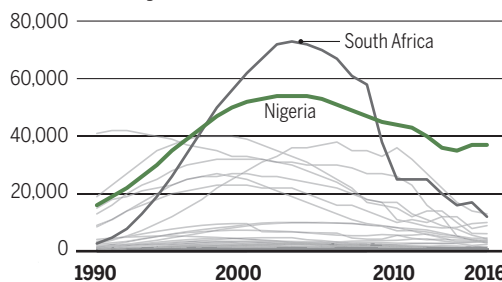
It's difficult to compare HIV/AIDS epidemics from one place to another because of differences in population size. These radar graphics capture this complexity by charting five different measures for each country or state. The graphic on the left compares Nigeria, Russia, and the United States with the rest of the world in raw numbers. The graphs on the right compare these epidemics with those in neighboring countries or states by showing rates and proportions, in addition to total numbers.



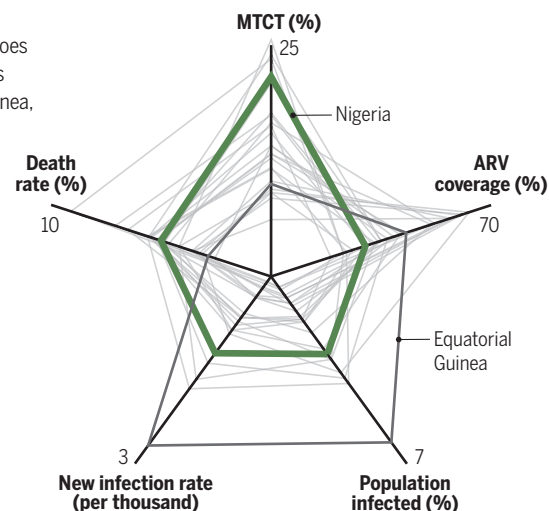
## Nigeria: children at risk

Since 2008, Nigeria has had more cases of mother-to-child transmission (MTCT) than any other country (below). Nigeria does not have the most severe epidemic in its region (right), but it has 3.2 million HIV-infected people—a huge challenge. Equatorial Guinea, in contrast, has the highest new infection rate and the highest prevalence of HIV, but it has a total population of only 1.2 million.

### Children newly infected with HIV, all countries



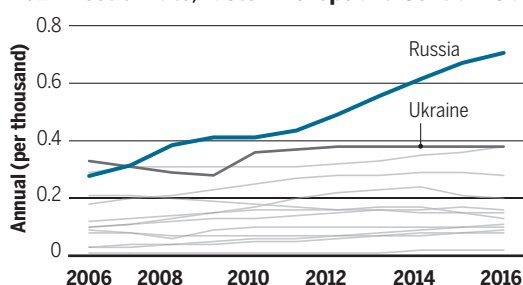
## HIV in Western and Central Africa



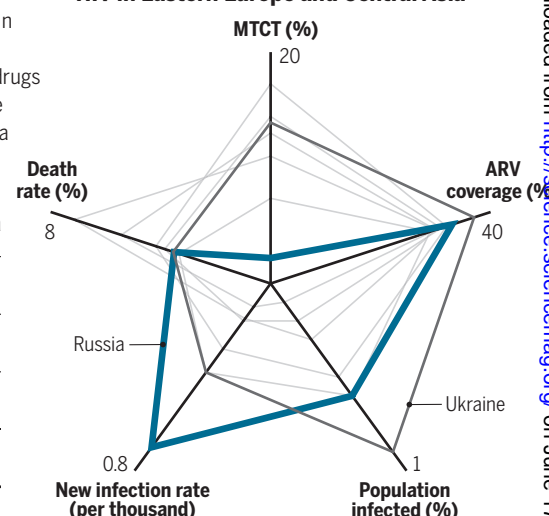
## Russia: a growing problem

Russia's rate of new infection outstrips every other country in Eastern Europe and Central Asia (below)—even Ukraine, which has more infected people per capita. Limited access to ARV drugs contributes to the country's high rate of new infections, because untreated people are more likely to transmit the virus. But Russia has succeeded in sharply reducing MTCT.

### New infection rate, Eastern Europe and Central Asia



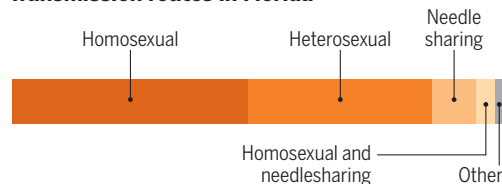
## HIV in Eastern Europe and Central Asia



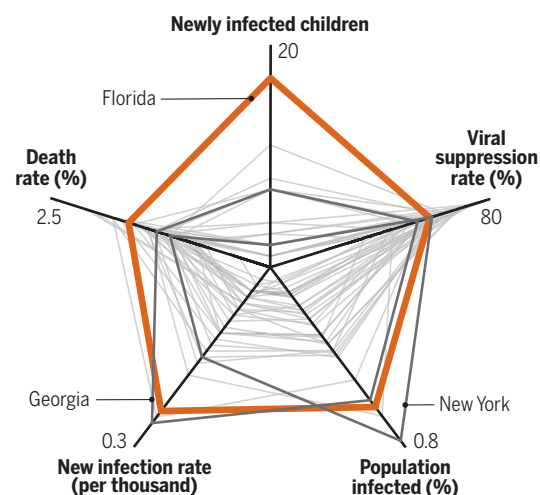
## Florida: high across the board

Compared with other U.S. states, Florida has a big problem. Georgia has the highest rates of new infection, but half the population of Florida. New York has a larger infected population, but has a lower death rate and fewer new infections. The United States only had 122 newly diagnosed children in 2016—some detected late and possibly born elsewhere—but Florida again stands out.

### Transmission routes in Florida



## HIV in U.S. states



These graphs rely on the best available data, which sometimes date back as far as 2014. Death, ARV coverage, and viral suppression are per HIV-infected population. The population infected is per total population. MTCT is per infected, pregnant women. Sources include the Joint United Nations Programme on HIV/AIDS, U.S. Centers for Disease Control and Prevention, and Russia's Federal Center for the Prevention and Control of the Spread of AIDS.



# THE MOTHER OF ALL CHALLENGES

Nigeria struggles to  
slow the spread of HIV to babies

By **Jon Cohen**, in Nigeria; Photography by **Misha Friedman**

**O**n a January morning, 12-year-old Yusuf Adamu slumps in his father's lap, head pressed against his chest. Infected at birth with HIV, he is tiny for his age and has birdlike limbs. He has been feverish for 3 days, which is why his father, Ibrahim, brought him to the pediatric HIV/AIDS clinic at Asokoro District Hospital in Abuja, Nigeria's capital. "He's been losing weight, he is not eating well, he's still taking his drugs, and he's complaining of chest pains and coughing," Ibrahim tells the nurse.

Yusuf's records show that at his last blood check 6 months ago, HIV had already ravaged the boy's immune system, even though he was receiving antiretroviral (ARV) drugs. When the doctor, Oma Amadi, examines his mouth, it is filled with white sores from candidiasis, a fungal infection. "The boy has been sick for so long," she says. "I'm going to admit him." When Amadi removes Yusuf's shirt to listen to his chest, the boy winces at the touch of her stethoscope. Amadi suspects Yusuf has tuberculosis, and after x-raying his lungs, the doctors put him in an isolation room.

Yusuf's mother was never tested for HIV before he was born: She received no prenatal care and delivered at home. Yusuf was not tested for the virus until she died of AIDS 3 years later. Ibrahim then learned that he, too, is HIV-positive, as are his two other wives. One ended up transmitting the virus to a second child, now 4.

The entire family receives ARVs, but Yusuf has only had intermittent access to the drugs. Dosing is based on weight, and Yusuf's has fluctuated so much that he has required monthly hospital visits. Ibrahim, a security guard, earns the equivalent of only about \$20 a month. The Adamus live 20 kilometers and three bus rides from the hospital. The round trip bus fare costs \$2, and Ibrahim has to miss a day of work for each checkup, when he also picks up his son's ARVs. Ibrahim simply can't afford regular treatment for his son. "There is no food at home," Ibrahim says.

Yet poverty alone does not explain the root of Yusuf's plight—which hundreds of thousands of other Nigerian children living with HIV now face. At a time when rates of mother-to-child transmission of HIV have plummeted, even in far poorer countries, Nigeria accounted for 37,000 of the world's 160,000 new cases of babies born with HIV in 2016. The most populous country in Africa, Nigeria does have an exceptionally large HIV-infected population of 3.2 million people. But South Africa—the hardest-hit country in the world, with 7.1 million people living with the virus—had only 12,000 newly infected children in 2016. The high infection rate, along with the lack of access to ARVs—coverage is just 30%—helps explain why 24,000 children here died of AIDS in 2016, nearly three times as many as in South Africa.

Downloaded from <http://science.sciencemag.org/>





Rose Kough and her 6-week-old baby received antiretroviral drugs to prevent mother-to-child transmission of HIV.



Mother-to-child transmission is only one part of Nigeria's HIV epidemic. But that route of transmission epitomizes the country's faltering response to the crisis, highlighting major gaps in HIV testing that allow infections to go untreated and the virus to spread. "Nigeria contributes the largest burden of babies born with HIV in the world—it's close to one in every four babies [globally] being born with HIV—and that's really not acceptable," says Sani Aliyu, who heads the National Agency for the Control of AIDS (NACA) in Abuja. And it is a solvable problem—even here. The key is to find and treat the relatively small population of pregnant, HIV-infected women, because those who receive ARVs rarely transmit the virus to their babies. Like most countries, Nigeria has made mother-to-child transmission a priority for more than a decade, and it has seen a reduction in children born with HIV. Still, the country stands out for its slow progress. "What we've realized is that we need to think outside the box," Aliyu says.

**A PREGNANT WOMAN** living with HIV has a 15% to 30% chance of transmitting the virus to her baby in utero or at birth, and breastfeeding will infect up to 15% more. In 1994, a study showed that one ARV drug, zidovudine, cut transmission rates by two-thirds if given to the mother before and after delivery and to the baby for 6 weeks. But few poor countries used that regimen because it was expensive and complex, requiring an intravenous drip of the drug during labor. Five

years later, a study in Uganda showed a single dose of another ARV, nevirapine, given to a mother in labor and a baby at birth, could reduce transmission by 50%, which soon became a standard of care. Countries all over the world began aggressive prevention campaigns. Nigeria launched a program in 2002 when it had 54,000 newly infected children, and transmissions began to slowly decline.

### ***"Nigeria contributes the largest burden of babies born with HIV in the world ... and that's really not acceptable."***

**Sani Aliyu**, National Agency for the Control of AIDS

Today, the standard of care is to treat all HIV-infected people, including pregnant women, with daily combinations of powerful ARVs. When treatment suppresses the virus in pregnant women and, as an additional safety measure, their newborn babies also receive ARVs for 6 weeks, transmission rates typically plummet to less than 1%. In the developed world and many developing countries, mother-to-child transmission is now rare. But the regimen can't be given if pregnant women don't know whether they are infected.

According to estimates from the Joint United Nations Programme on HIV/AIDS,

21.58% of HIV-infected, pregnant Nigerian women transmitted the virus to their children in 2016. Nigeria's central problem is that some 40% of women give birth at home or in makeshift clinics run by traditional birth attendants, where women are unlikely to get tested. The reasons women do not seek care at more formal health care facilities like Asokoro Hospital are many and overlapping: poverty, fear of stigma and discrimination for simply seeking an HIV test, lack of education, tradition, and husbands wary of health care.

Another barrier is the "formal" fee that the government levies for care at a clinic. Deborah Birx, director of the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) in Washington, D.C., which has invested more than \$5 billion in preventing and treating HIV in Nigeria, says the fee "opens the door" for others to tack on more insidious "informal" fees. "If you want to get your lab results back or you want to get your blood drawn, that nurse may charge you," Birx explains. Those fees, she says, "are very hard to police." When one Nigerian state eliminated the formal fee, the number of women who came to clinics for antenatal care doubled, she says.

Muktar Aliyu, an HIV/AIDS researcher at Vanderbilt University in Nashville who is Sani Aliyu's identical twin, says corruption is a major factor. "It's a big elephant in the room," says Muktar Aliyu, who still conducts studies in his home country. Scams such as informal fees are just part of the problem. The Global Fund to Fight AIDS, Tuberculosis and Malaria in 2016 suspended payment to the country after detecting what it called "systematic embezzlement" by Ministry of Health staff as well as improper auditing.

Conducting large-scale HIV testing is also hard because the virus is dispersed unevenly across the country, with some states having a much lower prevalence than others. In Niger, a state in the central part of the country, it is just 1.7%, according to 2015 estimates. "We'd test 1000, 2000 individuals and we'd get barely 20, 30 positive," Muktar Aliyu says. But Benue, an east-central state that has been hardest hit, has an estimated adult prevalence of 15.4%.

Several people at the front of Nigeria's HIV/AIDS response link the shortcomings to the government's lack of "ownership" of the epidemic. Foreign assistance—mainly from PEPFAR and The Global Fund—pays for nearly the entire HIV/AIDS response. Health Minister Isaac Adewole, an OB-GYN who worked in HIV/AIDS, says the "No. 1 challenge" is for Nigeria to move "from a donor-dependent program to a country-owned program." To give an example of the problem, Muktar Aliyu notes that foreign assistance often focuses on bolstering programs, including testing, at large treat-



Ibrahim Adamu sits with his son Yusuf in an isolation room at Asokoro District Hospital in Abuja.

ment centers, not the 800 or so smaller clinics spread across the country. "In the next 5 years, at the most, country ownership will come through for HIV programs in Nigeria," Sani Aliyu promises. "It's my job to make sure that money is available."

Since taking over NACA in 2016, Sani Aliyu has made some progress. For the first time, the federal government has been taking steps to prevent mother-to-child transmission, and state governments have devoted up to 1% of their budgets to efforts against HIV/AIDS. President Muhammadu Buhari, who appointed Sani Aliyu, authorized federal funds to pay for 60,000 new HIV-infected people to receive ARVs and vowed to add that same number to the treatment rolls each year. "The program, if successful, will serve as the exit gateway for PEPFAR as future programs acquire national ownership status," Sani Aliyu says.

Perhaps most important, NACA—with \$120 million in funding from PEPFAR and The Global Fund—now is working on a massive epidemiologic survey that many hope will bolster the country's efforts. Because HIV testing is so spotty, Bix explains, it's possible that official estimates of Nigeria's number of newly infected children are too high—or too low—and that HIV/AIDS workers target the wrong regions. "Our epidemic data from Nigeria is the weakest of all the countries," Bix says. The nationwide survey of HIV now underway, the largest ever done in the world, should be completed by the end of the year. "I'll be frank: I used to be upset with Nigeria," she says. "Now, I'm just waiting for the data."

Still, no one doubts that children are getting infected far too often. And some innovators are taking action.

**ON A SUNDAY MORNING** in mid-January, about 1000 parishioners fill the pews at the St. Vincent de Paul Catholic Church in Aliade, a remote, agricultural area in Benue state. Today, Reverend Emmanuel Dagi is leading celebrations called Baby Shower and Baby Reception, programs tailor-made to steer around the obstacles that keep so many pregnant women here from seeking an HIV test and receiving care.

Near the end of the church service, Dagi asks women who are pregnant or who have recently given birth to come forward for a blessing with their husbands. More than 50 people cluster around the pulpit, some women with large pregnant bellies, others with swaddled babies in their arms. "Defend these mothers and these fathers and their children from every evil," Dagi says. He walks from one end of the pulpit to the other, sprinkling the faithful with holy water.

The priest then asks the expecting cou-



Davida Anderson poses for friends behind a clinic that doubles as a hangout.

## Building TRUST in an LGBTQ-hostile country

**F**ive years ago, David Anderson, then 18, knew that he was at high risk of HIV infection. But he had never had a test for the virus because he feared that people at any clinic he visited would make assumptions about his sexuality and status. "I'm a bit feminine," says Anderson, who now openly identifies as both a gay man and a transgender woman (Davida). But then a friend told him about an unusual community center in Abuja, where he lives. "It was nice and safe for people of my kind to come," he says. Anderson tested negative and remains uninfected.

The clinic, opened by the Institute of Human Virology, Nigeria, in 2012, was the first in the country to cater to men who have sex with men (MSM) and transgender people, communities so heavily stigmatized that treating them is technically illegal. The clinic nevertheless provides state-of-the-art HIV services and conducts some of the most authoritative research on those communities in Nigeria. "We offer one-stop shopping," says Man Charurat, an epidemiologist heading a multifaceted research project at the clinic with virologist William Blattner, both of whom are based at the institute's main facility in Baltimore, Maryland.

The study is known as Trusted Community Center to Reduce HIV Infections by Engaging Networks of Friends and Partners to Support Safe Sex, or TRUST—the nickname of the clinic. The study looks at, among other things, the prevalence of HIV, the rate of new infections, and how the virus spreads in networks of sexual partners. Each person who enrolls is given a small incentive to bring five friends. TRUST—which has a second clinic in Lagos run by the U.S. Military HIV Research Program—now has more than 2000 participants, a startling 45% of whom tested positive for HIV.

TRUST also is looking at how antigay laws and other stigmatizing social factors impede care and treatment. In 2014, a new Nigerian law called the Same-Sex Marriage (Prohibition) Act made providing services to MSM a punishable offense. (To avoid legal complications, the TRUST clinic is located out of the way, in a residential neighborhood, and bills itself simply as "MSM friendly.") Stigma and discrimination against that community were already pervasive the year the law was enacted: According to a survey of 3500 undergraduates in Lagos, nearly 40% thought health care workers should not provide services to MSM, as reported in the August 2017 issue of *LGBT Health*. Researchers and affected communities alike contend that the law has made things worse.

Charurat, Blattner, and colleagues assessed their clients' attitudes toward seeking health care before and after enactment of the law. Fear of seeking health care jumped from 25% to 38%. Today, only 25% of TRUST participants infected with HIV know their status when they first come in. Of those, fewer than half are on treatment. TRUST encourages all HIV-infected clients to start antiretrovirals, and a new program is offering the drugs to uninfected participants—a proven prevention strategy called pre-exposure prophylaxis.

The TRUST clinic has evolved into a place to hang out as well. Anderson sometimes lets Davida shine, donning a wig, high heels, and an elegant gown. "That's the reason we set this up," says Charurat as Davida struts around the TRUST clinic grounds, "voguing" with friends. "It's not just a clinic. It's a community." —Jon Cohen



## Babies who dodge HIV may not be unscathed

Up to half of babies born to HIV-infected mothers become infected themselves either in utero, during labor, or through breastfeeding, unless the mothers and babies are treated with antiretroviral (ARV) drugs. But several studies suggest an additional burden for the many children in Nigeria who dodge transmission from their mothers, whether by luck or treatment: They may still suffer from growth retardation, immune abnormalities, and even an increased risk of mortality. "Exposed, uninfected children are a growing population, and they're neglected," says virologist Nicaise Ndembu of the Institute of Human Virology, Nigeria (IHVN) in Abuja.

At the University of Benin Teaching Hospital in Benin City, Ndembu is a co-principal investigator (co-PI) of a study exploring what happens to those exposed, uninfected children and why. "There's a growing body of data that suggests they're very different from normal infants," says another co-PI, epidemiologist Man Charurat of IHVN.

Theoretically, the harm could come from the mother's poor health, or from exposure to viral proteins or to the ARV drugs given to treat the mother and the baby to prevent transmission. "It's a really complicated picture," says Claire Thorne, an epidemiologist at University College London who is trying to separate the effects of HIV exposure from exposure to ARV drugs.

The IHVN team is exploring whether an HIV-caused condition called "leaky gut" may transfer from mother to uninfected child. HIV preferentially destroys immune cells in the gut shortly after infection, leading to an overstimulated immune system, inflammation, and a permeable gut. That could alter the mother's microbiome, and she might pass the altered bacteria to her child during birth or breastfeeding. The Benin researchers suspect the child's altered microbiome then increases diarrhea, impedes proper absorption of nutrients, and slows growth.

The IHVN study, which began in 2014 and will run through next year, compares 300 uninfected children, half of whom were exposed to HIV during their first 2 years of life. The researchers are analyzing stool samples from mothers and their children to see whether HIV alters the genera of bacteria that live in their guts. Researchers also are giving the children a dose of two sugars and then testing their urine to see whether their intestines have compromised permeability. "If we can identify problems, we can begin to look at interventions," Ndembu says. —Jon Cohen



Researchers measure head circumference to compare growth in children born to HIV-positive and negative mothers.

ples to attend Baby Shower, where they receive a gift bag and have blood drawn for tests for hepatitis B, sickle cell anemia, and HIV—casually lumped into the mix to sidestep stigma. People with newborns attend a separate celebration, Baby Reception, where they, too, receive gift bags. At the same time, health workers discreetly check with all people who tested HIV positive at earlier services to make sure they've been following proper procedures: taking ARVs for themselves, administering them to their newborns, and bringing the babies in for a blood test at 6 weeks of age, the earliest the virus can reliably and efficiently be detected.

Those celebrations are part of the Healthy Beginning Initiative, funded by the U.S. National Institutes of Health and the Centers for Disease Control and Prevention. It has now expanded to more than 115 churches, some of which also test for malaria, syphilis, and anemia. Led by Echezona Ezeanolue, a pediatrician who works with the Healthy-Sunrise Foundation based in Las Vegas, Nevada, the intervention takes advantage of the fact that some 90% of Nigerians regularly attend either church or mosque services. "That was a perfect place to test this," says Ezeanolue, a Nigerian who left the country 20 years ago. (Ezeanolue resigned from the University of Nevada in Las Vegas in March after a prolonged dispute with the school about what it alleged were financial irregularities with the maternal HIV program he ran there. He insists he did nothing wrong.)

Sani Aliyu adds that Baby Shower has another advantage. "The religious leaders don't carry the burden of politicians," he says. "People believe in what they say, and people follow them."

In a study of the project in 40 churches, half of which received the intervention, Baby Shower increased HIV testing in pregnant women from 55% in the control churches to 92%, the researchers reported in the November 2015 issue of *The Lancet Global Health*. A study published in *AIDS and Behavior* last year shows that in women's male partners, testing jumped from 38% to 84%. "With Baby Shower, you don't have to go to the hospital for an HIV test—you go to the church and nobody suspects anything," says Amaka Ogidi, coordinator of the project here. A follow-up study is assessing the actual impact on HIV transmission rates to the babies.

Ogidi says she at first had reservations because the idea of a baby shower is a U.S. concept. "We're not used to celebrating pregnancy—we're used to celebrating childbirth," she says. But the intervention has steadily grown in popularity, especially since the Baby Reception component was added. "The program is just like a sweet-smelling perfume," Ogidi says. "You smell it and say,





In Aliade, Nigeria's St. Vincent de Paul Church, Reverend Emmanuel Dagi blesses pregnant parishioners and encourages them to attend the Baby Shower program.

'Oooh, can I have some for myself?' You see smiles on faces and it's infectious."

**THE MAMA METTA** Traditional Clinic and Maternity Home in the Iyana Ipaja neighborhood of Lagos also is introducing prevention of mother-to-child transmission into a familiar setting. The clinic, set on a street wide enough only for foot traffic, resembles a small house. Feyami Temilade, who runs the clinic, is a traditional birth attendant, and she is known as Mama Metta because she becomes something of a second mother to every woman in her care.

This Friday morning, 16 big-bellied women sit on wooden benches in the waiting room. The walls are cluttered with framed certificates from courses Temilade has completed over the past 35 years, awards, fading photos of herself in celebratory garb, calendars, and pregnancy infographics. A poster above Temilade's desk says, "Know Your HIV Status" in English, Yoruba, and Nigerian Pidgin. A certificate next to it notes that she participated in a training workshop for traditional birth attendants run by the Society for Women and AIDS in Africa in—astonishingly—1991.

The women wait their turn for check-ups and to fill bottles with two herb concoctions she has prepared. The cost of the visit, says Temilade, who will listen to babies through a metal fetoscope

called a Pinard horn, is a mere 200 nairas—about \$0.55.

Birth attendants are unregulated in much of Nigeria. But Lagos state, which includes Nigeria's largest city, has a Traditional Medicine Board that accredits and monitors practitioners. Since 2012, the board has offered regular HIV/AIDS training, which includes an internship at a hospital, for Temilade and some 2000 other birth attendants.

Temilade has been in business here so long that two of the women here today were born in the facility, which has a birthing

room with two beds. "If you're pregnant, you're eating for two people," Temilade tells the group. She also warns them to avoid high heels and not to sit in the same position for too long. "Come here for HIV tests," she says, explaining that she'll link anyone who tests positive to a hospital.

That's a start, Sani Aliyu says. But, ultimately, he believes, getting women to antenatal clinics in established health facilities will be key to stopping mother-to-child transmission here. "In most cases, at least 80% of them will get tested and commence treatment," he says. One state has experimented with paying incentives to traditional birth attendants to bring pregnant women to formal health care settings.

Pregnant women make up only a fraction of Nigeria's huge HIV-infected population, and Sani Aliyu well recognizes that halting the epidemic here, as in other countries, depends on treating nearly everyone who is living with the virus—not just mothers and babies. "It's going to be a lot of work to put everybody with HIV on treatment," he says. But protecting babies from infection, he says, "should be a low-hanging fruit that can be reached." ■



Mama Metta, a traditional birth attendant in Lagos, Nigeria, listens to a woman's belly with a Pinard horn.

*Reporting for this story was supported by the Pulitzer Center.*





# DARK NIGHTS, BRIGHT STARS

Russia's explosion of HIV reflects stigma and neglect, but it spotlights courageous actions by a few

By **Jon Cohen**; Photography by **Misha Friedman**

In 2015, a dermatologist in Russia's fourth largest city, Yekaterinburg, diagnosed Katia with herpes. "I had no idea what it was," says Katia, who asked that her last name not be used. But because she had suffered repeated illnesses over the preceding 2 years and had an alcoholic ex-boyfriend who simultaneously had other girlfriends, she sus-

pected that something more serious might be wrong. She asked the doctor to give her a referral for an HIV test. "Why?" he asked. "Are you going to marry a foreigner?"

"I persisted and said, 'I'm not leaving here until I get a referral.'"

The day Katia learned the test results, she walked the streets for hours crying, unable to even find her car. The literature she re-

ceived explained she had the AIDS virus and said the outlook was grim. She read that the virus, if untreated, could kill her in as few as 3 years. Katia, then 30, had a young daughter. Who would raise her? And Katia had to hide her infection from her employer—her father. "If I told my father I had HIV, he wouldn't understand. He'd run away from me. Overall, the attitude here is horrible. When I look



Volunteers from the Andrey Rylkov Foundation distribute free needles and condoms in Moscow.

at [Western] Europe or the United States, I don't understand why it's so different here."

For Katia, the differences between east and west would soon become far starker.

Since 2015, the World Health Organization (WHO) in Geneva, Switzerland, has recommended treating everyone who tests positive for HIV, but doctors—despite what the literature suggested—assured Katia her immune system hadn't suffered enough damage to warrant antiretrovirals (ARVs). Two years later, Katia was married to a new partner and wanted to have another child. She had read on social media that HIV was a scam by Big Pharma, but that made no sense to her because the government promises free treatment. Other sites said proper treatment would almost eliminate her risk of transmitting the virus to her husband, who was uninfected, or their baby. She went to the AIDS center and requested ARVs. "They told me, 'Don't worry, your level of virus is low, you won't infect your husband, he's in the military and he's strong.'" Finally, in February 2017, after more than 2 months and a battery of tests, the doctors agreed to treat her, and in May, she became pregnant.

In June, the pharmacy ran out of ARVs.

Through social media, Katia contacted an HIV-infected woman, 2000 kilometers away in St. Petersburg, who was part of a "reserve pharmacy"—a network that gathers and redistributes ARVs from people who have either switched treatments or died (see sidebar, p. 1174). The woman told Katia whom to contact in Yekaterinburg. "I asked how much the pills would cost," Katia says. "She was like, 'Are you out of your mind? Just go and get your pills.'"

Almost anywhere in the world, an HIV-infected woman who has an uninfected partner and wants to have a baby would be first in line to receive ARVs. The challenges Katia faced in getting treatment amid Russia's epidemic highlight the country's faltering response, which critics have blasted as misguided, lackadaisical, and downright dismissive. Some federal health officials even question the term epidemic. "This is a very large and very serious epidemic, and certainly one of the few epidemics in the world that continues to get worse rather than get better," says Vinay Saldanha, the Moscow-based regional director for the Joint United Nations Programme on HIV/AIDS (UNAIDS) in Eastern Europe and Central Asia. "This is a public health crisis."

Yet the informal network of HIV-infected people that ultimately supplied Katia with ARVs highlights another, less recognized side of Russia's response. The darker the night, as

Fyodor Dostoevsky wrote in *Crime and Punishment*, the brighter the stars. Bold, committed HIV/AIDS advocates in Russia are pushing hard for change—and a few places in the country show signs, albeit modest, of mounting effective responses.

**UNAIDS CALCULATES** that between 2010 and 2015, Russia accounted for more than 80% of the new HIV infections in the entire Eastern European and the Central Asian region. By Russia's own estimates, the epidemic grew

10% per year during that period, with the new infections roughly split between people who inject drugs and heterosexual transmission. At about the same time, new infections in the rest of Europe and North America dropped by 9%. By the end of 2017, the Russian Ministry of Health estimates, just shy of 1 million people were living with HIV. Even official accounts acknowledge that only one-third of those receive ARVs.

Many people, even within the government, think those estimates understate the



## Status symbol

In 2015, Pavel Lobkov, a news anchor on TV Rain, an independent station in Moscow, confronted a predicament unlike any he had faced in more than 2 decades working as a journalist. It was World AIDS Day, and he was slated to host a show. His guest was the head of the Federal Center for the Prevention and Control of the Spread of AIDS, Vadim Pokrovsky. For 12 years, Lobkov had been Pokrovsky's patient.

Lobkov had never spoken publicly about his HIV status. No well-known person in Russia ever had—the country had no Magic Johnson. He asked a close friend for advice, who said to either call in sick or reveal his status. Lobkov switched off his cellphone. "Otherwise, I can call a second friend, a third friend, and somebody would distract me from doing this," he says.

Lobkov opened the show, *Hard Day's Night*, by explaining that Pokrovsky was more than a guest. "You are the one I came to with my trouble in 2003," he said, recounting the heartless way another doctor told him his diagnosis. His on-air revelation made headlines, even in the United States. And the fact that he was then nearly 50 and didn't have spots on his cheek, look exhausted, or show any signs that he was infected sent an important message to Russians, he says: You can live a normal life with HIV. Pokrovsky's AIDS center saw a jump in people coming in for HIV tests, and Lobkov received a flood of questions on social media seeking his advice.

Why did it take until 2015 for a prominent Russian to finally reveal that he was living with HIV? "We are still Soviet," Lobkov says. "And we are still trying to keep our problems inside us." —Jon Cohen





## Poster couple

Last year, on the streets of St. Petersburg, a poster appeared, featuring an attractive young man and woman looking into each other's eyes. "People with HIV are just like you and me," it read. It was part of a provocative campaign to encourage HIV testing in a country where the infection is still highly stigmatized. In an extra twist, the woman, Tatiana Vinogradova, is the prominent deputy director of the city's AIDS center, and the man, Andrey Skvortsov, who is living with the virus, is her husband.

They decided to do the poster, says Skvortsov, who runs a nongovernmental organization called Patients in Control, "to do the maximum we could" for both the patient community and public health. "I try to motivate people, I'm an example," says Skvortsov, who was diagnosed in 1997 with an HIV-ravaged immune system and today, thanks to antiretrovirals (ARVs), is healthy.

Vinogradova had another goal as well: to reduce bias and fear within the health care community. "I have a lot of friends who are doctors because we studied together at university, and they are still afraid of HIV patients," says Vinogradova, whose mother once ran the city's AIDS center and whose grandmother diagnosed the city's first HIV case. "They're not afraid of TB [tuberculosis]. They're not afraid of hepatitis C. In the late '80s, AIDS was the plague of the 20th century and it's still in their heads."

That stigma hit home 3 years ago, she says, when Skvortsov broke his collarbone in a bicycle accident. As a surgeon described the operation needed to repair his fracture, the couple told him that Skvortsov was living with HIV. Vinogradova offered to bring his medical records to show that he was on ARVs, that the virus was undetectable in his blood, and that his immune system was normal. The surgeon left the room, and when he came back, he said Skvortsov needed only a sling. They went to another hospital, where a surgeon had no concern about Skvortsov's HIV status and scheduled the operation for the next day.

Skvortsov and Vinogradova feel the situation is improving, if slowly, because the increasing severity of the epidemic has rattled complacent officials, and a new generation of doctors and activists understands the disease and what needs to be done. But they believe the situation will have to become bleaker before the epidemic receives the urgent, full-force response that's needed. "We have to be facing something terrible, and then there will be change," Skvortsov says. "It's the Russian mentality." —Jon Cohen

epidemic. Vadim Pokrovsky, who heads the Federal Scientific and Methodological Center for Prevention and Control of AIDS in Moscow, an epidemiological surveillance arm that's independent from the Ministry of Health, does battle with the divisions that oversee HIV and tuberculosis (TB). His group estimates that between 1.1 million and 1.4 million Russians are infected with HIV (see sidebar, p. 1175). And Michel Kazatchkine, special adviser to UNAIDS in Eastern Europe and Central Asia, and colleagues concluded in a study published last year that the true number of infected people may be as high as 2 million. They based their number on an estimate by Pokrovsky's team that only half the infected people in Russia knew their HIV status in 2013.

Kazatchkine, who earlier headed The Global Fund to Fight AIDS, Tuberculosis and Malaria, also in Geneva, says Russia has unnecessarily let a containable epidemic explode. "Having gone through years of AIDS in Europe and then in Africa and everywhere as head of The Global Fund, I just cannot accept that there have been so many missed opportunities," he says.

HIV took off in Russia in the mid-1990s, later than in Western European countries, and at first the spread was limited mainly to heroin users sharing needles and syringes. Even before he took over The Global Fund in 2007, Kazatchkine and others urged the government to use proven "harm reduction" prevention strategies, including needle and syringe exchange programs and opiate substitutes such as methadone. Several nongovernmental organizations (NGOs) started programs for drug users, but the government offered no funding for what it saw as "Western" ideas that went against conservative Russian culture—and to this day opiate substitutes remain outlawed. The government also did little for another highly stigmatized group, men who have sex with men (MSM), who also are at high risk of infection. Since 2013, a "propaganda law" has made posting or discussing information geared toward MSM illegal. They rely on NGOs for many HIV services, as do sex workers.

Compounding the problem, several NGOs left the country starting a decade ago after Russia, emboldened by a strengthening economy, began phasing out The Global Fund's support—which to date has totaled \$378 million (*Science*, 9 July 2010, p. 168). "They said, 'We don't need The Global Fund or money from outside; we'll cover everything ourselves,' but they didn't do it," Pokrovsky says. "We not only stopped outreach work, we lost many



people working in NGOs because they had no federal support." Other NGOs pulled up stakes after President Vladimir Putin in 2013 enacted a law that forced many of them to register as "foreign agents."

"They basically let the epidemic grow because of lack of prevention and because of very low access to treatment," Kazatchkine says. "In short, they did it all wrong."

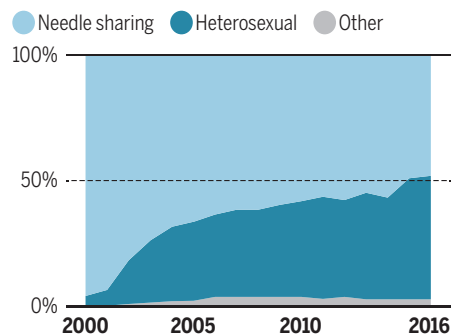
No program yet provides ARVs to high-risk uninfected people, a prevention strategy called pre-exposure prophylaxis (PrEP) that has had remarkable success, particularly with MSM in Western Europe, Australia, and the United States. "Who would PrEP be for?" Kazatchkine quips. "There are no MSM."

On top of all those challenges, Russia has a dysfunctional medical system, says Olga Bogolyubova, a psychologist who did HIV/AIDS research in St. Petersburg but, fed up, moved to the United States in 2015. Bogolyubova, who works at Clarkson University in Potsdam, New York, says Russia's broken system vastly complicates attempts to add HIV/AIDS programs for vulnerable populations. "The Russian medical system can be difficult to navigate and abusive even to people who aren't vulnerable," she says, noting long waits to see specialists and short supplies of medicines for many conditions. "A navy admiral recently shot himself because he couldn't get access to cancer care."

**TEREZA KASAEVA**, a deputy director of the Ministry of Health who coordinates the country's HIV/AIDS programs (after meet-

## A changing epidemic

The HIV/AIDS epidemic in Russia began in people who inject drugs, but heterosexual transmission has become a major driver. In small surveys, men who have sex with men have extremely high infection rates, but they rarely are studied.



ing with *Science* in November 2017, she became head of WHO's Global TB Programme), acknowledges that Russia "didn't pay much attention" to the problem until the past 5 years, but she says the current criticism is overblown. "Everybody is talking about the need to avoid stigma," Kasaeva says. "We are against the stigma on the Russian Federation about the situation."

Kasaeva and her colleagues stress that the number of people getting treatment has increased severalfold over the past few years, and they note that they recently completed their first strategic plan to battle the disease. "We understand that now we have a

problem and we're trying to solve it," she says. Everyone who requests treatment receives it, she insists, although some "people are trying to hide" and don't request it. Yes, shortages of ARVs occurred in some places last year, but that was only because of speed bumps in a shift away from regional drug purchase and distribution to a more streamlined, cost-effective federally run program.

Kasaeva recognizes that harm reduction programs might slow HIV's spread in the short term, but they don't address underlying problems, she says. "They're very popular and they look so smart and pretty, but in fact if we look into the problem deeply, it's not solving the problem." Harm reduction, "according to the opinion of a huge number of Russian experts," focuses on the symptoms and not the causes of addiction, she says. In contrast, the country's "systematic program of rehabilitation for drug users," she insists, has long-term effectiveness.

Yevgeny Roizman, who until last month was the powerful mayor of Yekaterinburg, has embraced that effort. (Roizman resigned in protest after regional lawmakers voted to end direct election of the mayor.) In 1999 he started a rehab program called City Without Drugs that, at the request of parents, forcibly removed teen addicts from their homes and locked them up while they went through withdrawal, chaining them to their beds at times. Criticism from human rights groups—featured prominently in Western European and U.S. media—and lawsuits ultimately shut the centers, but Roizman offers no apol-



Yevgeny Roizman, former mayor of Yekaterinburg, Russia, has opened a museum of religious icons. He has taken a hardline stance against drug users.



## The pill exchange

In 2012, when Alexander Chebin was dying from AIDS, he met a Pentecostal bishop who persuaded him to seek care and start taking antiretrovirals (ARVs). Chebin, who had served time in prison for a heroin-related offense and did not believe that HIV causes AIDS, realized "it was time to change not only my mind, but my faith." Now a pastor himself, Chebin is part of an informal network of activists in seven Russian cities who give other people with HIV hope through what they call a "reserve pharmacy."

People who can't access ARVs, mainly because government clinics have run out, post their problems on social media sites that Chebin and other activists monitor. Some people use private chats on Facebook, but many others post to a public website called pereboi.ru that is run by Patients in Control, "a movement of people living with HIV and other socially significant diseases." Chebin and other like-minded activists, or "consultants," in turn distribute ARVs to people in need in their areas.

At his home, Chebin keeps a large stock of ARVs donated by people who have switched drugs or by relatives of people who have died. He also swaps ARVs through the informal network with activists such as Yulia Vereschagina, who lives in St. Petersburg and keeps her own large stash of medicine in a cabinet under her TV. Vereschagina, a consultant with pereboi.ru, says she filled 600 requests between March and November 2017, sending 140 packages to Chebin and suppliers in other cities.

Vereschagina got involved with the network in March 2017, when she had side effects from her own ARVs and her clinic said it could not take back the pills that she wanted to return. She says people ask for help not only because of drug shortages, but also because the official process of registering for treatment can take months and is sometimes impossible. "Until the situation with the medication improves, I'm often the only source," she says. "There's always a shortage somewhere."

—Jon Cohen

ogies. He insists the program helped his region, the Sverdlovsk Oblast, eliminate heroin addiction and slow the spread of HIV. "My goal was to stop youth from using drugs," Roizman says. Although he offers few hard numbers to back his claims that the program worked, he has an abundance of anecdotes, proudly noting that he recently located 22 of the first teens treated at City Without Drugs who now have 38 kids of their own. "They're not doing drugs ever since they got to us," he says.

Roizman sees no need for harm reduction programs for drug users; NGOs that offered needle exchange in Yekaterinburg failed to make a dent, he says. He visited Ukraine, which legalized opiate substitution therapy, to see how it worked. "An army of heroin addicts," he says, was "reconfigured into an army of methadone addicts."

In other ways, however, Roizman's agenda is not entirely at odds with the international HIV/AIDS community, and he has a progressive bent. A published poet who opened a free museum of religious icon painting and sculpture, he strongly supported the opposition to Putin in the March election and is critical of "the main government officials" for not speaking about HIV. Roizman promotes condoms and sex education for teens. And he publicly took an HIV test himself to encourage people to know their status and start treatment if needed. "There are more serious problems and issues that we have to battle, but bearing in mind the dynamics,

we do understand that HIV might just steal the future."

**A DIFFERENT APPROACH** is in view on the banks of the Volga River in Kazan, the capital of the Republic of Tatarstan. Kazan's ornate kremlin (which means "fortress"), replete with both a mosque and a cathedral, dates back to Ivan the Terrible in the 16th century. The city has fanciful buildings of more recent vintage, too: a 30-meter-tall, copper-colored cauldron, a wedding palace built in 2013; and a swank soccer arena that opened the same year and is hosting World Cup games this month. The city is well-heeled, with an unusually large Muslim population for this part of Russia and a long history of independence—which helps explain why it stands out in its HIV/AIDS response.

Kazan in 1999 began aggressive needle and syringe exchange programs and other harm reduction efforts with support from NGOs, the semiautonomous republic itself, and eventually The Global Fund. "We've had a stable epidemic," says Larisa Badrieva, an epidemiologist and researcher there. "We really reached a lot of people very, very quickly." The city recorded about 1000 new HIV cases in 2001, she says, and that number fell to about 150 in 2008, with little evidence that the virus was spreading to the general population. A relatively high percentage of HIV-infected people in Tatarstan—about 50%—is being treated with ARVs. Yet Badrieva is worried about the future. With the pullout of The Global Fund and with other outside as-



Vyacheslav Ignatenko (right), holding a tray of opiate cooking equipment, receives bags of syringes from a needle exchange in Kazan, Russia.



sistance drying up, Badrieva says only one of seven centers in the city dedicated to drug users is still operating, and she is unsure whether it will remain open much longer. "Unless there's a dramatic shift in the drug scene, we'll see an increase of HIV infection across all groups," Badrieva predicts.

Like Kazan, St. Petersburg is often out of step with Moscow and has made solid headway against HIV/AIDS. The most westernized city in the country, St. Petersburg "is kind of an oasis," says Gregory Vergus, who works with an HIV/AIDS NGO called the International Treatment Preparedness Coalition. As a federal city, St. Petersburg directly receives funding from Moscow for HIV prevention, and Vergus says it invests the money wisely, focusing on vulnerable groups. "Most regions spend their HIV prevention money on balloons, songs, and working with grannies," he says.

Saldanha says he's particularly heartened about the latest data from the city. "We're hoping that St. Petersburg has now turned the corner," he says. In 2016, new infections in the city dropped below 2000 people a year for the first time since the epidemic took off, according to the city's AIDS center. Of the 36,000 residents who know they are living with HIV, about half now receive ARVs, and 82% of that group have undetectable levels of virus in their blood, meaning they are adhering to their treatment regimens. "We're doing something right," says Tatiana Vinogradova, deputy director of the St. Petersburg AIDS Center (see sidebar, p. 1172).

St. Petersburg still faces serious challenges, Vinogradova notes. Recent surveys that tested large groups of sex workers and MSM both showed double-digit prevalence of HIV, and only about 5% of the people already knew their status. Many migrants who have HIV also come to the region from former Soviet countries and do not have the proper St. Petersburg registration necessary to receive treatment and care. "It's a big problem, and we have no instruments to do anything about it," she says.

Still, Saldanha says, the fact that "St. Petersburg is finally pulling its head above water" gives him hope for the country as a whole. "It's very much a sign that in contemporary Russia, you can implement evidence-based HIV prevention programs," he says. But St. Petersburg is only one city in a highly populated, geographically expansive country. "They have a treatment coverage level that is half what it is in Zimbabwe," he stresses. "This epidemic is not going to go away on its own." ■

*Reporting for this story was supported by the Pulitzer Center.*



## The loyal opposition

**V**adim Pokrovsky, head of the Federal Center for the Prevention and Control of the Spread of AIDS in Moscow, is—to borrow from a Russian saying—deft enough to “shoe a flea”: He’s a top government official who speaks critically about its policies and yet keeps his job. “At the present moment, there are no real attempts to overcome the epidemic,” he says, noting that Russian President Vladimir Putin has not mentioned HIV/AIDS in a decade. “We need something new.”

Pokrovsky’s to-do list runs long. The government now spends 24 billion rubles (about \$400 million) on HIV/AIDS prevention and treatment, a sum he thinks should be quadrupled. He dismisses the government’s boast that it has expanded testing to some 32 million people each year. “Who was tested?” he asks. “Old women and children.” The conservative Russian Orthodox Church, he complains, has stifled sex education and condom campaigns. “I think it’s better first to save the life and then to save the soul,” he says.

Gregory Vergus, the St. Petersburg–based regional coordinator of a non-governmental organization, the International Treatment Preparedness Coalition in Eastern Europe and Central Asia, says Pokrovsky’s voice is essential because the entire system, from testing to delivering drugs, is badly broken. “Nobody in the government cares at every step of the way,” says Vergus, whose group advocates to make anti-HIV drugs more accessible in Russia. “He’s the only one.”

But other people question the extent of Pokrovsky’s commitment. One U.S. AIDS researcher sharply criticized Pokrovsky’s plenary talk in April at an HIV/AIDS conference in Moscow. “He addressed no modern preventive interventions beyond the era of the condom, did not speak to treatment as prevention, did not address the urgency around harm reduction, and did not address pre-exposure prophylaxis for those at risk,” says Chris Beyrer of the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. Beyrer says Pokrovsky may have held his tongue because the conference included many health officials from countries Russia wants to influence. “It was difficult not to come away with the take-home that the federal authorities were again refusing to take an evidence-based approach to the expanding HIV epidemic in Russia.”

Pokrovsky concedes his job is “rather difficult. ... Every morning I say it’s necessary to quit this work if there are no results, no progress. But in the evening, I’m thinking about what problem to go after the next day.” —Jon Cohen



# THE SUNSHINE STATE'S DARK CLOUD

New efforts aim to curb  
Florida's startlingly high HIV infection rate

By **Jon Cohen**, in Florida; Photography by **Misha Friedman**

It's a Tuesday afternoon in April, and doctors at the adult HIV/AIDS clinic at Jackson Memorial Hospital in Miami face their usual onslaught of patients. There's the young, recently diagnosed gay man from Venezuela here for his first appointment. An older gay man who emigrated from Colombia and has been treated at the clinic for 18 years. A 37-year-old Massachusetts native who is battling a heroin addiction, has a drug-related heart condition, and has done time for selling sex. Rounding out the queue are an undocumented grandmother from the Dominican Republic, a mentally challenged and occasionally homeless African-American woman, and an elderly Haitian woman in a wheelchair.

The mosaic of patients represents the major drivers of HIV's spread and the communities hard hit by AIDS in the United States. And it helps explain why in 2016, Miami had the highest new infection rate per capita of any U.S. city: 47 per 100,000 people, according to the Centers for Disease Control and Prevention (CDC). That's more than twice as many as San Francisco, New York City, or Los Angeles.

"Miami is the epicenter of the epicenter of HIV/AIDS in the United States," says Mario Stevenson, a virologist with a thick Scottish brogue who heads the infectious disease department at the University of Miami (UM) Miller School of Medicine, which shares a campus with Jackson Memorial. "There's no abatement in our upward slope."

The rest of the state isn't faring much better. Fort Lauderdale, Orlando, and Jacksonville also made the list of top 10 U.S. cities for rate of new HIV diagnoses. And

more HIV infections progress to AIDS here than in any other state, in part because many infected people who start taking antiretroviral (ARV) drugs don't stick with them.

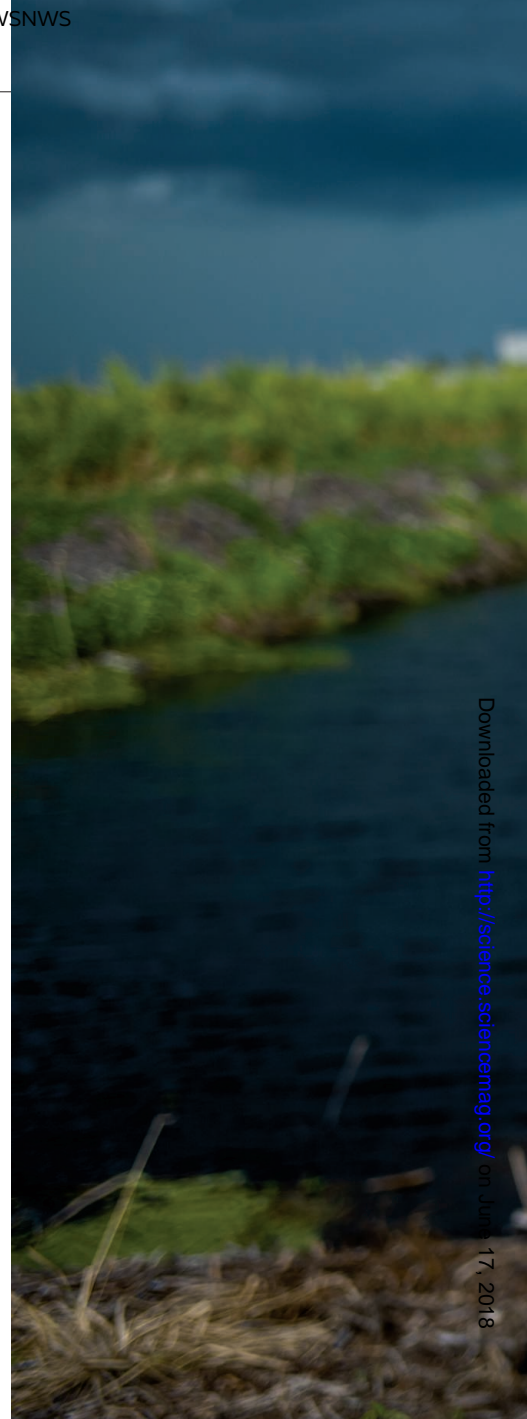
Stevenson, who specializes in HIV cure research, left Massachusetts for Florida in 2010. Since then, he has reached beyond probing how HIV hides in chromosomes to also addressing why the virus still infects and sickens so many people walking the streets outside his lab. "Stemming the tide is going to need more than just working at the bench doing molecular biology," says Stevenson, who has pushed UM to bring together affected communities, legislators, and the Florida Department of Health (DOH) to better coordinate the response to their epidemic. "We're saying, 'Hey, guys, we're in a mess.'"

As Florida's HIV/AIDS caseload keeps growing, more leaders are calling for change. Stevenson helped spearhead a task force that last year issued a report called *Getting to Zero* for Miami. The report built on the treatment-as-prevention principle of the international Ending AIDS movement: Infected people rarely transmit the virus or develop AIDS if they know their status and take ARVs. The report urged Miami to strengthen its testing and treatment efforts—as has happened in San Francisco, New York state, and Vancouver, Canada (*Science*, 17 July 2015, p. 226)—and it promoted a proven prevention strategy called pre-exposure prophylaxis (PrEP): giving ARVs to uninfected people who are at high risk. At the state level, the same agenda is at the heart of DOH's plan to eliminate HIV transmission and AIDS deaths, which

it has rolled out, albeit slowly, over the past few years.

Florida has made some progress, but Stevenson knows the path ahead will be rough. He and others complain that state officials have turned away federal funding for prevention and care, limited sex education, and generally downplayed the problem. The reformers say that in addition to politics, they must tiptoe around cultural minefields, slash through excessive bureaucracy, and tackle the stigma that compromises HIV/AIDS efforts everywhere.

The push for reform also faces indifference, especially among state and local legislators. For them, other illnesses—heart







disease, cancer, diabetes, and Alzheimer's, each of which takes more lives—often eclipse the need for HIV/AIDS resources, Stevenson says. And the epidemic can feel like old news. “Everyone recognizes the problem, but when it comes time on the microphone to say, ‘This is what we’re doing with HIV,’ there’s a disconnect,” he says. “We know how to address the issue. We know how to fix it. We just have to get up to speed.”

**WHY IS HIV SPREADING RAPIDLY** across the Sunshine State, famous for its tropical weather, wide beaches, the Everglades, and amusement parks? The stream of visitors and transplants drawn to those attrac-

tions is one piece of the puzzle. More than 100 million tourists visit Florida each year, some lured by the party-hearty, bacchanalian reputation of Miami, Key West, and other beach towns. Since the AIDS epidemic surfaced in 1981, the state’s population has doubled to nearly 21 million people, many of them immigrants from Latin America or the Caribbean—and in several of the island nations in the region, HIV is substantially more prevalent than in the United States. Florida is also in the deep South, which, because of a surge of HIV infections among black and Latino men who have sex with men (MSM), is home today to nearly half the estimated 1.1 million people living with HIV

Timothy Dean says concerns about stigma and discrimination keep people in his rural, sugar-growing town from visiting the local clinic.

in the country (*Science*, 13 July 2012, p. 168).

According to DOH modeling estimates, Florida in 2016 had 135,986 HIV-infected people, just behind California and New York. But 21,214 of those people—15%—did not know their status, and that fuels the state’s high transmission rate. “A significant fraction of the new infections come from people who don’t even know they have HIV,” says Michael Kolber, clinical director of UM’s HIV/AIDS program.

Florida’s complex demographics mag-



nify the challenge of reaching that 15%. A campaign that might prompt a Haitian immigrant to seek testing differs from what's needed to reach Latinos, blacks, MSM, heterosexuals, drug users, the homeless, or the mentally ill. Urban and rural communities also have vastly different lifestyles and testing options. "It's such a diverse state from the Panhandle to Key West that it's almost like different countries," says Jeffrey Beal, medical director of DOH's HIV/AIDS section, based in Fort Myers.

Many doctors, especially in rural areas, still don't routinely test for HIV, and some are slow to offer infected people treatment. "There's been a lack of attention on the part of my fellow physicians—I have to be very honest about that," says Beal, who began his career in Oklahoma at the start of the AIDS epidemic and burned out from watching so many patients die. He even quit practicing medicine for a year and moved to Florida to become a housing developer with his partner before the shortcomings of the state's HIV/AIDS care in rural areas led him back to the clinic.

At the same time, not enough people seek out testing, further undermining the promise of treatment as prevention. Some fear being ostracized if they're infected: Florida falls in the country's Bible Belt, which CDC notes suffers from "homophobia and transphobia, racism, and general discomfort with public discussion of sexuality." In conservative Clewiston, a few hours' drive north of Miami, Timothy Dean, who is openly gay and has his HIV-positive status tattooed on his arm, sees doctors at the nearby county health department. But he knows several local, infected people who avoid it.

"This town is very tight-knit, and rumors can start," says Dean, who works for DOH to connect HIV-infected people to care. "Your children may not want nothing to do with you because they find out that you're positive." And Clewiston, a sparsely populated community built around growing and processing sugar cane, has no public transportation, so Dean says seeking care outside of town simply isn't a realistic option for some.

Complacency, rather than fear, is the issue for other groups, such as immigrants

from the many countries where HIV is relatively uncommon. MSM from Latin America, for example, may not realize that moving to Florida could boost their risk of infection, says UM HIV/AIDS clinician Susanne Doblecki-Lewis. "The idea that the same behavior can have different consequences in different locations is a little bit difficult to process," Doblecki-Lewis says.

Other public health measures also have lagged here. Needle and syringe exchanges that now are in 32 other states—they were endorsed by the U.S. National Commission on AIDS in 1991—only became legal here in 2016. "We were decades late," says UM

When people test positive, Florida's system routinely links them to care. But the system falls short at keeping them there. Nearly one in three people are, in medical jargon, "lost to follow-up" for reasons such as transportation difficulties, stigma concerns, relocation, lack of social support, substance abuse, and poverty. In New York City, by contrast, which had nearly 90,000 people who knew their positive status in 2016, 88% were retained in care. As a consequence, only about half of the HIV-infected people here are fully suppressing the virus with ARVs, compared with 76% in New York City. Florida's shortfall has two steep costs: People are more likely to spread the virus to others, and they have a greater risk of developing AIDS themselves.

"THIS IS DR. SUZIE, the doctor of the Haitian community!" announces the host of "Radio Coin," a talk show on WLQY-AM in Miami that caters to the city's large Haitian population. Suzie Armas, who runs Miami's New Health Community Center in a neighborhood known as Little Haiti, takes to the microphone like a preacher at the pulpit. In a mix of Creole and English, she urges listeners to come to her clinic, regardless of whether they have insurance or legal immigration status, and receive a free HIV test.

"If you become positive, next step is to seek treatment," says Armas, a physician's assistant who was a licensed doctor in Haiti and Mexico. "We will take care of you."

Haitian immigrants have one of Florida's highest new infection rates for people

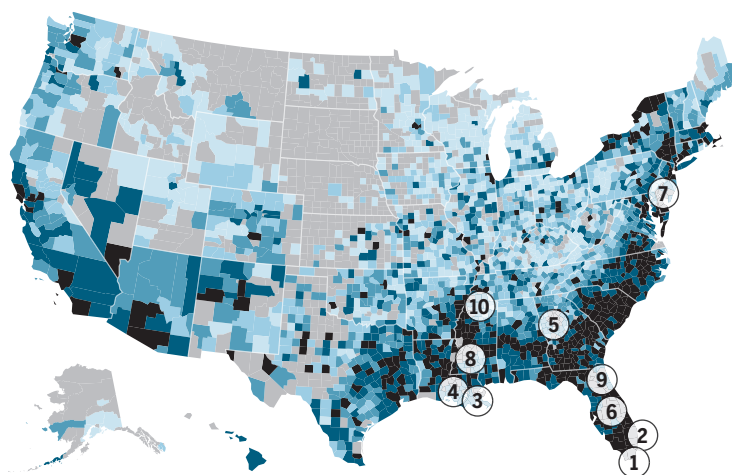
born in other countries, and Armas wants to make sure that people with the virus stay on treatment. But that takes a special effort with clients such as Chimens Point du Jour, 57, who worked as a high school administrator in Haiti but speaks limited English and tested positive for HIV in 2016 during a routine checkup at Armas's clinic. Because of language issues, most newspaper articles or public health campaigns have little impact on people like him, Armas notes, and many other Haitians are uneducated, even illiterate. "But every Haitian living in the U.S. listens to some

## Geography of infection

The southern U.S. states, including Florida, are home to nearly half the people living with HIV in the country, as shown in blue. Four of the 10 metropolitan areas with the highest rates of new infection per 100,000 (numbered dots) are in Florida.

### HIV prevalence (percentile)

0–19 20–39 40–59 60–79 80–99 No data



1. Miami Division, FL 47.0
2. Ft. Lauderdale Division, FL 40.1
3. New Orleans-Metairie, LA 33.3
4. Baton Rouge, LA 30.2
5. Atlanta-Sandy Springs-Roswell, GA 29.4
6. Orlando-Kissimmee-Sanford, FL 27.4
7. Philadelphia Division, PA 25.1
8. Jackson, MS 25.0
9. Jacksonville, FL 22.9
10. Memphis, TN-MS-AR 22.9

clinician Hansel Tookes, who, while still a med student at the school, led a 4-year lobbying campaign that persuaded legislators to change the law.

Similarly, Florida has been slow to promote PrEP, unlike many other locales. Florida Surgeon General Celeste Philip mandated in December 2017 that all 67 county health departments offer PrEP at no cost by the end of 2018. On a smaller scale, UM has a mobile clinic named Gamechanger that will start driving around Miami in the next few months, offering both HIV testing and PrEP.





Physician Hansel Tookes (left), who oversees a needle exchange program in Miami's Overtown neighborhood, meets with Natasha Dixon (right) and Erik Olivero (center).

type of radio station," she says. That makes "Radio Coin" a potent tool.

For Point du Jour, another draw is that the clinic offers him free transportation to appointments. "*Bél bagay!*" he says, which roughly translates to "This place is awesome!"

The center's program for Haitians is but one small-scale attempt to help HIV-infected Floridians stay in care. Another is the Infectious Disease Elimination Act, known as IDEA Exchange, which Tookes and his team began in December 2016. IDEA Exchange offers Miami drug users clean needles, HIV tests, and referrals to rehab and treatment. The exchange also has a mobile unit that travels to the city's drug hot spots.

Shortly after the IDEA Exchange van rolls into Miami's Overtown neighborhood early one April morning, a steady stream of clients shows up. Natasha Dixon, 33, was diagnosed with HIV last year at the van. "I can barely take care of myself," says Dixon, who brings along a kitten that the IDEA Exchange staff later agrees to adopt. A sometimes homeless mother of three, Dixon ran out of ARV pills when she left town to visit her mom. Tookes sees that Dixon urgently needs care and tells her they'll arrange a clinic visit and restart her on the drugs, along with helping her find a place to live. "I want to get my life back together," Dixon says. "This is crazy."

But the Florida health system's bureaucracy works against that sense of urgency. A few weeks later, when Dixon visits Tookes at his clinic, he says he cannot prescribe ARVs to her. Regulations stipulate that she first visit the county health department, meet with a case manager, and then enroll in the federally funded AIDS Drug Assistance Program before receiving treatment. Tookes attributes the breakdown to

### ***"Miami is the epicenter of the epicenter of HIV/AIDS in the United States."***

**Mario Stevenson.** University of Miami Miller School of Medicine

DOH policies, as well as Florida's decision to not accept the federal government's offer of "expanded" Medicaid, which makes it easier for low-income, younger people to receive care.

"There literally should be a conveyor belt, with a red carpet, to move her from the mobile unit into clinic with me sitting there smiling to welcome her into care," Tookes says. "She made a major first step by coming to see us at clinic, but the barriers within our system are almost insurmountable." Instead, after more than 6 weeks had passed, Dixon still was not

back on medication.

Beal says his team recognizes that Florida must revamp its response. He's particularly buoyed by DOH's recent moves. The department has begun to use some new federal money and repurpose state funds to find patients lost to follow-up, make HIV testing a routine part of care, promote PrEP, and immediately offer newly infected people a free, 30-day supply of ARVs while they clear insurance hurdles.

Change is underway. Aside from the new PrEP mandate from the state surgeon general, HIV testing is becoming more routine after a law change that simplifies the consent process. "Disease intervention specialists" at DOH have begun to comb through databases to find and then contact people lost to follow-up. And even Tookes, who is sharply critical of the system in Florida, was heartened in late May when DOH started a new project with IDEA Exchange to rapidly offer treatment to people his team diagnoses.

It's early days. But once those barriers are knocked down, Beal says, "No one can really look me in the face in the state of Florida and say, 'I can't get the medicines I need.'" He does not expect to see an immediate decline in new infections. But, he says, "Hopefully, by 2019, we will." ■

*Reporting for this story was supported by the Pulitzer Center.*



# INSIGHTS

## PERSPECTIVES



Whale sharks are one of many marine species that are endangered but difficult to monitor.

### OCEANS

## *A tool for finding rare marine species*

Environmental DNA analysis shows promise for studying rare and elusive marine species

By **Ellen K. Pikitch**

**B**y their very nature, rare marine species are challenging to study. They often elude detection by established research methods because of their scarcity and patchy distribution. Their small populations demand the use of sampling techniques that minimize the risks of injury or death and that conserve the habitats and ecological communities that support them. At the same time, ever more once-common marine species are becoming rare as threats such as overfishing, invasive species, and environmental destruction reduce their abundance (1). Analysis of environmental DNA (eDNA) shows promise as a rapid, safe, sensitive, and cost-effective means for detecting and studying rare marine species.

Application of eDNA in marine systems is relatively recent, with studies of microorgan-

isms preceding those of macroscopic organisms (2). In the ocean, eDNA is obtained from seawater that contains sloughed skin cells, scales, secretions, and other suspended tissues. Extraction and analysis of the genetic material contained in these tissues can reveal the presence of otherwise hidden marine life. Analysis may be aimed at detection of a single species for which the genome or a portion thereof is known (barcoding), or a broader inspection for many species (metabarcoding).

The use of eDNA in marine studies is challenging because the ocean's sheer size, extreme depths, and high dynamism mean that genetic material is very dilute and is quickly dispersed (3). However, the fact that eDNA persists only briefly in marine environments—on the order of days to weeks (4)—and thus illuminates the “here and now” is an advantage for many research and policy applications (3). In other environments—including cave soils, permafrost, and lake sediments in anoxic basins—eDNA can persist for years or centuries and, in some cases, date as far back as the Pleistocene (5).

eDNA has many advantages over traditional biological sampling methods, such as increased sensitivity, decreased cost, less resource-intensive sampling, less reliance on taxonomic expertise, and noninvasive sampling; however, it also poses challenges, such as unknown and potentially substantial false-positive and false-negative rates and other read errors, possible conflation of multiple vectors, unknown error rates, and unknown source and location of genetic material (3). Recent eDNA studies of a wide variety of marine taxa in habitats, ranging from nearshore kelp forest ecosystems (6) to extreme ocean depths (7), have shown major advances in the design, conduct, analysis, and interpretation of eDNA studies. Recently developed protocols, when followed carefully, greatly reduce the likelihood of false positives and negatives and other potential errors, enhancing confidence in the results. Occupancy and other statistical models enable quantification of detection probabilities across factors such as season, temperature, habitat characteristics, and life stage.

School of Marine and Atmospheric Sciences,  
Stony Brook University, Stony Brook, NY 11794, USA.  
Email: ellen.pikitch@stonybrook.edu



Reference DNA databases may limit application of eDNA methodologies for some species. However, rapid advances are being made to fill gaps through special purpose studies and by broader initiatives, including Genbank, the BarCode of Life, the International Barcode of Life (iBOL) program, and the Earth BioGenome Project (EBP).

### USE OF eDNA TO DETECT SPECIES PRESENCE

Recent studies generally confirm the greater sensitivity of eDNA compared with traditional survey approaches for detecting the presence of species in marine environments. For example, Thomsen *et al.* (2) found eDNA to be superior to each of nine conventional sampling methods. Similarly, Yamamoto *et al.* (8) detected 40% more fish species using eDNA metabarcoding from water samples collected over a 6-hour period than they had recorded during 14 years of underwater visual surveys conducted in the same area. However, eDNA does not always outperform conventional survey methods. For example, Bakker *et al.* (9) did not find nurse shark eDNA in samples taken from an area where the species was known to be present.

Incomplete species detection may be resolved in future studies by using alternative markers and primers, improving reference databases, and using multiple markers. There are also circumstances in which eDNA may be the only practicable approach for studying rare marine species. One such group are chondrichthyans (cartilaginous fishes that include sturgeon, sharks, skates, rays, and chimera fishes), whose life history characteristics make them particularly vulnerable to overexploitation and extinction (10). About a quarter of these species are classified as threatened, endangered, or critically endangered, and for nearly half (46.8%), existing data are insufficient to determine their conservation status (10). Study methods that involve capture of individuals are less efficient and more resource-intensive than eDNA and can cause mortality and sublethal effects such as abortion (11).

Cetaceans (whales, dolphins, and porpoises) are another threatened and difficult-to-study species group for which eDNA may be a highly effective approach. Despite extensive efforts conducted over the span of 30 years, only a quarter of the ocean's surface has been surveyed for cetaceans from ships or aircraft (12). eDNA of killer whales has been detected in seawater samples taken up to 2 hours after their passage and despite water mass movements of several kilometers (13). This indicates the promise of eDNA as an alternative to standard genetic sampling methods for cetaceans; the latter involve invasive sampling and typically require close

approach of a vessel while the animal is at the surface. Given its greater sensitivity and the fact that samples can be obtained from a wide variety of platforms, eDNA has the potential to rapidly fill data gaps for cetaceans.

### USE OF eDNA TO QUANTIFY ABUNDANCE

Several natural and mesocosm studies have shown a relationship between the amount of DNA present in seawater samples and species count data or measures of relative abundance. Weltz *et al.* (14) found higher concentrations of eDNA in tanks that held a greater biomass of marine skates. Correlations of eDNA con-

Despite these promising studies, eDNA alone cannot yet be used to estimate population abundance reliably and with precision. Many factors influence eDNA concentration, including environmental parameters, life stage, size and age composition, changes in adult body condition, seasonal changes in activity levels, shedding rates, eDNA persistence rates, and changes in population size; these factors need to be better understood and quantified. Recent eDNA studies, including (14), explicitly account for and measure the influence of factors that determine eDNA concentration in relation

## How eDNA outperforms other methods

Comparison of the effectiveness, safety, and feasibility of eDNA methods, other noninvasive methods (such as underwater visual survey line transects), and traditional sampling methods involving capture of organisms (such as bottom trawls or longlines), based on the author's assessment. eDNA performs other methods for most attributes, except those that are not discoverable through genetic analysis. The scores reflect current rankings; rankings for eDNA are expected to improve as the method develops. See supplementary materials.

| Attribute   | eDNA    | Other noninvasive | Established capture method |
|---|---------|-------------------|----------------------------|
| <b>Effectiveness</b>  |         |                   |                            |
| Detect species presence   | ● ● ● ● | ● ● ●             | ● ●                        |
| Assess species range distribution   | ● ● ● ● | ● ● ●             | ● ●                        |
| Assess species trends in abundance  | ● ● ●   | ● ● ●             | ● ●                        |
| Estimate species abundance  | ● ● ●   | ● ● ●             | ● ●                        |
| Assess associated ecological community (biodiversity)                                   | ● ● ● ● | ● ● ●             | ● ●                        |
| Determine individual attributes (size, age, sex, diet composition)                      | ● ● ● ● | ● ● ●             | ● ● ● ●                    |
| Assess biological/ecological attributes over time (tag and live release of individuals) | ● ● ● ● | ● ● ●             | ● ● ● ●                    |
| <b>Safety</b>   |         |                   |                            |
| Prevent injury or mortality of individuals sampled                                      | ● ● ● ● | ● ● ● ●           | ● ●                        |
| Prevent injury or mortality of nontarget species  | ● ● ● ● | ● ● ● ●           | ● ●                        |
| Prevent damage to important habitat   | ● ● ● ● | ● ● ● ●           | ● ●                        |
| <b>Feasibility</b>  |         |                   |                            |
| Degree of expertise (taxonomic) needed to conduct sampling                              | ● ● ● ● | ● ● ●             | ● ●                        |
| Labor and time required to sample   | ● ● ● ● | ● ● ●             | ● ●                        |
| Expense of sampling platform and its operation  | ● ● ● ● | ● ● ●             | ● ●                        |
| Cost of sample processing   | ● ● ● ● | ● ● ●             | ● ●                        |

● Least ● Moderate ● ● Best

centrations with abundance or biomass have also been found for marine species in situ, including Japanese jack mackerel (8) and various species caught in bottom trawls (7). Sigsgaard *et al.* (15) used eDNA to determine genetic variation among individual whale sharks (see the photo) within a group and then calculated the female effective population size on the basis of these data. Mitochondrial DNA control region sequences from eDNA samples and tissue samples taken at the same location showed similar haplotype frequencies; estimated effective population sizes for both sample types were on the same order of magnitude.

to population, auguring well for substantial progress in the years ahead. Ancillary studies conducted in laboratories, microcosms, and mesocosms, comparisons of eDNA and traditional methods conducted in parallel, and research conducted over fine temporal and spatial scales and under various environmental conditions will assist progress toward abundance estimation.

### COMPARISON OF eDNA AND OTHER METHODS

Overall, eDNA analysis is more effective, safe, and feasible than other noninvasive sampling techniques and conventional capture proce-



dures (see the figure). Given its advantages over conventional survey methodologies, eDNA is likely to quickly become the method of choice for detecting rare and elusive marine species and determining their distributions and ranges.

Estimation of the relative and absolute abundance of mobile aquatic organisms is an ongoing challenge for eDNA and conventional study methods alike. Because of their long history of use, conventional techniques currently have clear advantages over eDNA for assessing abundance. However, given its rapidly increasing use and the expected concomitant advances in knowledge, the utility of eDNA may catch up quickly and may ultimately overtake that of conventional techniques because of its greater sensitivity and nondestructive character. In the interim, finding ways to incorporate eDNA information into existing comprehensive analysis frameworks is likely to improve abundance estimates.

The greatest limitation of eDNA methodologies is that they can only inform us about attributes of the living world that are discoverable from genetic material. Many species, community, and ecosystem traits and interactions are not currently discernible from analysis of eDNA, including life stage, size, age, growth rates, physiological processes, behaviors, trophic dynamics, and the relationships between organisms and their environment. Understanding these complex aspects of our living world is vital, and eDNA can thus only complement, rather than replace, other research approaches. ■

#### REFERENCES AND NOTES

1. D. J. McCauley *et al.*, *Science* **347**, 1255641 (2015).
2. P. F. Thomsen *et al.*, *PLOS ONE* **7**, e41732 (2012).
3. R. P. Kelly *et al.*, *Science* **344**, 1455 (2014).
4. M. A. Barnes, C. R. Turner, *Conserv. Genet.* **17**, 1 (2016).
5. V. Slon *et al.*, *Science* **356**, 605 (2017).
6. J. A. Port *et al.*, *Mol. Ecol.* **25**, 527 (2016).
7. P. F. Thomsen *et al.*, *PLOS ONE* **11**, e0165252 (2016).
8. S. Yamamoto *et al.*, *Sci. Rep.* **7**, 40368 (2017).
9. J. Bakker *et al.*, *Sci. Rep.* **7**, 16886 (2017).
10. N. K. Dulvy *et al.*, *eLife* **3**, e00590 (2014).
11. K. R. Adams, L. C. Fetterplace, A. R. Davis, M. D. Taylor, N. A. Knott, *Biol. Conserv.* **217**, 11 (2018).
12. K. Kaschner, N. J. Quick, R. Jewell, R. Williams, C. M. Harris, *PLOS ONE* **7**, e44075 (2012).
13. C. S. Baker, D. Steel, S. Nieuwkerk, H. Klinck, *Front. Mar. Sci.* **5**, 133 (2018).
14. K. Weltz, J. M. Lyle, J. Ovenden, J. A. T. Morgan, D. A. Moreno, *PLOS ONE* **12**, e0178124 (2017).
15. E. E. Sigsgaard *et al.*, *Nat. Ecol. Evol.* **1**, 0004 (2016).

#### ACKNOWLEDGMENTS

I thank P. Tompkins and R. Silver for help with background literature assembly and for preparation of the manuscript. I am also grateful for helpful conversations with S. Baker, M. Shivji, L. Crowder, and F. Chavez as well as constructive comments from several anonymous reviewers. Any errors or misinterpretations are my own.

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6394/1180/suppl/DC1

10.1126/science. aao3787

## NEUROSCIENCE

# Crystallizing a memory

Researchers identify and modulate synaptic correlates of a memory engram

By Steve Ramirez

**W**hat is the physical basis of memory? What does it take to retrieve a memory in the brain? What would it take to activate or erase memories? In the early 20th century, the German zoologist Richard Semon coined the term “engram” to denote the physical manifestation of a memory in the brain (1). Two decades later, Canadian psychologist Donald Hebb posited a physiological correlate for learning and recollection: The process of learning strengthens the connections, or synapses, between neurons, which leads to the development of brain-wide cell assemblies that undergo changes in their structural and functional connectivity (2). The coordinated activity of these assemblies—called ensembles, traces, or engrams—that occurs during learning (memory formation) is thought to be reengaged during recall and thereby forms a stable neuronal correlate of memory (2). As subsequent memories are formed, the dynamics of these assemblies evolve and provide preexisting scaffolds to influence how the brain processes the variety of memories an organism forms. Studies by Abdou *et al.* (3) on page 1227 of this issue and by Choi *et al.* (4) develop new technologies to visualize discrete engrams in the brain and modulate them in a synapse-specific manner to understand memory strength and memory restoration from an amnesic state. This improved understanding could eventually be translated to modulate memories to alleviate maladaptive memory states.

Hebb's conceptualization of memory in the brain became an oft-quoted creed in brain science: Neurons that fire together wire together. In the spirit of Semon, cells that are active during learning, that undergo enduring learning-induced changes, and that facilitate recollection are referred to as engram cells. A physical manifestation of Hebb's principle, of engram cells communicating and linking with one another during learning, has been recently demonstrated in mice (2): A discrete ensemble of hippocampal cells that were

simultaneously active during learning preferentially strengthened their structural and functional connectivity relative to quiescent cells. It was a remarkable demonstration of hippocampal engram cells firing together at the time of learning and physically interlinking together to facilitate memory retrieval.

Choi *et al.* developed an activity-dependent strategy to tag and visualize not just active engram cells but also active synapses between engram cells and non-engram cells. They used a clever trick: They engineered a system in which presynaptic and postsynaptic membranes of a neuron

**“...modulating memories... reveal both how memory naturally works and, when artificially controlled, how memory can work.”**

had complementary green fluorescent protein (GFP) fragments that reconstituted a functional GFP on synapse formation. Excitingly, by using fluorescent proteins of different colors, the researchers were able to visualize two different presynaptic neurons that projected to a single postsynaptic cell. This allowed the authors to measure how learning modulates connectivity between engram cells, engram to non-engram cells, between non-engram cells, and non-engram to engram cells. Their results are striking: Learning induced preferential increases in synaptic connectivity specifically between engram cells and not between non-engram cells. They also found that, although weak and strong fear memories activated a similar proportion of cells in the hippocampus, a stronger fear memory elicited stronger connectivity (that is, a higher density of synapses and potentiation) specifically between engram cells.

A cell ensemble can process multiple memories, but how the same population of cells can encode separate memories has remained unclear. Abdou *et al.* combined cutting-edge techniques to visualize and directly modulate discrete memories in a synapse-specific manner. They used an

Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA. Email: dvsteve@bu.edu

auditory fear conditioning task in which a tone is paired with a foot shock such that mice subsequently display fear responses to hearing the tone without foot shock. Of the myriad of neural circuits involved in a memory, the auditory cortex (AC), the medial geniculate nucleus (MGN), and the lateral amygdala (LA) are key nodes involved in processing an auditory fear memory.

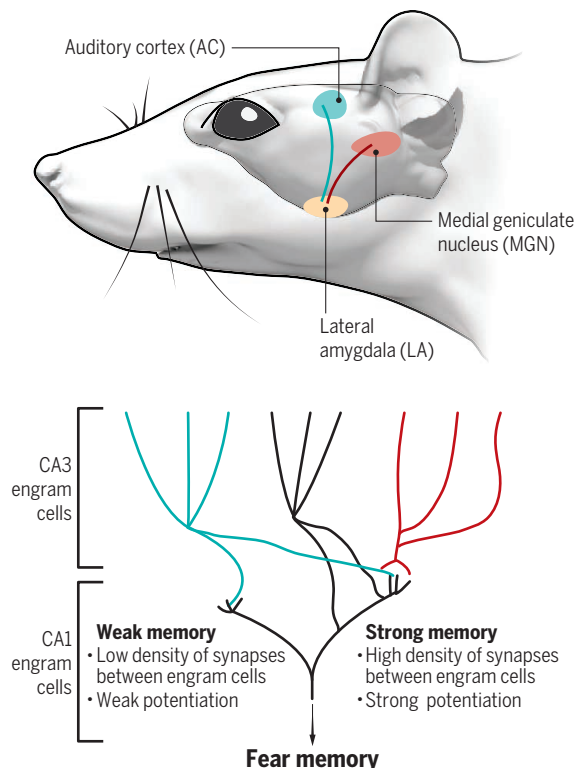
Abdou *et al.* leveraged recent findings supporting the idea that blocking protein synthesis at the time of memory formation induces partial amnesia by impairing memory retrieval (5). They tagged engram cells in the AC and MGN and found that stimulating their axonal terminals in the LA still elicited fear responses, even under partial amnesic conditions. In an attempt to induce complete amnesia, the authors both blocked protein synthesis and induced autophagy—a degradation process in which the cytosolic constituents of a cell are recycled. Under these conditions, stimulation of AC or MGN terminals in the LA failed to induce fear. Within these parameters, the memory had been erased—a compelling finding given that, previously (2), memories had been fully erased by ablating LA cells rather than by inhibiting cellular processes involved in memory formation.

The authors permanently rescued memories specifically in the partial amnesia cases by repeatedly strengthening, or potentiating, the connections between AC or MGN terminals in the LA. However, a memory could not be recovered after inducing complete amnesia. These fascinating results indicate that memories can be artificially restored, though only under specific cases of partial amnesia when retrieval rather than learning is defective.

In a technical tour de force, Abdou *et al.* then modulated two memories that recruited the same assembly of neurons. They observed that two temporally overlapping fear memories elicited activity in overlapping sets of LA cells, consistent with previous data (6). Importantly, they demonstrated that the acquisition of two fear memories engaged two different sets of synapses from the AC and MGN, which interacted with the overlapping neural ensembles in the LA—as such, each memory could be individually, and lastingly, restored or suppressed by modulating the activity levels of each of their respective synapses in the LA.

## The microstructure of recollection

The auditory cortex, medial geniculate nucleus, and lateral amygdala are regions in the mouse brain that process auditory fear memory. Strong or weak fear memories depend on the synaptic connectivity between neuron ensembles (engrams) in these regions.



These data provide a tantalizing demonstration of linking and controlling an individual memory amid the ocean of experience that a mouse remembers. They provide a glimpse into the microstructure of recollection. Embedded within and across subregional activity are discrete circuits with unique histories that sculpt the morphological and physiological properties of a neuron and, by extension, of a memory. Engrams are not localized to a single X-Y-Z coordinate in the brain, but rather appear to be distributed with key nodes in the brain being necessary, sufficient, or both, to regulate individual components of memory (see the figure). Semon's engram recruits Hebb's assemblies in a brain-wide manner.

Modern neuroscience has isolated cells active during defined periods of learning and recollection to observe their behavior and to test for their necessity and sufficiency for a given behavioral readout of memory. The advances in modulating memories in mice have been extraordinary: Researchers have been able to visualize stable neural correlates of memory (7–9), to allocate and erase a specific memory (2), to reactivate a memory (10, 11), to

temporarily inhibit a memory (12, 13), to connect and create artificial memories and spatial maps (14, 15), and to bring back memories once thought to be lost to amnesia (3–5). They provide conceptual scaffolds for subsequent research aimed at modulating memories to reveal both how memory naturally works and, when artificially controlled, how memory can work. Moreover, researchers can now simultaneously attempt to modulate memories in a therapeutic context to alleviate clusters of symptoms underlying maladaptive states. Although the same techniques cannot yet be used in humans, it is important to translate these concepts from rodent to human memory and back. Once this is achieved, the goals are abundant: for example, to prevent the return of fear memories in posttraumatic stress disorder; to reawaken a memory in Alzheimer's disease; and to manipulate positive memories to boost mood.

The field is burgeoning with questions: In any given circuit, what are the real-time physiological kinetics of engram cells? How do these cells behave as subsequent learning experiences occur? How does the passage of time influence

the physical properties of engram cells? How would mimicking the endogenous, learning-induced firing patterns of the brain affect the behavioral expression of a memory compared to artificial parameters that yield a behavioral response? How do different neural circuits differentially process the mnemonic content associated with an experience? Choi *et al.* and Abdou *et al.* have given us exciting new tools and concepts that, together, bring us closer to unweaving the magnificent neural knot we call memory. ■

## REFERENCES

1. R. Semon, *The Mneme* (Allen & Unwin, 1921).
2. S. A. Josselyn *et al.*, *Nat. Rev. Neurosci.* **16**, 521 (2015).
3. K. Abdou *et al.*, *Science* **360**, 1227 (2018).
4. J.-H. Choi *et al.*, *Science* **360**, 430 (2018).
5. T. J. Ryan *et al.*, *Science* **348**, 1007 (2015).
6. A. J. Rashid *et al.*, *Science* **353**, 383 (2016).
7. L. G. Reijmers *et al.*, *Science* **317**, 1230 (2007).
8. K. M. Christian *et al.*, *Learn. Mem.* **10**, 427 (2003).
9. N. M. Weinberger, *Nat. Rev. Neurosci.* **5**, 279 (2004).
10. X. Liu *et al.*, *Nature* **484**, 381 (2012).
11. K. K. Cowansage *et al.*, *Neuron* **84**, 432 (2014).
12. K. Z. Tanaka *et al.*, *Neuron* **84**, 347 (2014).
13. C. A. Denny *et al.*, *Neuron* **83**, 189 (2014).
14. S. Ramirez *et al.*, *Science* **341**, 387 (2013).
15. S. Trouche *et al.*, *Nat. Neurosci.* **19**, 564 (2016).

10.1126/science.aau0043



## NEUROSCIENCE

# Scaling of human brain size

Higher cognitive regions are preferentially expanded in individuals with larger brains

By David C. Van Essen

What makes humans unique as a species and as individuals? Our uniqueness stems from language, tool use, reasoning, and other cognitive abilities that are largely mediated by specialized regions of the cerebral cortex. These regions of higher cognitive function have expanded disproportionately during human evolution (compared with nonhuman primates) and during postnatal maturation, when cortical surface area expands threefold between infancy and adulthood (1). Our uniqueness as individuals reflects countless differences in brain structure, function, and connectivity. One basic anatomical difference between similarly aged individuals is a more than 1.5-fold variation in total brain size (and total cortical volume) (2). On page 1222 of this issue, Reardon *et al.* (3) bring this aspect of individual variability under the umbrella of "differential scaling" by showing that human brains of different sizes do not scale uniformly across all regions. Rather, larger brains show greater expansion in regions associated with higher cognition and less expansion in regions associated with sensory, motor, and limbic (emotion- and affect-related) functions.

A simple *a priori* hypothesis is that brains of different size might be linearly scaled versions of one another. However, there is already evidence against this hypothesis, insofar as the cerebral cortex is a mosaic of many cortical parcels (areas) that each show more than twofold individual variability in size (4, 5). But are individual differences in the size of various parcels correlated systematically, for example, according to function? Reardon *et al.* analyzed data from more than 3000 healthy individuals, drawn from three independent

cohorts: the Philadelphia Neurodevelopmental Cohort (PNC) and National Institutes of Health (NIH) cohort (each comprising children and young adults), as well as the Human Connectome Project (HCP) cohort (comprising young adults only). By using *in vivo* structural magnetic resonance imaging (MRI) scans of individual brains, surface models of the cerebral cortex were generated and aligned to a surface-based cortical atlas. Local cortical surface area was then expressed in relation to individual differences in total cortical surface area (see the figure). Notably, both age and sex were ruled out as confounding factors, even though average brain size differs by age and sex (2). Areal scaling maps show broad similarities across

the three cohorts in terms of which regions are expanded in larger brains (positive scaling) and which are less expanded in larger brains (negative scaling). However, there are many differences across the three cohorts, and regions that pass statistical significance for only one or two cohorts might not reflect genuine neuroanatomical effects. Given concerns about reproducibility (6), it is notable that Reardon *et al.* carried out what is effectively a multicohort reproducibility analysis.

The authors compared areal scaling maps to independent measures related to cortical evolution, development, function, and gene expression. They found modest correlations with maps of evolutionary expansion (compared to nonhuman primates) and postnatal developmental expansion (7). Comparisons with neuronal networks when at rest, mapped using functional MRI (8), reveal that the "default-mode" network (which is more active at rest) tends to show positive areal scaling, whereas the limbic network tends to show negative areal scaling. Comparisons with gene expression maps from postmortem human brains (8)

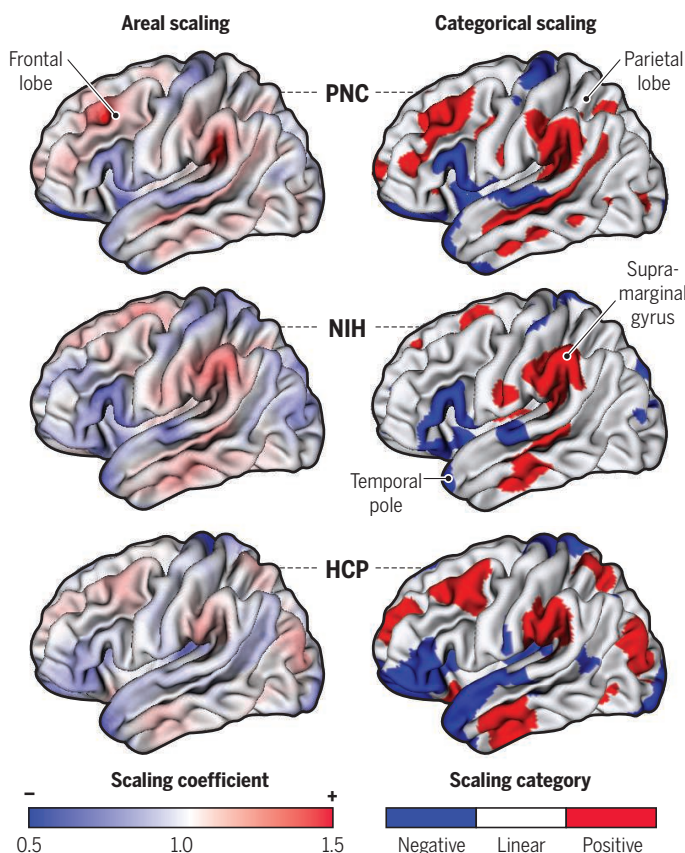
reveal that areal scaling is more positive in regions enriched in mitochondria-related genes and show high energy consumption at rest. These regions have lower myelin content within gray matter and greater synaptic plasticity (9) and have neurons with larger dendritic arbors and more dendritic spines (10).

Reardon *et al.* also examined other brain structures using a subcortical surface-based approach. They found positive- and negative-scaling regions within each structure, indicating that size-related differences are not restricted to the neocortex. For technical reasons, they excluded the cerebellum from their analysis, but this would be interesting to examine in the future, as the cerebellum is involved in cognition as well as coordination of movement (11).

One important issue is whether some of the underlying neuroanatomical scaling patterns are sharper than the relatively smooth gradations shown in their maps. Functionally corresponding regions were not consistently aligned between individuals because alignment was driven by cortical-folding patterns, which are variable and imperfectly correlated with cortical areal

## Brain size variation

Maps of areal scaling relative to total brain size for the PNC, NIH, and HCP cohorts. The supramarginal gyrus is positively scaled and the temporal pole is negatively scaled in all three cohorts, but differences in scaling between the cohorts also exist. IQ correlates with overall cortical surface area.



Washington University, St. Louis, MO, USA.  
Email: vanessen@wustl.edu

boundaries (4, 5, 12); in addition, the authors extensively smoothed the data for methodological reasons. The HCP dataset is well suited for further exploration because it was aligned using areal features rather than only cortical folding (5, 13) and it also includes a 180-area-per-hemisphere multimodal cortical parcellation that has been accurately delineated in individuals and as a group average (5). This should enable analysis of scaling relationships determined for each cortical parcel, which would circumvent the confounds of imperfect intersubject registration.

Another issue is the possible relationship between the size of different brain regions and behavior. Reardon *et al.* found that the intelligence quotient (IQ) significantly correlates with overall cortical surface area (higher IQ is observed in individuals with more cortex, after factoring out age and sex). Others have reported that variability in "functional connectivity" in individuals at rest appears to be greatest in regions of higher cognitive function, including those associated with personality, intelligence, visual perception, and memory performance (14). Intersubject variations in behavior and lifestyle that are predictable from functional connectivity may largely reflect individual differences in the spatial arrangement of functional regions, perhaps including their size, particularly in regions of higher cognitive function (15).

Questions of how areal scaling emerges during brain development and maturation are also intriguing to consider. Are areal-scaling differences driven by genetic factors and/or influenced by environmental factors? Do regions that are larger in some individuals have a greater number of neurons and/or a larger fraction of neuropil (dendritic, axonal, synaptic, and glial arborizations)? These questions are amenable to analysis through large-scale human neuroimaging projects combined with advances in postmortem and in vivo anatomical methods. ■

#### REFERENCES AND NOTES

1. J. Hill *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13135 (2010).
2. J. N. Giedd *et al.*, *Neuropsychopharmacology* **40**, 43 (2015).
3. P. K. Reardon *et al.*, *Science* **360**, 1222 (2018).
4. K. Amunts *et al.*, *J. Comp. Neurol.* **412**, 319 (1999).
5. M. F. Glasser *et al.*, *Nature* **536**, 171 (2016).
6. M. Baker, *Nature* **533**, 452 (2016).
7. B. T. Yeo *et al.*, *J. Neurophysiol.* **106**, 1125 (2011).
8. M. J. Hawrylycz *et al.*, *Nature* **489**, 391 (2012).
9. M. F. Glasser *et al.*, *Neuroimage* **93**, 165 (2014).
10. G. N. Elston, I. Fujita, *Front. Neuroanat.* **8**, 78 (2014).
11. R. L. Buckner, *Neuron* **80**, 807 (2013).
12. D. C. Van Essen *et al.*, *Cereb. Cortex* **22**, 2241 (2012).
13. E. C. Robinson *et al.*, *Neuroimage* **167**, 453 (2017).
14. S. Mueller *et al.*, *Neuron* **77**, 586 (2013).
15. J. D. Bijsterbosch *et al.*, *eLife* **7**, e32992 (2018).

#### ACKNOWLEDGMENTS

The author thanks C. Donahue for comments.

10.1126/science.aat8948

#### ECOLOGY

## Animals feel safer from humans in the dark

### Mammals shift their activities to twilight and night hours in response to human disturbance

By Ana Benítez-López

About 75% of Earth's land surface is currently modified by human activities (1). The expanding footprint of human activities is not only causing the loss of habitat and biodiversity but also affecting the dynamics of wildlife populations. Researchers have long examined human-induced spatial shifts in the distribution of wildlife, but temporal adjustments in animal activity have received less attention. On page 1232 of this issue, Gaynor *et al.* (2) present a comprehensive meta-analysis to quantify the increase in wildlife nocturnality due to human disturbance.

About 50 years ago, Walther suggested that animals perceive human disturbance similarly to predation risk (3). According to this risk-disturbance hypothesis, animal responses to predation risk and anthropogenic disturbance stimuli create similar trade-offs between avoiding perceived risk and pursuing other fitness-enhancing activities, including feeding, parental care, or mating displays (4). Although these responses allow human-wildlife coexistence, they may have important effects on animal fitness through indirect effects on survival and reproduction.

Humans, as day-active (diurnal) apex "superpredators" (5), instill fear in other wildlife like no other predatory species. Animals usually respond by reducing movement rates (6) and spatially avoiding anthropogenic activities (7). However, as wilderness areas disappear, there is little opportunity for animals to spatially avoid humans. In highly disturbed areas, animals may substitute spatial refuges by temporal refuges, with animals shifting daytime activities to the twilight or night hours (see the figure). Such temporal partitioning is a common response in animal communities that allows coexistence between competitive species and shapes predator-prey dynamics. However, until recently, the effect of human disturbance on animal temporal activities

has been difficult to assess, particularly in secretive wildlife.

In recent decades, the advent of technologies, such as satellite and global positioning system (GPS) telemetry or camera traps, has made it possible to monitor wildlife activity more accurately. Gaynor *et al.* have now collated data from 76 studies of 62 mammal species from different locations across the world and have quantified the effects of several forms of human disturbance on wildlife. They conclude that nocturnality is a universal behavioral adaptation of wildlife in response to humans.

In their analysis, the authors compared activity patterns of mammals in areas with high and low human disturbance. They report that mammals increased their nocturnality by a factor of 1.36 across continents, habitats, taxa, and human activities. This means an increase of ~20% in nocturnality on average. Furthermore, out of 141 identified mammal responses to human disturbance, 83% corresponded to an increase in nocturnality; larger mammals exhibited a slightly stronger response than smaller mammals, either because they are more likely to be hunted (8) or as a result of an increased chance of human encounter. Lethal (hunting) and nonlethal activities had comparable effects on mammal activity, supporting Walther's seminal idea on the similarity in animal perception of predation risk and human disturbance (3, 4).

The consequences of human-altered patterns in the activity of wildlife are manifold and are not necessarily limited to the disturbed species or population. The diminished ability of apex predators to hunt at night may impair their role as top-down regulators, and predator-prey interactions may change drastically, whereas prey species that become more nocturnal to avoid humans may be more susceptible to predation by nocturnal predators. Further, human impacts may constrain the temporal partitioning of carnivores exploiting the same prey communities, thus altering competitive dynamics among carnivores by increasing temporal overlap while hunting. Human-altered interspecific competition dynamics may in turn enable human-tol-

Department of Environmental Science, Institute for Water and Wetland Research, Radboud University, 6525 HP Nijmegen, Netherlands. Email: a.benitez@science.ru.nl



erant species to outcompete human-sensitive ones. Additionally, lethal activities may drive rapid evolutionary changes in seed size through the functional extinction of large seed dispersers in tropical forests (9), with concomitant shifts in ecosystem composition and structure. Finally, there is growing evidence of microevolutionary changes associated with adaptive responses to urban environments, including changes in bird song frequency (in response to anthropogenic noise) and increased sedentari-ness in urban blackbirds (10).

Wildlife responses are not limited to spatial or temporal avoidance of human disturbance. Although not included in their analysis, Gaynor *et al.* highlight other behavioral adjustments to cope with human activities, including an increase in vigilance rates and reduction in foraging activity, which may be detrimental for individual condition and reproductive success. However, there are many other responses that organisms use to cope with increased disturbance. Indeed, some animals may not display any visible behavioral adjustment in response to humans, giving the impression that human disturbance has no remarkable effect in this species. Yet, at the physiological level, animals may increase the heart rate and the rate of glucocorticoid production (a physiological indicator of stress in wildlife) through the activation of stress responses (11, 12). The short-term release of glucocorticoids is an adaptive response that redirects energy from nonvital activities toward survival, but prolonged exposure to stressors and the cumulative effects of maintaining elevated glucocorticoid levels may impair immune and reproduction functions (13). Eventually, long-term disturbances may result in lower fitness, lower juvenile survival, or lower reproduction rates (14), with negative consequences at population level.

Given the continuing increase of the global human footprint (1), Gaynor *et al.*'s study is timely and of paramount importance for understanding the influence that humans may have on the behavior of diurnal, twilight-active, and nocturnal wildlife. Similar studies are needed to assemble the multiple responses to human disturbance of nonmammalian taxa. It is also important to keep in mind that the modulation of activity patterns is just one of the multiple responses of wildlife to anthropogenic activities. Holistic approaches that take into account behavioral, physiological, population, and evolutionary responses to human disturbance across taxa are urgently needed to fully understand the consequences of human encroachment for the persistence of wildlife populations. This knowledge will be crucial to develop new tools in conservation planning that address temporal human-wildlife interactions, similar to the way in which spatial ecology is currently incorporated for land planning or spatial conservation prioritization. ■

#### REFERENCES AND NOTES

1. O. Venter *et al.*, *Nat. Commun.* **7**, 12558 (2016).
2. K. M. Gaynor, C. E. Hohnowski, N. H. Carter, J. S. Brashares, *Science* **360**, 1232 (2018).
3. F. R. Walther, *Behaviour* **34**, 184 (1969).
4. A. Frid, L. Dill, *Conserv. Ecol.* **6**, 11 (2002).
5. C. T. Darimont, C. H. Fox, H. M. Bryan, T. E. Reimchen, *Science* **349**, 858 (2015).
6. M. A. Tucker *et al.*, *Science* **359**, 466 (2018).
7. A. Benítez-López, R. Alkemade, P. A. Verweij, *Biol. Conserv.* **143**, 1307 (2010).
8. A. Benítez-López *et al.*, *Science* **356**, 180 (2017).
9. M. Galetti *et al.*, *Science* **340**, 1086 (2013).
10. M. J. McDonnell, A. K. Hahs, *Annu. Rev. Ecol. Syst.* **46**, 261 (2015).
11. B. G. Walker *et al.*, *Conserv. Biol.* **19**, 1571 (2005).
12. R. Tarjuelo *et al.*, *Behav. Ecol.* **26**, 828 (2015).
13. R. M. Sapolsky, L. M. Romero, A. U. Munck, *Endocrine Rev.* **21**, 55 (2000).
14. E. H. Strasser, J. A. Heath, *J. Appl. Ecol.* **50**, 912 (2013).

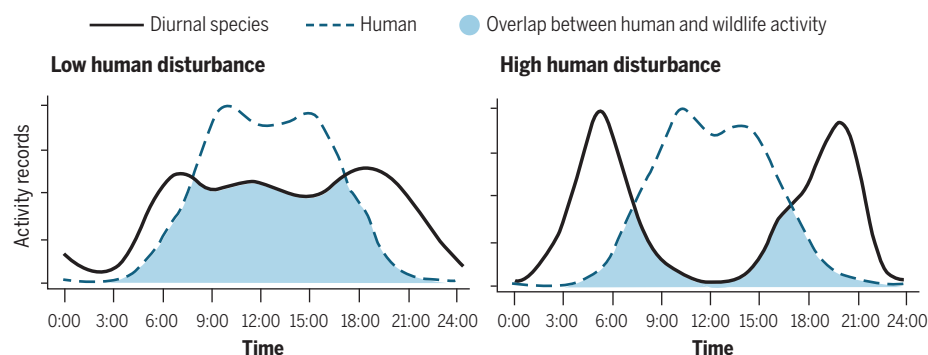
#### ACKNOWLEDGMENTS

M. Huijbregts, L. Santini, and A. Schipper provided valuable comments on the manuscript.

10.1126/science.aau1311

## Human disturbance shifts animal activity into the night

Animals that are active during the day in areas with low human disturbance (left) change their activity patterns in areas with high human disturbance (right). Instead of a broad distribution of activity throughout the day, their activity peaks in the early morning and again in the early evening.



## NEUROSCIENCE

# Facing your fears

## Activation of remote fear memory promotes fear attenuation

By Paul W. Frankland and Sheena A. Josselyn

Remembering traumatic fearful events is adaptive. However, treating no-longer-threatening situations as dangerous may be maladaptive and lead to anxiety disorders, including phobias and posttraumatic stress disorder. Central to many forms of therapy designed to tackle these anxiety disorders is the idea that to overcome fear, one needs to face it. For instance, cognitive behavioral therapy and exposure therapy allow patients to confront the objects or situations that provoke their anxiety in the controlled environment of the therapist's office. With repeated exposures, the patients' anxiety levels gradually decline, and the objects or situations that they once feared no longer trouble them. On page 1239 of this issue Khalaf *et al.* (1) provide a neural mechanism in mice for "facing one's fears." These findings may inform the development of more effective forms of treatment for anxiety disorders.

The therapeutic strategy of facing your fear has its roots in Ivan Pavlov's studies of classical fear conditioning in dogs. In his experiments, Pavlov paired an otherwise innocuous stimulus (such as a buzzer) with an aversive stimulus (such as an electric shock). Subsequent presentation of the buzzer (the now-conditioned stimulus), by virtue of its association with the electric shock (the unconditioned stimulus), evoked fearlike behaviors (conditioned responses) in the dogs. However, similar to patients in therapy, repeated presentations of the conditioned stimulus alone (in the absence of the unconditioned stimulus) eventually led to a decline in conditioned fearful responding (called behavioral extinction) (2).

Khalaf *et al.* provide fundamental insights into the neural mechanisms underlying behavioral extinction in mice. During memory formation, populations of neurons (neuronal ensembles) become co-

Department of Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada.  
Email: paul.frankland@sickkids.ca; sheena.josselyn@sickkids.ca

active. Almost 70 years ago, Donald Hebb proposed that memory formation involves the strengthening of synaptic connections between neurons in these ensembles (3). This synaptic strengthening was postulated to ensure that, given an appropriate retrieval cue, the pattern of activation in the ensemble during memory formation would be recapitulated, promoting memory recall. These neuronal ensembles are thought to correspond to engrams—the neural representation of an encoded event or experience (4). Engrams exist in one of two states. Most of the time, engrams are dormant, with only the potential for memory retrieval. However, an appropriate retrieval cue reactivates the engram to promote memory recall (5).

Until recently, discussions about how engrams were engaged in memory formation and extinction were largely theoretical. However, the development of new tools has provided more direct evidence that neuronal ensembles active during memory encoding correspond to the engram. In mice, researchers are now able to permanently “tag” active neuronal ensembles at, for example, the time of learning with genetically encoded fluorescent reporters or actuators of neuronal activity [optogenetic opsins or chemogenetic DREADDs (designer receptors exclusively activated by designer drugs)]. This tagging process allows ensembles to be visualized or manipulated (inhibited or activated) at later time points (6). Consistent with Hebb’s proposal, neuronal populations engaged during learning are reengaged at above-chance levels during memory recall (7). Moreover, tagged ensembles appear to be both necessary and sufficient for memory recall, because inhibiting the activity of these ensembles prevents memory recall (8), whereas activation of these populations leads to the artificial (and involuntary) expression of the memory in the absence of retrieval cues (9).

Khalaf *et al.* take advantage of engram-tagging techniques to examine how fear memories are extinguished. They used a contextual fear-conditioning task, in which mice learned an association between a conditioning box (or contextual conditioned stimulus) and a mild foot shock (the unconditioned stimulus) (see the figure). Mice froze when placed in the box 1 month later, indicating that they recognized and feared this context. Active neuronal ensembles were permanently

tagged at this remote memory test. Mice were then repeatedly reexposed to the conditioning box in the absence of foot shock to induce extinction, allowing the authors to ask whether the tagged neurons were selectively reactivated during extinction learning. “Engram neurons” in the dentate gyrus, a structure within the hippocampus, were preferentially reactivated during extinction. Moreover, the degree of reactivation correlated with levels of fear reduction, suggesting that engram reactivation may drive extinction.

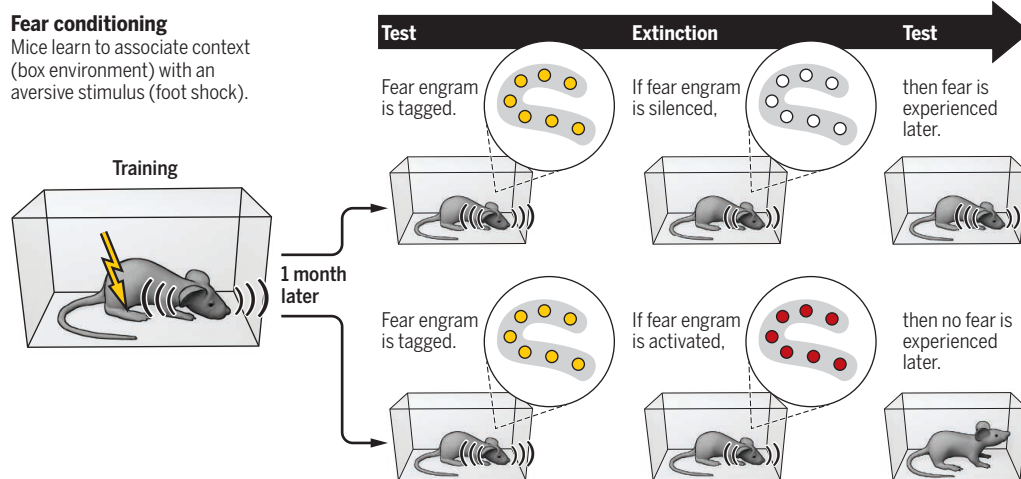
Khalaf *et al.* directly tested whether engram reactivation and fear attenuation are

fear extinction, it may be impossible to develop pharmacological agents to facilitate behavioral therapy for anxiety disorders or to optimize behavioral interventions.

After extinction training, fear frequently returns. In the clinic, relapse represents a major hurdle in therapies designed to reduce anxiety and fear. In the lab, extinguished conditioned fear responses may return with the passage of time (spontaneous recovery), after an aversive event (reinstatement) or in a new unextinguished context (renewal), suggesting that extinction produces new learning, rather than erasing the original memory engram (10–12). Whether

## Facing fear is key for extinction

One month after fear conditioning in a box, mice exhibit engram activation in the dentate gyrus and fear when placed back into the same box. Mice in which the fear engram was silenced during extinction continued to exhibit fear in the box, whereas mice in which the fear engram was activated during extinction showed no fear in the box.



causally related. Inactivating tagged neuronal ensembles in the dentate gyrus slowed behavioral extinction, suggesting that engram reactivation plays a necessary role in fear extinction. Conversely, when the activity of engram neurons in the dentate gyrus was artificially increased during extinction training, fear attenuation was facilitated. No effects on extinction were noted if the activity of a similarly sized population of neurons that were activated by placement in a new context (not the feared context) was manipulated. These results are consistent with the idea that directly activating the engram supporting the fearful memory (akin to facing your fears) is necessary for fear extinction. In many respects, these findings confirm what any accomplished therapist already knows—that, to a large degree, patients with anxiety disorders must relive their trauma to overcome it. Importantly, however, they also identify the neural mechanism underlying this effect. Without understanding the neural basis of

direct activation of the engram decreases the return of fear after extinction training is an important unanswered question that may have direct clinical relevance and answer long-standing questions as to whether extinction changes or erases the original fear memory. ■

## REFERENCES

1. O. Khalaf *et al.*, *Science* **360**, 1239 (2018).
2. I. P. Pavlov, *Conditional Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* (Oxford Univ. Press, 1927).
3. D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory* (Wiley, 1949).
4. S. A. Josselyn, S. Kohler, P. W. Frankland, *Nat. Rev. Neurosci.* **16**, 521 (2015).
5. M. Moscovitch, in *Science of Memory: Concepts*, H. L. Roediger III, Y. Dudai, S. M. Fitzpatrick, Eds. (Oxford Univ. Press, 2007).
6. S. Tonegawa *et al.*, *Neuron* **87**, 918 (2015).
7. L. G. Reijmers *et al.*, *Science* **317**, 1230 (2007).
8. K. Z. Tanaka *et al.*, *Neuron* **84**, 347 (2014).
9. X. Liu *et al.*, *Nature* **484**, 410 (2012).
10. M. E. Bouton, *Learn. Mem.* **11**, 485 (2004).
11. J. Ji, S. Maren, *Hippocampus* **17**, 749 (2007).
12. K. M. Myers, M. Davis, *Mol. Psychiatry* **12**, 120 (2007).

10.1126/science.aau0035



## MACHINE LEARNING

# Understanding spatial environments from images

An algorithm that is trained to predict views of a spatial environment also infers its 3D structure

By **Matthias Zwicker**

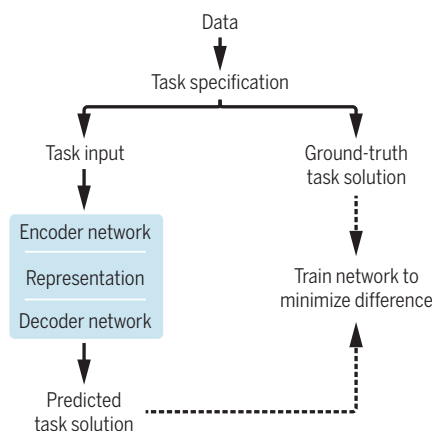
The ability to understand spatial environments based on visual perception arguably is a key function of the cognitive system of many animals, including mammals and others. A common presumption about artificial intelligence is that its goal is to build machines with a similar capacity of “understanding.” The research community in artificial intelligence, however, has settled on a more pragmatic approach. Instead of attempting to model or quantify understanding directly, the objective is to construct machines that merely solve tasks that seem to require understanding. Understanding can only be measured indirectly, for example, by analyzing the ability of a system to generalize the solving of new tasks, which is sometimes called transfer learning (1). Transfer learning is particularly appealing in an unsupervised setting, which means that the objective of the original task is defined in terms of the input data itself, without requiring additional, task-specific information (see the figure). On page 1204 of this issue, Eslami *et al.* (2) present an important step toward building machines that learn to understand spatial environments using unsupervised transfer learning. Remarkably, they develop a system that relies only on inputs from its own image sensors, and that learns autonomously and without human supervision.

Although the research community has made limited progress toward a theoretical analysis of unsupervised transfer learning, it has achieved some notable successes in empirical experiments. Loosely inspired by the structure of biological brains, the most successful techniques rely on artificial neural networks (3) as their computational structures. Like biological brains, the neurons in artificial neural networks (ANNs) receive multiple inputs from other neurons, perform a simple computation, and forward outputs to other neurons. Neuron connections in ANNs are fixed; they are trained by adjusting the computation of each neuron's output.

One approach (4) for unsupervised transfer learning using ANNs is to build on an intuitive network architecture consisting of an encoder network (called a “representation network” by Eslami *et al.*) that obtains visual inputs and encodes them in a compact representation. This step is followed by a decoder network (called a “generation network” by Eslami *et al.*) that interprets the representation to provide a solution for the task at hand. Transfer learning reuses

## Unsupervised representation and transfer learning

A user specifies a task and an automatic process to obtain task inputs and corresponding ground-truth solutions from a dataset. By designing suitable training tasks, the learned representation can be shown experimentally to “understand” the data, in the sense that it can be used solve new tasks.



the encoder and representation, potentially with a different decoder, to solve a new task. For example, unsupervised transfer learning using such an encoder-decoder approach has been used to solve visual puzzle problems (5), and the learned representation has shown to be effective at recognizing objects in images.

The approach of Eslami *et al.* takes this concept one step further by addressing the problem of understanding three-dimensional (3D) environments instead of 2D images. However, a key challenge in understanding

a 3D scene is that most sensory devices such as digital cameras cannot acquire 3D scenes directly. Instead, these devices only record 2D views. Eslami *et al.* circumvent this problem in an elegant manner by simulating, or rendering, 2D views of virtual 3D environments, which they themselves constructed using standard computer algorithms. Specifically, this approach allows them to produce 2D views from any desired viewpoint. They then train a neural network with an encoder-decoder architecture to predict new views of a 3D scene from two nearby views of the same scene, which are provided to the encoder as inputs. This step is the key idea that facilitates transfer learning, as Eslami *et al.* show in their experiments. In a training step, the neural network learns to solve this view prediction task on millions of virtual 3D scenes and their 2D views.

What is probably most fascinating about the study by Eslami *et al.* is that their system is not limited to predicting new views of 3D scenes, as it was asked to do during training. Indeed, they show that the internally encoded representation captures the structure of 3D scenes, for example, in terms of the number and types of objects contained in the scenes. Eslami *et al.* demonstrate this capability by performing scene algebra, which means that linear combinations of scene representation vectors, when decoded into images, correspond to semantically meaningful and predictable changes in scene structure. This representation allows them to replace a certain type or color of an object with another one. They also show how their learned representation can be used to control a virtual robot arm to navigate a simple 3D scene based on its 2D views acquired by a simulated camera.

The work of Eslami *et al.* does not come without limitations. Most important, their experiments are restricted to simple 3D scenes consisting of a few basic geometric objects. Hence, it remains unclear how close their approach could come to understanding complex, real-world environments, which would make it useful, for example, to implement practical control of robots. Nonetheless, their technique introduces a number of crucial contributions that will likely make this possible in the future. ■

## REFERENCES

1. S. J. Pan, Q. Yang, *IEEE Trans. Knowl. Data Eng.* **22**, 1345 (2010).
2. S. M. A. Eslami *et al.*, *Science* **360**, 1204 (2018).
3. Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**, 436 (2015).
4. Y. Bengio, *Proceedings of ICML Workshop on Unsupervised and Transfer Learning* (2012), vol. 27, pp. 17–36.
5. M. Noroozi, P. Favaro, in *European Conference on Computer Vision 2016*, Lecture Notes in Computer Science, B. Leibe, J. Matas, N. Sebe, M. Welling, Eds. (Springer, Cham, 2016), vol. 9910, pp. 69–84.

Department of Computer Science, University of Maryland, College Park, MD, USA. Email: zwicker@cs.umd.edu

## POLICY FORUM

## OCEAN POLICY

# Preparing ocean governance for species on the move

Policy must anticipate conflict over geographic shifts

By **Malin L. Pinsky**<sup>1</sup>, **Gabriel Reygondeau**<sup>2</sup>, **Richard Caddell**<sup>3,4</sup>, **Juliano Palacios-Abrantes**<sup>2</sup>, **Jessica Spijkers**<sup>5,6</sup>, **William W. L. Cheung**<sup>2</sup>

**T**he ocean is a critical source of nutrition for billions of people, with potential to yield further food, profits, and employment in the future (1). But fisheries face a serious new challenge as climate change drives the ocean to conditions not experienced historically. Local, national, regional, and international fisheries are substantially underprepared for geographic shifts in marine animals driven by climate change over the coming decades. Fish and other animals have already shifted into new territory at a rate averaging 70 km per decade (2), and these shifts are expected to continue or accelerate (3). We show here that many species will likely shift across national and other political boundaries in the coming decades, creating the potential for conflict over newly shared resources.

A shifting fish stock exacerbates existing fisheries challenges because it contravenes the “clear boundaries” principle for sustainable governance of common pool resources, eroding incentives for conservation when new free riders, having no agreed-upon responsibilities for shared conservation and management, gain access to a resource (4, 5). Stock shifts can incentivize regional overharvesting as actors scramble to exploit a perceived disappearing resource. A stock that upon moving straddles national boundaries may find itself in “double jeopardy,” exposed to

unsustainable competitive harvesting (5). Governance challenges posed by shifting marine animal distributions have been recognized in certain cases, but the scope and magnitude of this problem have remained unclear, and there have been few efforts to address the issues.

## SHIFTING FISHERIES DRIVE CONFLICTS

International law recognizes that cooperation is necessary for management of shared stocks, yet fisheries disputes remain commonplace and are a leading cause of militarized disputes between democratic states in the post-World War II period (6). The so-called “mackerel war” erupted in 2007 when the northeast Atlantic mackerel stock (*Scomber scombrus*)—a fishery then managed by the European Union, Norway, and

***“...widespread noncooperative management... risks...fractured international relationships, and political conflicts that could spill over into other, nonfishery areas of international politics.”***

Faroe Islands—shifted into Iceland’s Exclusive Economic Zone (EEZ) (7). Conflict arose over appropriate allocations among the actors, compounded by disagreement about the drivers and therefore the expected duration of the shift. In the absence of cooperation, the mackerel stock became increasingly overfished (7).

Shifting species have caused conflict even between countries that historically cooperate closely. During a period of warmer-than-average regional ocean temperatures in the 1980s and 1990s, U.S. catches of Pacific salmon (*Oncorhynchus* spp.) increased more than 10-fold and included increased interceptions of Canadian-bound salmon (5). Canadian fisheries retaliated by targeting salmon migrating to spawn in the United States. Six years of rancorous disagreement passed before a new joint management agreement was concluded.

Shifting species distributions also present internal challenges for nations. In the United States, Blueline tilefish (*Caulonotus microps*) were historically caught and managed south of the Virginia–North Carolina border. When tilefish appeared farther north, a fishery exploited the stock for nearly a decade without regulation. This situation only changed in 2015 with emergency rules from the National Marine Fisheries Service.

These cases exemplify a general pattern: Existing fisheries management and governance are largely predicated on population geographies that remain broadly static through time. Challenges emerge when stock distributions become less predictable and are compounded when states act unilaterally to exploit the resultant windfall.

## MAGNITUDE OF FUTURE CHALLENGES

The oceans have already absorbed 93% of the heat from anthropogenic climate change (8), and if future species geographic shifts exceed historical variation, adjustment to existing ocean governance will be needed. Alternatively, future geographic shifts could be sufficiently limited to retain stocks primarily under the jurisdiction of those countries currently managing them. The extent to which future shifts in species distributions will generate newly shared fish stocks and increase the potential for conflict, however, has not been clear.

We therefore projected future shifts in the distribution of 892 commercially important marine fish and invertebrates in relation to 261 of the world’s EEZs (see supplementary materials). Instead of precisely forecasting future changes, the projections help delineate plausible scenarios that illustrate the extent of future challenges. Comparing 1950–2014 with 2090–2100, we found that many of the world’s EEZs are likely to receive one to five new, climate-driven transboundary stocks by the end of the century (see the first figure). Up to 10 new stocks were projected for some EEZs in east Asia, a region where new transboundary stocks could worsen maritime relations already complicated by disputed territories, overlapping EEZ claims, and illegal fishing.

The number of EEZs with new transboundary stocks was expected to reach  $46 \pm 8$  ( $\pm$ standard error) or  $60 \pm 4$  by 2060 ( $57 \pm 4$  or  $85 \pm 22$  by 2080) under strong mitigation [representative concentration pathway (RCP) 2.6] or business-as-usual (RCP 8.5) greenhouse gas emissions scenarios, respectively (see the second figure). Limit-

<sup>1</sup>Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, NJ, USA. <sup>2</sup>Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, BC, Canada. <sup>3</sup>Netherlands Institute for the Law of the Sea, Utrecht University, Utrecht, Netherlands. <sup>4</sup>School of Law and Politics, Cardiff University, Cardiff, UK. <sup>5</sup>Stockholm Resilience Center, Stockholm University, Stockholm, Sweden. <sup>6</sup>Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Queensland, Australia. Email: malin.pinsky@rutgers.edu



ing greenhouse gas emissions would therefore reduce the potential for new fisheries conflicts. In total, new transboundary stocks were projected to be present in 23% (RCP 2.6) to 35% (RCP 8.5) of global EEZs by 2100 (see the figures). In the tropics, fisheries will likely move out but not in, a process that creates additional food security concerns.

Most countries were projected to receive 1 to 30% of their potential fisheries catch from new stocks by 2100, but percentages were higher in temperate regions (e.g., Australia or countries around the Baltic and Bering seas) and highest in shared Antarctic fishing grounds (92%). We note that past conflicts over even a single species with low catch volumes have been substantial.

the fisheries regime of UNCLOS was buttressed by the UN Fish Stocks Agreement (UNFSA), which specifically applies to straddling and highly migratory stocks and entered into force in 2001. The UNFSA reinforced national obligations to cooperate and to apply a precautionary approach to fisheries. Notwithstanding its constructive influence on international fisheries law, the UNFSA has not focused attention upon stocks that shift to occupy territory beyond the areas they have occupied historically.

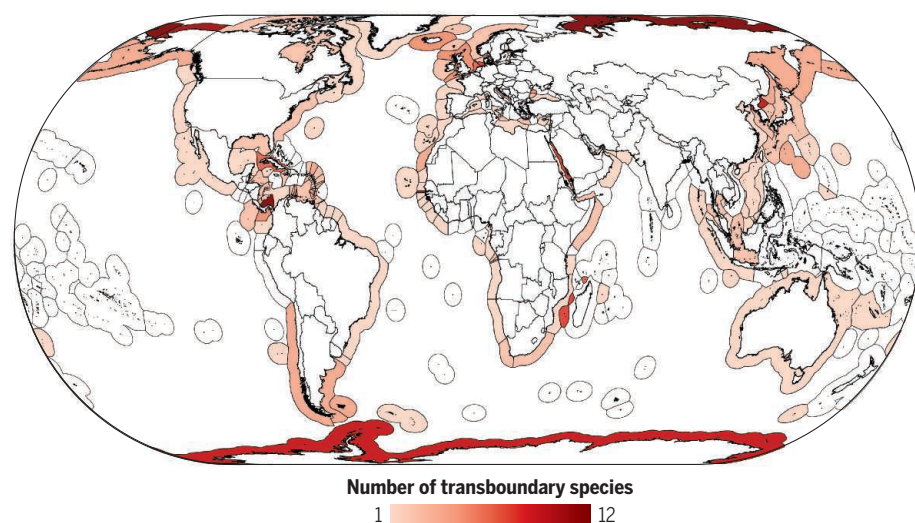
Regional fisheries management organizations (RFMOs) remain the primary vehicle through which fish stocks that straddle multiple EEZs are managed. Many RFMOs address single species such

impacts on species not directly managed by the RFMOs.

Attempts to resolve conflicts judicially are largely untested, although shifting stocks could prompt judicial consideration in the future (legal processes in the mackerel dispute were discontinued). International courts and tribunals have been receptive to calls for more responsible stewardship of fish stocks and have adopted far-sighted allocation practices in individual cases. Nevertheless, they have historically accorded little consideration to environmental factors (climate or otherwise) in territorial disputes, and shifting stocks have not played an overt role in boundary decisions. Likewise, international law does not facilitate the adjustment of national jurisdictional boundaries in response to changing ocean conditions. Finally, judicial decisions do not always resolve conflicts: China refused to participate in recent arbitration concerning the South China Sea, rendering fisheries relations vulnerable to unilateral actions around stock shifts in this region.

## Exclusive Economic Zones projected to contain one or more new fishery stocks by 2100, compared to the distribution in 1950–2014

Projections represent an ensemble average across three Earth-system models under the high-greenhouse gas emissions scenario (RCP 8.5). See supplementary materials.



### MAJORS GAPS IN GOVERNANCE

The current legal framework for the international regulation of fisheries does not directly account for fluctuating or changing distributions. The primary source of international obligations for the governance of global fisheries resources remains the 1982 United Nations Convention on the Law of the Sea (UNCLOS), which entered into force in 1994. Under UNCLOS, states must ensure that fisheries in their EEZs are not endangered by overexploitation; hence, national regulations for fishing could provide a basis for far-sighted management of shifting stocks.

For “straddling stocks” occurring in two or more EEZs, or within an EEZ and the high seas, UNCLOS obliges states to cooperate to establish necessary conservation and management measures. In 1995,

as tuna or salmon, however, and an influx of additional species lies beyond their individual remits. Despite recent progress, fish stocks in large parts of the global oceans are weakly managed—a trend that may be exacerbated by shifting distributions. Few bodies have established a clear position on the elaboration of regulations for new fisheries, a loophole that often allows newly fished stocks to be heavily exploited before meaningful standards are developed (9). Moreover, there has been little to no cooperation between RFMOs on the potential for future shared stocks, and limited interactions with other regional ocean regulators or global treaties. Concerns also remain over the limited application of ecosystem-based management principles by RFMOs, including limited consideration of

### GOVERNANCE SOLUTIONS FOR SHIFTING FISH

Past conflicts, the projected widespread emergence of new transboundary stocks, and the gaps in current governance frameworks all suggest that substantial new approaches are needed to forestall future conflict. The first step is for management authorities to plan ahead for cooperative management, which demands an emphasis on acquiring reliable projections of species shifts and associated uncertainties. Negotiations over shared stocks are easier with mutually agreed facts, which can be facilitated by data from multilateral or independent scientific bodies, including the Intergovernmental Panel on Climate Change (IPCC). All projections should be interpreted cautiously, however, given the potential for thresholds and surprises in ecological systems. These inherent uncertainties complicate localized evaluations of the costs and benefits of cooperation (10, 11).

For RFMOs, performance reviews provide an established process for consideration of species shifts, although success depends on capacity and a culture of critical reflection. Data sharing with other bodies is also vital. The Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) has been exemplary in this regard and has established collaborative arrangements with neighboring RFMOs to monitor the movement of stocks across regulatory frontiers. CCAMLR has also forged arrangements with other sectoral regulators to consider the prospective ecological footprint of a moving fishing industry.

Cooperation must then extend beyond data sharing to inform genuinely collaborative management. For example, to meet their responsibilities under UNCLOS and UNFSA, RFMOs must accept the prospect of shared oversight and agree on regulatory responsibilities for species with an increasing presence in neighboring areas. Overlapping stocks have already generated tensions, such as those between the Northwest Atlantic Fisheries Organization (NAFO) and the North East Atlantic Fisheries Commission (NEAFC) in the 1990s until shared responsibilities were implemented. CCAMLR has taken a different and more constructive approach to cooperation, in part by imposing greater precautionary

the case of the U.S.-Canada Pacific Salmon Treaty, however, contributions to a conservation fund helped stabilize relations, creating an alternative avenue for compensation often termed a "side payment" (13). Similar approaches are illustrated by Norway and Russia's swaps of fisheries access within EEZs to balance shifts in shared stocks, an important example of flexibility in comanaging Arctic resources. Trading herring, blue whiting, or other fishery access to help resolve the Icelandic mackerel dispute has also been suggested (14). The utility of side payments suggests that new bilateral or multilateral agreements concerning shifting fisheries will be more effective if negotiated at political levels above simply fisheries management. Presently, however, multilateral processes generally focus on discrete issues to help secure support. For instance, it is being actively debated whether fisheries should be included in United Nations negotiations on marine biodiversity beyond national jurisdiction (BBNJ), despite this process expressly seeking integration across governance sectors.

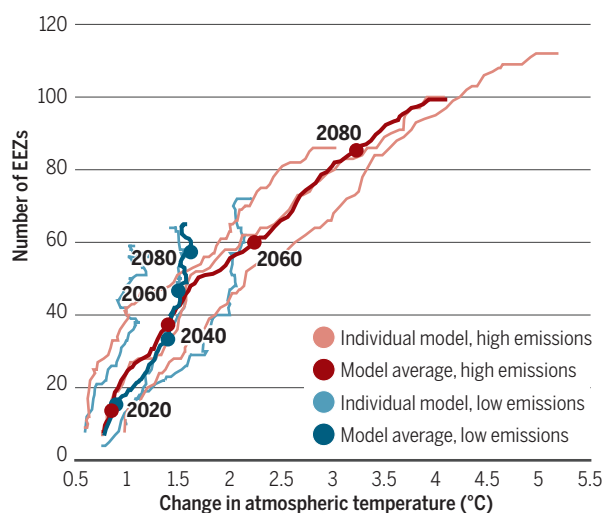
Compounding this proprietorial approach are concerns that access to current and prospective RFMOs is restricted to those with a "real interest" in the stock, with participatory rights zealously guarded by current constituents (14, 15). The North Atlantic RFMOs—which are facing geographic shifts in a number of important fisheries—are currently closed to new

consideration in the boundary delimitation process or bespoke arrangements between states over contested resources.

Climate-driven shifts in marine species distributions represent a growing governance issue affecting states in all regions of the world. An alternative future of widespread noncooperative management over new transboundary stocks risks extensive overfishing, decline in global food and livelihood provisioning from the ocean, fractured international relationships, and political conflicts that could spill over into other, nonfishery areas of international politics. However, the challenges can also be mitigated through far-sighted governance strategies. With adaptable agreements between states, we have hope that ocean fisheries can continue to provide the myriad nutritional, livelihood, and economic opportunities relied upon by billions of people around the world. ■

## The number of EEZs with new transboundary stocks increases with global temperature

The extent of warming and number of EEZs were greater under a high-greenhouse gas emissions scenario (RCP 8.5, red) and lower under a low-emissions scenario (RCP 2.6, blue). See supplementary materials.



oversight of new fisheries, including for tuna displaced by ocean warming (9). Data sharing and collaborative management can inform vital regulatory approaches such as area-based management and no-take zones to reduce pressure on shifting stocks.

Prevailing management mentalities also remain a fundamental challenge, notably the perception that one party "wins" and the other "loses" when a stock shifts geographically, an asymmetry that can undermine cooperation (12). Game theory provides lessons for incentivizing cooperation, including broadening the scope of negotiations to include non-fish resources (5). Broader negotiations, however, risk reducing fisheries to a bargaining chip to be leveraged against other political and economic priorities, as suggested by the Brexit negotiations. In

members (9, 15). Many existing fisheries are based on principles of zonal attachment and relative stability, wherein national allocations of fish catch are based upon historical patterns of presence in each country and geographic area. A first step toward more adaptable fisheries would be objective and regularly updated allocations of catch or effort to reflect changes in stock distributions. An intriguing, alternative approach would be to develop fisheries permits that are tradable across political boundaries, as considered to some extent by the International Commission for the Conservation of Atlantic Tunas (ICCAT) and by the Northwest Atlantic Fisheries Organization (NAFO) (15). Regions with disputed maritime boundaries will remain especially prone to conflict, and shifting stocks may require additional

## REFERENCES AND NOTES

1. C. J. Costello et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5125 (2016).
2. E. S. Poloczanska et al., *Nat. Clim. Change* **3**, 919 (2013).
3. W. W. L. Cheung, G. Reygondeau, T. L. Frölicher, *Science* **354**, 1591 (2016).
4. E. Ostrom, J. Burger, C. B. Field, R. B. Norgaard, D. Policansky, *Science* **284**, 278 (1999).
5. K. A. Miller, G. R. Munro, U. R. Sumaila, W. W. L. Cheung, *Can. J. Agr. Econ.* **61**, 309 (2013).
6. S. M. Mitchell, B. C. Prins, *Int. Stud. Q.* **43**, 169 (1999).
7. J. Spijkers, W. J. Boonstra, *Reg. Environ. Change* **17**, 1835 (2017).
8. M. Rhein et al., in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker et al., Eds. (Cambridge Univ. Press, Cambridge, 2013), pp. 255–316.
9. R. Caddell, *Int. J. Mar. Coast. Law* **33**, 199 (2018).
10. K. Miller et al., *Prog. Oceanogr.* **87**, 338 (2010).
11. S. Polasky, S. R. Carpenter, C. Folke, B. Keeler, *Trends Ecol. Evol.* **26**, 398 (2011).
12. G. R. Munro, *Can. J. Econ.* **12**, 355 (1979).
13. K. A. Miller, G. R. Munro, *Mar. Resour. Econ.* **19**, 367 (2004).
14. P. Ørebech, *Int. J. Mar. Coast. Law* **28**, 343 (2013).
15. A. Serdy, *The New Entrants Problem in International Law* (Cambridge Univ. Press, Cambridge, UK, 2015).

## ACKNOWLEDGMENTS

We thank C. Payne and K. Miller for helpful comments and discussions. Funding: This is a contribution to the Nippon Foundation–University of British Columbia Nereus Program, and we acknowledge support from it. We also acknowledge support from National Science Foundation OCE-1426891 and DEB-1616821, New Jersey Sea Grant R/6410-0011, an Alfred P. Sloan Research Fellowship, the National Oceanic and Atmospheric Administration's Coastal and Ocean Climate Applications program, and the project "Green Growth Based on Marine Resources: Ecological and Socio-Economic Constraints (GreenMAR)" funded by Nordforsk no. 61582. Author contributions: Conceptualization, M.L.P.; Data curation, G.R., J.P.-A.; Formal analysis, G.R., J.P.-A.; Funding acquisition, W.W.L.C., M.L.P.; Methodology, G.R., M.L.P., W.W.L.C.; Supervision, M.L.P., W.W.L.C.; Visualization, G.R.; Writing – original draft, M.L.P., J.S., W.W.L.C., G.R., R.C.; Writing – review & editing, M.L.P., J.S., W.W.L.C., G.R., R.C. Competing interests: None declared. Data and materials availability: See supplementary materials.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6394/1189/suppl/DC1



BOOKS *et al.*

## HISTORY OF SCIENCE

# Asperger's chilling complicity

The physician behind the eponymous syndrome was an active participant in Nazi eugenics

By **Meredith Wadman**

**O**n 1 July 1941, a young Austrian physician named Hans Asperger signed a document transferring a toddler named Herta Schreiber to Spiegelgrund, an asylum for mentally ill children on the outskirts of Vienna. Two-year-old Herta had suffered diphtheria and meningitis, leaving her severely disabled. She “must present an unbearable burden to her mother,” Asperger, then the director of the Curative Education Clinic at the University Children’s Hospital in Vienna, wrote on the transfer document. Herta’s placement at the asylum, called Spiegelgrund, was “absolutely necessary,” he directed.

A haunting photograph of the crying toddler, her head shaven, taken soon after her arrival at Spiegelgrund, is all that survives of Herta. The facility’s murderous medical director, Erwin Jekelius, a former colleague of Asperger’s at the university clinic, soon sought Reich authorization to kill the girl as

part of the Nazi drive to rid the gene pool of undesirables. She died two months after her admission to Spiegelgrund, where the Nazis would kill nearly 800 children between 1940 and 1945.

Unlike the millions who were gassed, the killings of the Spiegelgrund children were prolonged and intimate, Edith Sheffer explains in her searing new book, *Asperger’s Children*. “Doctors personally examined the children they condemned. Nurses personally fed and changed [their] sheets.... Death came slowly, painfully, as children would be starved or given overdoses of barbiturates until they grew ill and died, usually of pneumonia.”

The man who sent Herta to her death is the same Asperger who is famous today for the eponymous syndrome, which was coined in 1981 by British psychiatrist Lorna Wing and by the mid-1990s had made its way into common usage. Wing resurrected Asperger’s description of a broad condition he called “autistic psychopathy” in his obscure 1944 postdoctoral thesis. It expanded the classical conception of autism to describe a broader class of behaviors. The

“high functioning” end of this spectrum became known as Asperger’s syndrome.

With the growing prominence of the syndrome came increased scrutiny of the man behind it. But until recently, historians writing in English have either skirted Asperger’s closeness to the Nazi regime or depicted him as a compassionate physician

who used his medical position to rescue disabled children otherwise bound for destruction. The 2016 book *In a Different Key: The Story of Autism*, by John Donvan and Caren Zucker, began to dismantle this portrait of Asperger. Now Sheffer, a senior fellow at the Institute of European Studies at the University of California, Berkeley, finishes the job.

Her meticulously researched yet readable account shines a dispassionate light on Asperger as a man actively complicit with Nazi eugenicists carrying out Hitler’s child “euthanasia” program. An important, complementary article by Herwig Czech at the Medical University of Vienna was recently published in the open-access journal *Molecular Autism* (1).



**Asperger's Children**  
The Origins of Autism  
in Nazi Vienna  
Edith Sheffer  
Norton, 2018. 317 pp.

The reviewer is on staff at Science magazine and is the author of *The Vaccine Race: Science, Politics, and the Human Costs of Defeating Disease* (Viking, New York, 2017). Email: mwadman@aaas.org

Seven hundred and seventy-two lighted pillars memorialize the children killed at Spiegelgrund.

A devout Catholic, Asperger never joined the Nazi party. But he joined several far-right, anti-Semitic groups in the mid-1930s. In 1934, at 28, he was promoted over the head of a senior Jewish colleague to direct the Curative Education Clinic, a child psychiatry clinic at the Vienna children's hospital. His patron there was the hospital director, Franz Hamburger, a Nazi ideologue who oversaw heinous medical experiments on children and advocated letting those with "poor constitutions" die. Asperger adulated Hamburger and credited his mentor's influence on his thinking as late as 1977.

In 1937, Asperger wrote that it is "impossible" to establish rigid criteria for diagnosing autism. But in 1938, soon after Germany annexed Austria, he did precisely that, defining "autistic psychopaths" as a "well-characterized group." By 1944, he was describing the most impaired children as dehumanized, "intelligent automata." The autistic child was "like an alien"; "not an active member of the greater organism"; and at risk of growing up to "roam the streets ... grotesque and dilapidated." Being thus diagnosed put children in grave peril. The Third Reich was a "diagnosis regime," Sheffer notes, with death as a treatment option.

Although Asperger offered intensive, individualized care to children he thought could be socially integrated, Sheffer writes that he also sent, or helped send, dozens of children to Spiegelgrund in full knowledge of the murders transpiring there. As well as dispatching children from his own clinic, Asperger publicly urged Viennese colleagues to transfer "difficult cases" to the asylum. And in 1942, he served on a seven-member city commission that evaluated 210 children from another psychiatric institution and dismissed 35 as hopeless cases, incapable of development or education. The commission sent them to Spiegelgrund, ordering "Jekelius action." All were killed.

*Asperger's Children* paints a dark portrait of Asperger's creeping complicity, as his patient care and his scholarship became inextricably woven into the fabric of the Nazi child extermination campaign. Asperger tried in later life to distance himself from that enterprise, painting himself as a Nazi resister who endangered himself by protecting children. This book should lay that saintly image to rest. ■

#### REFERENCE

1. H. Czech, *Mol. Autism* **9**, 29 (2018).

10.1126/science.aat9801

## COMPUTER SCIENCE

# Platforms of power

A Microsoft researcher confronts how companies shape what we see and say online

By Anna Lauren Hoffmann

**O**n 1 March 2018, Twitter announced a new initiative aimed at measuring and evaluating the platform's "conversational health," that is, how well it promotes lively conversation and critical thinking while also minimizing the social impact of abuse, spam, and manipulation. The initiative is one piece of a larger effort by social media companies both to respond to the proliferation of disinformation and propaganda exposed in the aftermath of the 2016 U.S. presidential election and to attend to ongoing issues of sexist, racist, and homophobic abuse and harassment.

In framing the initiative as a matter of "health," the company put a therapeutic spin on much longer-standing practices for evaluating, filtering, and shaping what we can see and say online. As Tarleton Gillespie captures it in his new book, *Custodians of the Internet*, "[t]he concerns around political discourse and manipulation on social media platforms feel like the crest of a larger wave that has been breaking for a few years now, a broader reconsideration, on so many fronts, of social media platforms and their power in society."

In the book, Gillespie, a principal researcher at Microsoft Research New England and an adjunct associate professor at Cornell University, goes beyond a mere account of the tools and practices employed by social media companies to address problems of harassment, obscenity, and hate speech on their platforms. He also aims to capture just what is at stake in debates over online expression—from the consequences of being "censored" online to the fate of social and democratic norms in the face of easily manipulable news feeds.

Although these questions have taken on a new urgency, they are not wholly new. As the opening chapters of *Custodians* make clear, content moderation is endemic to exchanging information over the internet.

There has never been a time when policy, social context, or technical realities have not informed the possibilities and limits of online communication.

Despite this long history, the work of moderation has become more pointed, and more challenging, in the 21st century. The explosion of platforms for expression—from message boards and chat rooms to large-scale social networking sites—and the increasing pervasiveness of those platforms in social, political, and economic life demand that we "reconsider what platforms are, and ask new questions about

their power in society." To this end, more than half the book is dedicated to demystifying platforms and their politics, making visible the human labor, internal policies, and automated tools that make content moderation at the scale of millions (or, in some cases, billions) of users possible.

In our current geopolitical moment, it would be easy to criticize parts of Gillespie's account as frustratingly apolitical. There are times when the reader might reasonably expect him to be more cynical about the relationship between content moderation, the collection of massive amounts of user data, and platforms' business models. As he himself notes, the mechanisms of moderation "cannot be disentangled from the economic imperatives that drive moderation too."

But what Gillespie sacrifices in critical detours early on, he gains in expository



**Custodians of the Internet**  
Tarleton Gillespie  
Yale University Press,  
2018. 296 pp.



Internet platforms work hard to maintain the illusion of being neutral facilitators of content.

The reviewer is at the Information School, University of Washington, Seattle, WA 98195, USA. Email: alho@uw.edu



efficiency. He moves quickly between topics that merit entire volumes themselves, from cataloging the social and technical means by which platforms moderate to capturing the lived experience of users under sometimes hostile systems. (The struggle of mothers attempting to share breastfeeding photos on Facebook, only to be thwarted by overly prudish or actively hostile editorial guidelines, is one example of the latter and features prominently in the middle of the book.)

These criticisms, however, are matters of degree, not category. As with the filtering mechanisms described in the book, readers from different backgrounds might calibrate the relative importance of various factors differently. Some might prefer to foreground the economics of content moderation while downplaying concerns over censorship and freedom of expression. Others might disfavor talk of algorithms and automation and instead emphasize the efforts of human laborers—the severely underpaid and overtaxed “clickworkers” who daily must wade through the internet’s most toxic content. That the book is able to fairly account for these myriad and

far-ranging dimensions of moderation is a credit to Gillespie’s detailed work.

For the reader looking for a more trenchant analysis, Gillespie’s final two chapters move in this direction. As Gillespie makes clear, online platforms work hard to keep their moderation practices invisible. It’s better for them if users (and, perhaps more pointedly, regulators) remain largely unaware of the constant negotiations and often fraught decisions they must make behind the scenes to maintain an image as neutral facilitators of content rather than being seen as a new kind of editorial power.

That moderation has, despite companies’ best efforts, found its way into broader public and regulatory conversations is, as Gillespie suggests, evidence that we’re seeing “social media’s slow and bumpy maturation, its gathering recognition that it is a powerful infrastructure for knowledge, participation, and public expression.” In view of this, he asks us to imagine a world in which social media platforms not only claim the right to moderate but also fully absorb “responsibility for their role in organizing, curating, and profiting from the activity of their users.”

Even if unrealistic, it’s an earned conclusion. It resonates not despite but precisely because it is backed by decades of online history, scholarship, and industry development. In that regard, the book is a notable departure from other entries in the genre, such as 2011’s *The Filter Bubble* or 2001’s *Republic 2.0*, that have been, at times, heavy on political prognostication but light on social scientific substance.

For scholars, activists, and practitioners already steeped in the politics of online platforms, *Custodians of the Internet* offers both a comprehensive retrospective and critical provocation. The book effectively documents where we’ve been, where we are now, and where—barring the sorts of changes Gillespie suggests—we’re headed.

But this book isn’t only for those already steeped in the practices of content moderation; it’s also for audiences who have only recently become aware of how deeply filtering, policy, and human judgment shape everyday life online. For newcomers, Gillespie offers an accessible and wide-ranging introduction to the dirty work of cleaning up the internet. ■

10.1126/science.aat8179

## LETTERS

Edited by Jennifer Sills

## Conservation accord: Let countries govern

In their Perspective "How to pay for saving biodiversity" (4 May, p. 486), E. B. Barbier *et al.* propose a global agreement for biodiversity conservation in which both countries and corporations would participate. We argue that corporate involvement is unrealistic. Only countries should serve as formal parties to the conservation agreement.

Each country joining the proposed conservation agreement would use national conservation data to create targets, policies, and regulations that rigorously standardize the activities of corporations as well as the government (1–4). This way, corporations would take part in the biodiversity conservation, regardless of whether they sign the agreement. If corporations independently join the agreement, they could pursue their own targets, policies, and timelines, which may differ from the national standard. Two conflicting standards would confuse the participants, complicate enforcement, and undermine the goal of saving biodiversity.

Corporate involvement may also be unrealistic in terms of logistics. There are vastly more corporations and organizations than there are countries (5). The manpower, financial support, and material resources required to add a large number of diverse corporations and organizations to the conservation agreement would not be economical. Therefore, countries should take the lead in governing biodiversity conservation.

**Ming Chen,<sup>1,2</sup> Yingzhu Sun,<sup>1,2</sup> Chunping Yang,<sup>1,2</sup> Guangming Zeng,<sup>1,2\*</sup> Zhongwu Li,<sup>1,2</sup> Yi Zhu,<sup>3</sup> Jiachao Zhang<sup>4</sup>**

<sup>1</sup>College of Environmental Science and Engineering, Hunan University, Changsha 410082, China. <sup>2</sup>Key Laboratory of Environmental Biology and Pollution Control, Ministry of Education, Hunan University, Changsha 410082, China. <sup>3</sup>School of Environmental Science and Engineering, Hubei Polytechnic University, Huangshi 435003, China. <sup>4</sup>College of Resources and Environment, Hunan Agricultural University, Changsha 410128, China.

\*Corresponding author. Email: zgming@hnu.edu.cn

## REFERENCES

1. E. D. Minin *et al.*, *Biol. Conserv.* **206**, 56 (2017).
2. H. M. Pereira *et al.*, *Science* **339**, 277 (2013).
3. F. T. Wetzel *et al.*, *Biodiversity* **16**, 137 (2015).
4. S. H. M. Butchart *et al.*, *Conserv. Lett.* **8**, 329 (2015).
5. The World Bank, "Listed domestic companies, total" (2017); <https://data.worldbank.org/indicator/CM.MKT.LDOM.NO>.

10.1126/science.aat7724



## Conservation accord: Cash is not enough

In their Perspective "How to pay for saving biodiversity" (4 May, p. 486), E. B. Barbier *et al.* call for a global biodiversity agreement and substantially increased corporate financing to conserve 50% of the global land area. However, additional financing will achieve little without first resolving the divergent perspectives of the various societal actors in conservation.

Barbier *et al.* refer to the Paris Climate Change Agreement, but the success of that accord reflects the near-universal recognition of the urgent need to respond to climate change. This is not yet the case for conservation. Cajoling society toward common conservation values and goals requires societal engagement rather than arbitrary global targets. Funds without such societal commitment and governance will simply be lost to corruption or rendered ineffective.

Barbier *et al.* also call for the agriculture industry to contribute financially to conservation in recognition of the estimated billions received from environmental services such as pollination. However, only a third of global agriculture area depends on pollinators (1). Moreover, the agriculture industry includes large transnational companies as well as millions of small farmers. The latter group depends most on ecosystem services (2, 3), yet has the least capacity to contribute financially to global conservation. Locally

orientated conservation action, driven by locally recognized conservation values, is a more credible approach that needs less funding and more engagement and capacity development.

Finally, massive allocation of funding to conserve extensive swathes of land in largely tropical countries will be subject to neocolonial interpretation and national sovereignty concerns, for which there are many unhappy precedents.

### Jaboury Ghazoul

Department of Environmental Systems Science, ETH Zurich, 8092 Zurich, Switzerland and Prince Bernhard Chair for International Nature Conservation, Department of Biology, Utrecht University, 3584 CH Utrecht, Netherlands. Email: jaboury.ghazoul@env.ethz.ch

## REFERENCES

1. M. A. Aizen *et al.*, *Curr. Biol.* **18**, 1572 (2008).
2. Millennium Ecosystem Assessment, "Ecosystems and human well-being: Synthesis" (Island Press, Washington, DC, 2005).
3. R. Vignola *et al.*, *Agric. Ecosyst. Environ.* **211**, 126 (2015).

10.1126/science.aau0798

## Conservation accord: Corporate incentives

In their Perspective "How to pay for saving biodiversity" (4 May, p. 486), E. B. Barbier and colleagues suggest that corporations should support global biodiversity conservation. They propose an international policy, similar to the Paris Climate Change Agreement, alongside an objective of conserving 50% of all habitats ["Half Earth"



(1)]. They suggest that corporations that benefit directly from increased biodiversity could buy into this agreement and help finance conservation efforts. We agree that corporations can play a larger role in conserving biodiversity. However, direct benefit is not the best incentive for corporations, and the Half Earth concept is not the best goal.

Simply suggesting that corporations finance conservation, in part because certain sectors stand to gain directly, is dangerous. Even if a sector benefitted overall, buy-in would be substantially eroded wherever this did not visibly translate into benefits for individual corporations, or more critically operational units within corporations. This argument could also marginalize action on biodiversity as a corporate social responsibility initiative, thereby diverting corporations' attention from addressing their environmental impacts and comprehensively managing their biodiversity risks (2, 3).

Corporations will be motivated to take effective, large-scale action only when biodiversity loss is perceived as a material risk to operations (including financial losses and damage to reputation) (4), not just as the opportunity for direct financial gain as Barbier and colleagues suggest. This will require stronger external market forces, such as environmental regulation, financial incentives, and public pressure, and the translation of science-based approaches to support businesses in measuring and evaluating business operations' impacts on biodiversity (3–5).

Barbier *et al.* also fail to acknowledge scientific criticism of the Half Earth concept (1, 6), which could make it a hard sell to corporations attempting to manage risk. More appealing might be a "no net loss or better" objective (7), integrating both existing international biodiversity targets (8) and those already adopted by leading corporations (9).

**Prue F. E. Addison<sup>1\*</sup> and Joseph W. Bull<sup>2</sup>**

<sup>1</sup>Department of Zoology, University of Oxford, Oxford, OX2 6GG, UK. <sup>2</sup>Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Canterbury, Kent, CT2 7NR, UK.

\*Corresponding author.

Email: prue.addison@zoo.ox.ac.uk

## REFERENCES

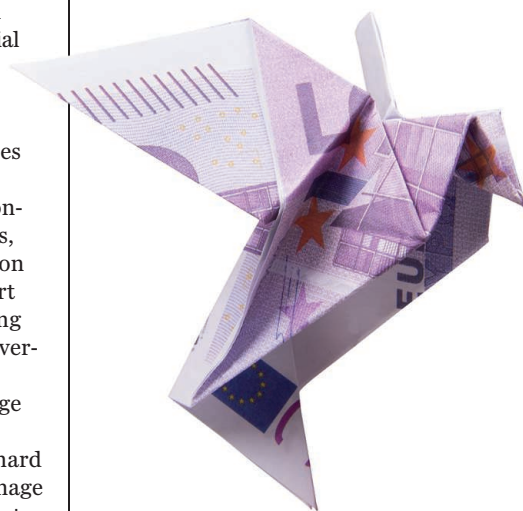
1. J. E. M. Watson, O. Venter, *Nature* **550**, 48 (2017).
2. K. H. Redford *et al.*, *Front. Ecol. Evol.* **3**, 1 (2015).
3. Cambridge Conservation Initiative, "Biodiversity at the heart of accounting for natural capital: The key to credibility" (Cambridge Conservation Initiative, Cambridge, UK, 2016); <https://naturalcapitalcoalition.org/biodiversity-at-the-heart-of-accounting-for-natural-capital/>.
4. J. Dempsey, *Geoforum* **45**, 41 (2013).
5. C. J. Vorösmarty *et al.*, *Science* **359**, 523 (2018).
6. B. Büscher *et al.*, *Oryx* **51**, 407 (2017).
7. W. N. S. Arlidge *et al.*, *Bioscience* **68**, 336 (2018).

8. Convention on Biological Diversity, "Strategic plan for biodiversity 2011–2020, including Aichi Biodiversity Targets" (CBD, 2011); [www.cbd.int/sp/default.shtml](http://www.cbd.int/sp/default.shtml).
9. H. J. Rainey *et al.*, *Oryx* **49**, 232 (2015).

10.1126/science.aau0788

## Response

Chen *et al.* argue against allowing corporations to participate formally in a Global Agreement of Biodiversity (GAB). Instead, they suggest an agreement in which only countries sign and enforce. Unfortunately, there is little evidence to indicate that current global agreements and financing are adequate to prevent continued decline in global biodiversity and critical habitats (1–3). It is unlikely that governments on their own will close the chronic funding gap in time to avert a global biodiversity crisis. Companies, especially those in key sectors, such as seafood, forestry, agriculture, and insurance, also have a financial stake in averting the global biodiversity crisis. These sectors could support the



GAB conservation targets as well as provide financial and technological assistance for conservation in developing countries.

Such a GAB would engage government and industry, and hopefully other nonstate actors. The evolution of the Paris Climate Agreement suggests that the strategy could succeed. Already some corporations, local governments, and other nonstate entities have announced voluntary pledges and low-carbon strategies to comply with the Paris Agreement (4), even though the private sector is not a formal participant and corporations do not contribute to the accord's climate financing. A recent proposal (4) advocates adding a mechanism to the Paris Agreement by which corporations, cities, and other nonstate actors could formally join the accord.

Ghazoul argues that the Paris Agreement is not a good model, because the global consensus and prioritization of climate change is more ingrained than conservation. He argues that local conservation efforts will be more effective. We agree that local and decentralized approaches to biodiversity conservation should be recognized and supported by any GAB and are fundamentally important for its success. However, we disagree that local conservation efforts will provide sufficient scale of conservation or resolve the chronic funding gap for global biodiversity conservation. As with any public good, biodiversity conservation suffers from a free-riding problem, in which governments have an incentive to provide less than the optimal level of funding in the hope that others will cover the costs. In particular, global funding to support conservation efforts in developing countries, which host most biodiversity, is woefully inadequate to prevent habitat loss and overexploitation. We maintain that the private sector can play a central role in addressing these problems. As we discuss in our Perspective, some corporations in natural resource-based sectors are already taking concrete actions.

Ghazoul also questions whether agricultural companies will have the incentive to contribute to a GAB, based on the benefits they receive from pollination services. We believe that this incentive does exist for agriculture and that the industry should be willing to contribute substantially to protect pollination services. Agriculture also receives other benefits from biodiversity protection, such as maintaining landraces that contain the genetic diversity used for generating new and improved crop varieties and assisting crop breeding programs (5).

Addison and Bull support private sector involvement in a GAB but propose that the conservation targets apply a "no net loss or better" objective rather than E.O. Wilson's Half-Earth concept. We agree that establishing targets for terrestrial and marine biodiversity conservation should be supported by sound scientific evidence. The Convention on Biodiversity (CBD) is already considering how to extend the existing Aichi Targets toward the 50% conservation goal as part of its post-2020 agenda (6).

Addison and Bell also raise concerns about corporate motivations for financing a GAB. We agree that there may be a need to enhance the incentives for corporations to finance conservation by adopting complementary policies, improving environmental regulations, and increasing public pressure.

A good example is the emerging science-business initiative for ocean stewardship (7). As we stated in our Perspective, the inclusion of corporations alongside governments in supporting and implementing a GAB could also help coordinate and align their incentives for greater conservation. A carefully crafted government-corporation GAB could create substantial opportunities for enterprise and innovation, align incentives by eliminating corporate free-rider problems, and revitalize the global system of biodiversity protection.

**Edward B. Barbier,<sup>1,2\*</sup> Joanne C. Burgess,<sup>1,2</sup> Thomas J. Dean<sup>2,3</sup>**

<sup>1</sup>Department of Economics, Colorado State University, Fort Collins, CO 80523, USA. <sup>2</sup>School of Global Environmental Sustainability, Colorado State University, Fort Collins, CO 80523, USA. <sup>3</sup>College of Business, Colorado State University, Fort Collins, CO 80523, USA.

\*Correspondence author. Email: edward.barbier@colostate.edu

#### REFERENCES

1. G. Ceballos, P. R. Ehrlich, R. Dirzo, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6089 (2017).
2. B. Worm *et al.*, *Science* **314**, 787 (2006).
3. E. Dinerstein *et al.*, *BioScience* **67**, 534 (2017).
4. D. C. Esty, P. Boyd, "To move Paris Accord forward, bring cities and companies on board" *Yale Environment* 360

- (2018); <https://e360.yale.edu/features/to-move-paris-accord-forward-bring-cities-and-companies-on-board>.
5. M. R. Bellon, in *Agrobiodiversity Conservation and Economic Development*, A. Kontoleon, U. Pascual, M. Smale, Eds. (Routledge, 2008), ch. 4; [www.taylorfrancis.com/books/e/9781134039104/chapters/10.4324%2F9780203890127-13](http://www.taylorfrancis.com/books/e/9781134039104/chapters/10.4324%2F9780203890127-13).
6. J. Watts, "Make half of the world more nature friendly by 2050, urges UN biodiversity chief," *The Guardian* (2018); [www.theguardian.com/environment/2018/apr/13/make-half-the-world-more-nature-friendly-by-2050-says-un-chief](http://www.theguardian.com/environment/2018/apr/13/make-half-the-world-more-nature-friendly-by-2050-says-un-chief).
7. H. Österblom *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9038 (2017).

10.1126/science.aau1746

#### TECHNICAL COMMENT ABSTRACTS

**Comment on "Satellites reveal contrasting responses of regional climate to the widespread greening of Earth"**

**Yue Li, Zhenzhong Zeng, Ling Huang, Xu Lian, Shilong Piao**

Forzieri *et al.* (Reports, 16 June 2017, p. 1180) used satellite data to show that boreal greening caused regional warming. We show that this positive sensitivity of temperature to the greening can be derived from the positive response of vegetation to boreal warming, which indicates that results from a statistical

regression with satellite data should be carefully interpreted.

Full text: [dx.doi.org/10.1126/science.aap7950](https://doi.org/10.1126/science.aap7950)

**Response to Comment on "Satellites reveal contrasting responses of regional climate to the widespread greening of Earth"**

**Giovanni Forzieri, Ramdane Alkama, Diego G. Miralles, Alessandro Cescatti**

Li *et al.* contest the idea that vegetation greening has contributed to boreal warming and argue that the sensitivity of temperature to leaf area index (LAI) is instead likely driven by the climate impact on vegetation. We provide additional evidence that the LAI-climate interplay is indeed largely driven by the vegetation impact on temperature and not vice versa, thus corroborating our original conclusions.

Full text: [dx.doi.org/10.1126/science.aap9664](https://doi.org/10.1126/science.aap9664)

#### ERRATA

**Erratum for the Report "Seasonal and daily climate variation have opposite effects on species elevational range size" by W.-P. Chan *et al.*, *Science* **360**, eaat9919 (2018).** Published online 4 May 2018; 10.1126/science.aat9919



## TECHNICAL COMMENT

## CLIMATE CHANGE

# Comment on “Satellites reveal contrasting responses of regional climate to the widespread greening of Earth”

Yue Li<sup>1</sup>, Zhenzhong Zeng<sup>1,2\*</sup>, Ling Huang<sup>1</sup>, Xu Lian<sup>1</sup>, Shilong Piao<sup>1,3,4</sup>

Forzieri *et al.* (Reports, 16 June 2017, p. 1180) used satellite data to show that boreal greening caused regional warming. We show that this positive sensitivity of temperature to the greening can be derived from the positive response of vegetation to boreal warming, which indicates that results from a statistical regression with satellite data should be carefully interpreted.

Forzieri *et al.* (1) presented a statistical analysis of satellite data to quantify the climate impact of vegetation greening. They used a huge number of observations combined with a method that divided global grids into bioclimatic spaces, making this study of particular importance for understanding the temporal covariations between leaf area index (LAI) and the surface biophysics across different climate zones and during extreme warm-dry/cold-wet years. However, one of their main con-

clusions, that “the increasing trend in LAI contributed to the warming of boreal zones,” raises concerns about whether the climate impact of boreal greening can be isolated from this kind of statistical regression with satellite data.

The conclusion that boreal greening has contributed to boreal warming was established on the basis of the statistically positive sensitivity of land surface temperature ( $T_s$ ) to LAI (i.e.,  $\delta T^{\text{LAI}}$ ) over the high latitudes including northern Canada, central Europe, and western Siberia (Fig. 1A).

We argue that it is not appropriate to directly apply the methodology in Wang *et al.* (2) to compute  $\delta T^{\text{LAI}}$  “as the partial derivative of surface temperature with respect to LAI in a multiple regression of surface temperature against LAI, precipitation and incoming shortwave radiation after detrending all variables” [supplementary text in (1)]. Note that the basic hypothesis in Forzieri *et al.* contrasts with that in Wang *et al.* (2), with the latter assuming that the atmospheric carbon dioxide growth rate (reflecting carbon flux fluctuations in tropical terrestrial ecosystems) is determined by the climate (i.e., temperature, precipitation, and incoming solar radiation). Therefore, the causality between vegetation and climate is reversed in Forzieri *et al.* Over the boreal cold regions, warming is the dominant driver of vegetation greening in the high latitudes (3, 4), where the positive  $\delta T^{\text{LAI}}$  may be a result of the vegetation response to climate change rather than a metric measuring the climate response to vegetation greening.

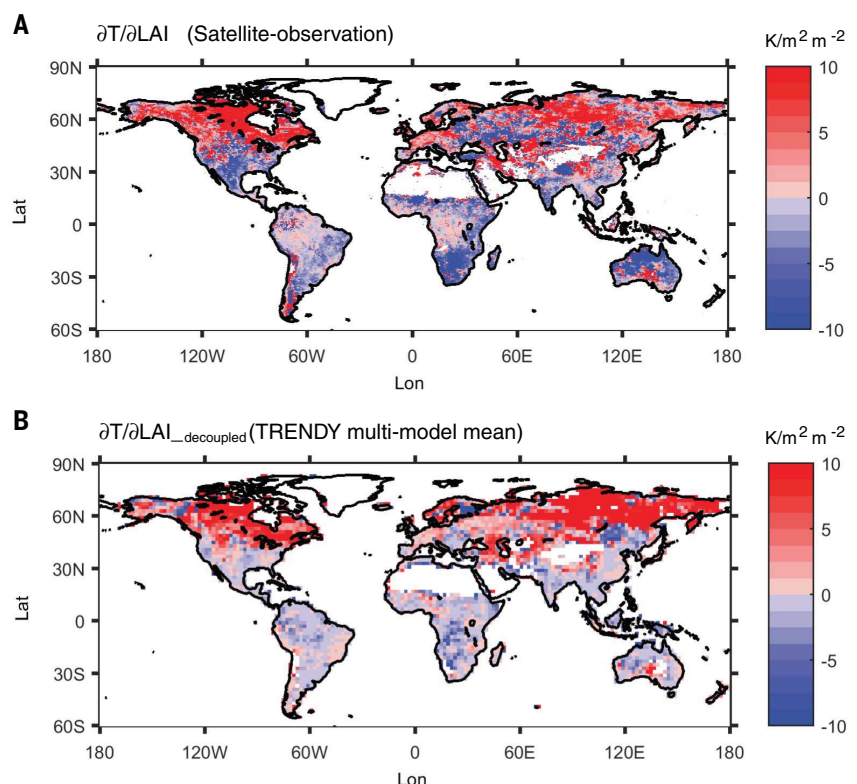
To demonstrate the unsuitability of  $\delta T^{\text{LAI}}$  for representing the climate impact of vegetation greening, we recomputed the sensitivity, applying the same method as in Forzieri *et al.* but

<sup>1</sup>Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. <sup>2</sup>Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA. <sup>3</sup>Key Laboratory of Alpine Ecology and Biodiversity, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100085, China. <sup>4</sup>Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100085, China.

\*Corresponding author. Email: zzeng@princeton.edu

**Fig. 1. Temperature sensitivity to leaf area index (LAI).** (A) Spatial patterns of the

sensitivity of land surface temperature ( $T_s$ ) to LAI from satellite observations ( $\partial T/\partial \text{LAI}$ ). (B) Sensitivity of observed surface air temperature ( $T_a$ ) to the simulated LAI from the multimodel mean of five global ecosystem models participating in the TRENDY project ( $\partial T/\partial \text{LAI}_{\text{decoupled}}$ ).  $T_s$  is the average of the daytime and nighttime temperature. The computation of  $\partial T/\partial \text{LAI}$  and  $\partial T/\partial \text{LAI}_{\text{decoupled}}$  followed the methodology in Forzieri *et al.* (1), using the satellite-observed  $T_s$  and LAI for  $\partial T/\partial \text{LAI}$ , and the observed  $T_a$  and the simulated LAI for  $\partial T/\partial \text{LAI}_{\text{decoupled}}$ , respectively.



replacing the data with the offline simulations from five global terrestrial ecosystem models (4). Because these offline models were driven by the observed climate and atmospheric CO<sub>2</sub> concentration, the causality between vegetation and climate in the simulations was unidirectional; that is, the calculated  $\delta T^{\text{LAI}}$  in the offline model simulations (i.e.,  $\delta T^{\text{LAI}}_{\text{decoupled}}$ ) mainly reflects the vegetation response to climate change. We found that over the boreal regions,  $\delta T^{\text{LAI}}_{\text{decoupled}}$  is positive and at the same magnitude as  $\delta T^{\text{LAI}}$  (Fig. 1B versus Fig. 1A), indicating that the positive  $\delta T^{\text{LAI}}$  regressed with satellite data, which has been used to reflect the climate feedbacks of vegetation greening in Forzieri *et al.*, can be derived from the positive response of vegetation to boreal warming. Therefore, it may not be possible to isolate the signal of regional climate response to vegetation greening from such a statistical regression with satellite data, the results of which should therefore be carefully interpreted.

Finally, it is worth mentioning that the uncertainty in regional climate change induced by vegetation change remains large in the boreal regions. For observation-based analyses, one limitation could be a lack of evidence to show the causality between vegetation and climate due to the statistical analysis often relying on correlations. For example, it is almost impossible to tell whether boreal forest warms the regional climate or whether boreal forest occurs in relatively warm regions when interpreting the results of the positive spatial relationship be-

tween vegetation types and climate in boreal regions (5), unless extra information—such as a plantation map (6) and/or the background climate variability (7)—is available to help exclude the climate impact on vegetation growth and isolate the signal of the vegetation impact on the regional climate. Recently, Zeng *et al.* (8) used coupled land-atmosphere global climate models integrated with satellite data to show that a 30-year vegetation greening could have slowed the global land-surface warming, including in the boreal zones, as a result of the dominant effect of evapotranspiration cooling. Relative to boreal afforestation, the observed boreal greening occurs during the growing season when the snow-albedo feedback is minimal [figure S4 in (8)]. Although the continuous summer-to-winter greening of coniferous forests and the advance of beginning of leaf growing of deciduous forests still trigger the snow-albedo warming feedback, the climate impact of growing season greening seems to be dominated by evapotranspiration cooling even in the boreal regions (8, 9). Unlike the statistical method used by Forzieri *et al.*, the experimental design in Zeng *et al.* (8) could isolate the climate effect of the greening of Earth from the background global warming and other confounding forcings, although substantial uncertainty was also present in their study (e.g., results were dependent on model parameter schemes). We suggest that it is necessary to combine the coupled land-atmosphere model (8, 9), as an important tool, with extra infor-

mation in observation-based analyses to reduce the uncertainty in quantifying the climate effects of the greening of Earth in future studies.

## REFERENCES AND NOTES

1. G. Forzieri, R. Alkama, D. G. Miralles, A. Cescatti, *Science* **356**, 1180–1184 (2017).
2. X. Wang *et al.*, *Nature* **506**, 212–215 (2014).
3. W. Lucht *et al.*, *Science* **296**, 1687–1689 (2002).
4. Z. Zhu *et al.*, *Nat. Clim. Change* **6**, 791–795 (2016).
5. Y. Li *et al.*, *Nat. Commun.* **6**, 6603 (2015).
6. S.-S. Peng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2915–2919 (2014).
7. R. Alkama, A. Cescatti, *Science* **351**, 600–604 (2016).
8. Z. Zeng *et al.*, *Nat. Clim. Change* **7**, 432–436 (2017).
9. Z. Zeng *et al.*, *J. Clim.* **31**, 2633–2650 (2018).

## ACKNOWLEDGMENTS

This study was supported by the National Key R&D Program of China (2017YFA0604702), the National Natural Science Foundation of China (41530528 and 41561134016), the 111 Project (B14001), and the National Youth Top-notch Talent Support Program in China. The data for computing the observed sensitivity of land surface temperature to the LAI were derived from moderate-resolution imaging spectroradiometer (MODIS) land surface temperature products (MYD11C2, [https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/myd11c2\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/myd11c2_v006)), GIMMS3g LAI datasets (<http://sites.bu.edu/cliveg/datacodes/>), precipitation from the Climate Research Unit (CRU) at the University of East Anglia ([http://badc.nerc.ac.uk/browse/badc/cru/data/cru\\_ts/cru\\_ts\\_3.21/data](http://badc.nerc.ac.uk/browse/badc/cru/data/cru_ts/cru_ts_3.21/data)), and downward shortwave radiation from the Clouds and the Earth's Radiant Energy System project (CERES, <https://ceres-tool.larc.nasa.gov/ord-tool/jsp/SYN1degSelection.jsp>). Computing the sensitivity of surface air temperature to simulated LAI required extra datasets of surface air temperature from the CRU and the LAI from five global ecosystem models (i.e., CLM4.5, JULES, OCN, ORCHIDE, and VISIT) participating in the TRENDY project (<http://dgvn.ceh.ac.uk/node/9>).

27 August 2017; accepted 20 April 2018  
10.1126/science.aap7950



## TECHNICAL RESPONSE

## CLIMATE CHANGE

# Response to Comment on “Satellites reveal contrasting responses of regional climate to the widespread greening of Earth”

Giovanni Forzieri<sup>1\*</sup>, Ramdane Alkama<sup>1</sup>, Diego G. Miralles<sup>2</sup>, Alessandro Cescatti<sup>1</sup>

Li *et al.* contest the idea that vegetation greening has contributed to boreal warming and argue that the sensitivity of temperature to leaf area index (LAI) is instead likely driven by the climate impact on vegetation. We provide additional evidence that the LAI-climate interplay is indeed largely driven by the vegetation impact on temperature and not vice versa, thus corroborating our original conclusions.

**O**n the basis of statistical analysis of satellite data, we argued that the observed vegetation greening has contributed to the warming of cold and humid boreal zones through a reduction of surface albedo (1). In their Comment, Li *et al.* (2) contend that “this positive sensitivity of temperature to the greening can be derived from the positive response of vegetation to boreal warming.” They base this claim on two major points: (i) The statistical analysis is un-

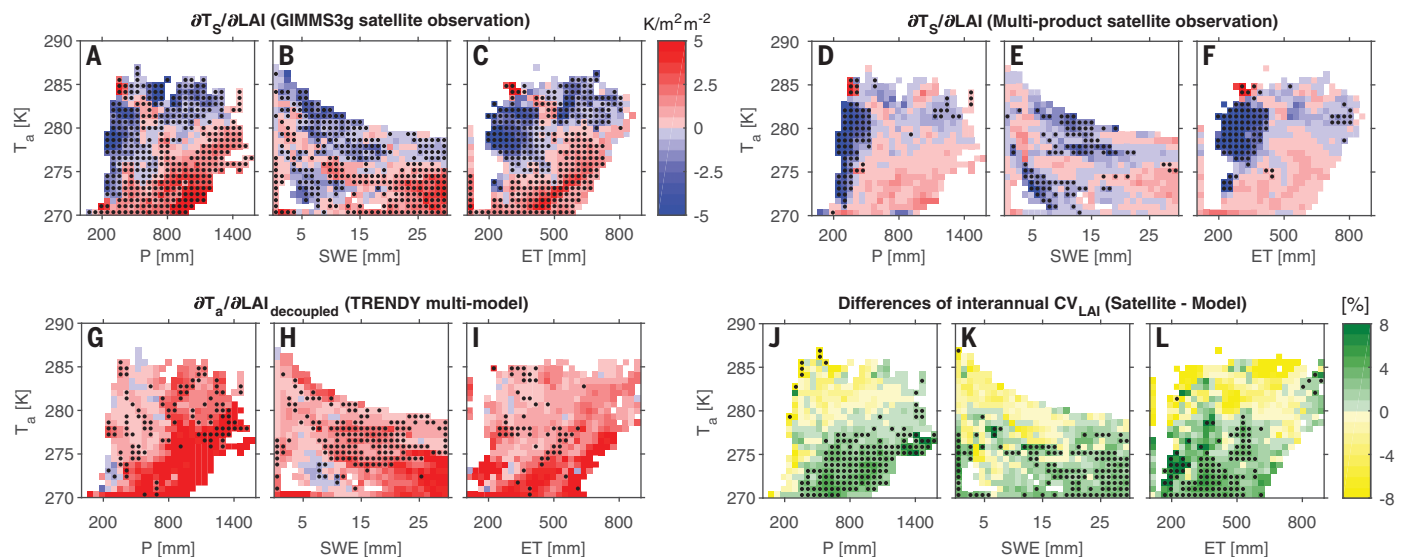
suitable to isolate the climate response to vegetation greening, and (ii) the importance of radiative processes associated with LAI-driven changes in boreal albedo is limited. In addition, Li *et al.* underpin a climate impact of greening dominated by evapotranspiration cooling in the boreal regions as suggested by model simulations (3), implicitly assuming that land models correctly represent the vegetation-climate interplay.

Li *et al.* support the first claim (i) by applying the statistical method used in (1) to decoupled simulations from five global terrestrial ecosystem models and conclude that “over the boreal regions,  $\delta T_a^{LAI}$  [i.e.,  $\partial T_a / \partial LAI$ ]<sub>decoupled</sub> is positive and at the same magnitude as  $\delta T_a^{LAI}$  [i.e.,  $\partial T_a / \partial LAI$ ]...”, indicating that the positive  $\delta T_a^{LAI}$  regressed with satellite data ... can be derived from the positive response of vegetation to boreal warming.” Even though we acknowledge the relevance of feedback loops in plant-climate interactions, we argue that this modeling experiment, being based on decoupled model runs, does not support the conclusions of Li *et al.* because it cannot exclude the direct impact of changes in LAI on surface temperature, and because it does not quantify the relative magnitude of the bidirectional effects and therefore cannot support any conclusion on which of the two dominates.

We stress that our methodology isolates the interactions between LAI and  $T_s$  and factors out the signal of covarying drivers such as precipitation and incoming radiation. Nonetheless, we recognize that the statistical approach cannot fully disentangle the short-term temperature control on boreal vegetation and separate it from the feedback. However, our results support the conclusion that at the interannual time scale, the

<sup>1</sup>European Commission, Joint Research Centre, Directorate for Sustainable Resources, Ispra, Italy. <sup>2</sup>Laboratory of Hydrology and Water Management, Department of Forest and Water Management, Ghent University, B-9000 Ghent, Belgium.

\*Corresponding author. Email: giovanni.forzieri@ec.europa.eu



**Fig. 1. Sensitivity of temperature to variations in LAI and interannual variations in LAI across climatological gradients. (A to C)** For boreal regions and based on satellite GIMMS3g observations, values of sensitivity of surface temperature to LAI ( $\partial T_s / \partial LAI$ ) are illustrated against the climatological median temperature ( $T_a$ , y axis) and precipitation (A), snow water equivalent (B), and evapotranspiration (C) (P, SWE, and ET; x axis). **(D to F)** Same as (A) to (C) but for multiproduct satellite observations (GIMMS3g, GLASS, GLOBMAP, LTDR, and MODIS). **(G to I)** Same as (A) to (C) but for TRENDY (CLM4.5, JULES, OCN, ORCHIDEE, and VISIT) model-derived

sensitivity of air temperature to LAI ( $\partial T_a / \partial LAI$ ). **(J to L)** Same as (A) to (C) but for the differences in interannual variations of LAI, expressed as the coefficient of variation ( $CV_{LAI}$ ), retrieved from each combination of satellite products and TRENDY models (25-member ensemble). Black dots show bins with average value statistically different from zero ( $P < 0.05$ ). Statistical significance in (A) to (C) is calculated for all the observed values falling into the given bin and reflects the robustness of the observed signal. Statistical significance in (D) to (L) is calculated across the members of the ensemble and reflects the robustness of their average signal.

land-climate interplay is largely affected by the LAI impact on temperature. Here, we extend our analyses to provide further evidence. Figure 1 unequivocally shows that the observed temperature variations associated with boreal greening depend in sign and magnitude on the background climate, with a pattern that is fully consistent with the expected biophysical impact of vegetation on temperature. As such,  $\partial T_s / \partial \text{LAI}$  is positive in cold-humid boreal regions with extended snow cover, where the greening can trigger large reductions in albedo but only limited increases in evapotranspiration. On the contrary, in boreal areas with limited snow cover and water availability, the increase in transpiration associated with the greening ultimately yields a negative  $\partial T_s / \partial \text{LAI}$  (Fig. 1, A to C). We note that despite the spread observed over multiple satellite LAI products (4), the multiproduct average signal of  $\partial T_s / \partial \text{LAI}$  shows largely consistent patterns (Fig. 1, D to F). If the sensitivity of temperature to boreal greening was mainly driven by the positive response of vegetation to warming, as suggested by Li *et al.*, we should expect a positive value of  $\partial T_s / \partial \text{LAI}$  over the whole boreal climate domain, consistent with the results of the modeling experiment (Fig. 1, G to I). On the contrary, the clear contrast between observations and decoupled simulations suggests that the interplay between climate and LAI at the interannual scale is largely modulated by the vegetation impacts on climate as suggested in (1), and not vice versa.

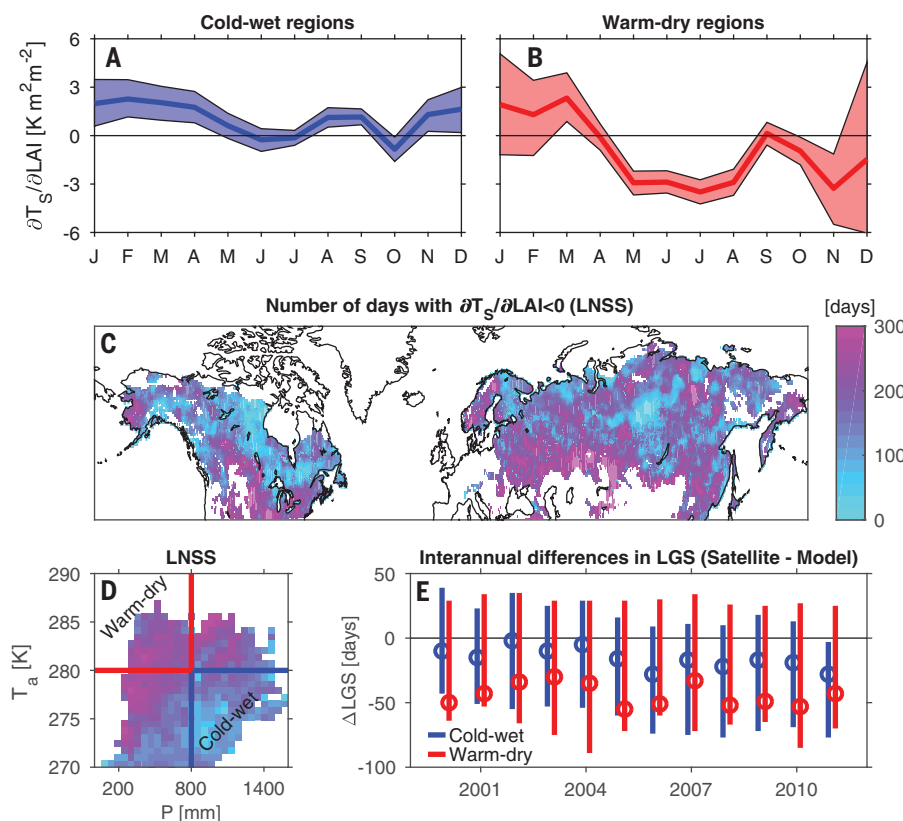
Furthermore, relative to observations, models substantially underestimate the interannual variability of LAI in cold-wet zones, as highlighted

by the large difference (up to 8%;  $P < 0.05$ ) in the coefficient of variation (Fig. 1, J to L). Similar discrepancies are also found for growing-season LAI (not shown). Such model limitations are likely to hamper their capacity to represent the interplay between biophysical processes and vegetation changes (5, 6). As acknowledged also by Li *et al.*, we stress that the large model uncertainty may hinder the assessment of the net impact of vegetation greening on boreal climate.

On the second point (ii), Li *et al.* claim that there is a limited impact of LAI on radiative biophysical processes (e.g., albedo) in the boreal zone: “the observed boreal greening occurs during the growing season when the snow-albedo feedback is minimal (3)..., the climate impact of growing season greening seems to be dominated by evapotranspiration cooling even in the boreal regions.” This conclusion is based on the rather uncertain predictions of land surface models on the relationship between vegetation and albedo off-season (7, 8). On the other hand, remote sensing observations report on the abundance and spreading of evergreen species in the boreal zone (9), whose variation in LAI may substantially affect albedo and temperature also during the dormancy season via the snow-albedo feedback (10). The relevance of off-season radiative biophysical processes is suggested by the positive values of winter/spring  $\partial T_s / \partial \text{LAI}$  in the whole boreal climate domain (Fig. 2, A and B). In warm-dry regions, the observed sensitivity becomes negative during the growing season, thanks to the stronger control that LAI exerts on transpiration in less humid ecosystems. These

seasonal dynamics cannot be the simple effect of temperature-driven greening, as claimed by Li *et al.*, because in that case we would expect a positive sensitivity throughout the year across the whole boreal region. On the contrary, the emerging seasonal patterns confirm that LAI variation can indeed influence in space and time the short-term net signal of the vegetation-climate interactions.

Ultimately, the net annual LAI impact on climate depends on the relative magnitude and duration of the periods dominated by radiative and nonradiative biophysical processes (11, 12), as clearly proved by the consistency between the climate patterns of  $\partial T_s / \partial \text{LAI}$  in Fig. 1A and the length of the season with negative sensitivity (LNSS) in Fig. 2, C and D. Unfortunately, models show severe limitations in predicting the seasonality of vegetation and the length of the growing season (LGS), here expressed as the number of days when LAI is larger than 30% of its amplitude (13) and derived after linear interpolation of monthly LAI values (Fig. 2E). In particular, models systematically overestimate LGS ( $16 \pm 2$  days and  $44 \pm 3$  days in cold-wet and warm-dry zones, respectively; multimodel average  $\Delta \text{LGS} \pm \text{SE}$ ). We highlight that the threshold-based approach used to derive LGS is not sensitive to missing LAI data during winter (14). The emerging model bias in the estimate of LGS is consistent with previous studies (15) and may be one of the causes of the overestimation of LAI-driven evaporative cooling, with potential influences on the derived net climate impact of boreal greening.



**Fig. 2. Seasonal patterns of the sensitivity of surface temperature to LAI and length of the season with negative sensitivity (LNSS).**

(A) Satellite-derived monthly sensitivity of surface temperature to LAI for cold-wet boreal regions ( $T < 280 \text{ K}$  and  $P > 800 \text{ mm}$ , in blue); lines and shaded areas denote the median values and confidence bounds ( $\pm \text{SE}$ ). Note that monthly sensitivity of surface temperature to LAI is computed by using annual LAI values and monthly-scale climate drivers in order to minimize the potential biases of satellite retrieval of LAI in snow cover conditions. (B) Same as (A) but for warm-dry boreal regions ( $T > 280 \text{ K}$  and  $P < 800 \text{ mm}$ , in red). (C and D) Spatial (C) and climate (D) pattern of satellite-derived LNSS (number of days with negative sensitivity of surface temperature to LAI). (E) Interannual differences in length of the growing season ( $\Delta \text{LGS}$ ) were estimated from observations and models; circles and bars reflect the ensemble median and the five-member ensemble data range, respectively. The GIMMS3g product is used as the reference observational dataset for consistency with (1).



Our findings in (1), further corroborated by the results reported here, suggest that the impacts of LAI variations on climate are modulating the overall signal of the interannual vegetation-climate interplay. They also highlight the importance of realistically representing the temporal variation of vegetation in land surface models for simulating present and future changes in the energy and water cycle (5, 6). We agree that at longer time scales, like those investigated in (3), the strong trend in temperature and covarying changes in atmospheric CO<sub>2</sub> and N deposition may have driven the northern greening. On the other hand, our analysis suggests that at shorter temporal scales, LAI variations affect the inter-annual spatial and seasonal variability of surface temperature. We agree with Li *et al.* that future studies based on the integration of models and observations are required to reconcile these findings and derive conclusive statements about

the climate effects of the widespread greening of Earth.

#### REFERENCES AND NOTES

1. G. Forzieri, R. Alkama, D. G. Miralles, A. Cescatti, *Science* **356**, 1180–1184 (2017).
2. Y. Li, Z. Zeng, L. Huang, X. Lian, S. Piao, *Science* **360**, eaap7950 (2018).
3. Z. Zeng *et al.*, *Nat. Clim. Change* **7**, 432–436 (2017).
4. C. Jiang *et al.*, *Glob. Change Biol.* **23**, 4133–4146 (2017).
5. Y. Zhang *et al.*, *Sci. Rep.* **6**, 19124 (2016).
6. G. Forzieri *et al.*, *J. Adv. Model. Earth Syst.* 10.1002/2018MS001284 (2018).
7. A. J. Pitman *et al.*, *Geophys. Res. Lett.* **36**, L14814 (2009).
8. J. P. Boisier *et al.*, *J. Geophys. Res. Atmos.* **117**, D12116 (2012).
9. Y. He *et al.*, *J. Clim.* **30**, 5021–5039 (2017).
10. F. S. Chapin III *et al.*, *Glob. Change Biol.* **6** (S1), 211–223 (2000).
11. G. B. Bonan, *Science* **320**, 1444–1449 (2008).
12. R. M. Bright *et al.*, *Nat. Clim. Change* **7**, 296–302 (2017).
13. A. Verger, I. Filella, F. Baret, J. Peñuelas, *Remote Sens. Environ.* **178**, 1–14 (2016).
14. Z. Zhu *et al.*, *Remote Sens.* **5**, 927–948 (2013).
15. G. Murray-Tortarolo *et al.*, *Remote Sens.* **5**, 4819–4838 (2013).

#### ACKNOWLEDGMENTS

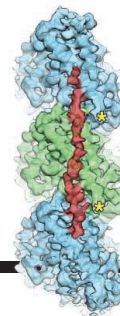
Supported by the Seventh Framework Program (FP7) LUC4C project (grant 603542). D.G.M. acknowledges support from the European Research Council under grant agreement 715254 (DRY-2-DRY). Satellite LAI products are provided by the Global Inventory Modeling and Mapping Studies (GIMMS3g, <http://sites.bu.edu/cliveg/datacodes/>), Global Land Surface Satellite (GLASS, <http://glcf.umd.edu/data/lai/>), Global Mapping (GLOMAP, [www.globalmapping.org/](http://www.globalmapping.org/)), NOAA Long Term Data Record (LTDR, <https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00898>), and Moderate Resolution Imaging Spectroradiometer (MODIS, [https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table/mcd15a3h\\_v006](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd15a3h_v006)). Satellite-derived evapotranspiration and snow water equivalent data are derived from the Global Land Evaporation Amsterdam Model Version 3a (GLEAM v3a, [www.gleam.eu/](http://www.gleam.eu/)) and from the Finnish Meteorological Institute (GLOBSNOW, [www.globsnow.info/index.php?page=SWE](http://www.globsnow.info/index.php?page=SWE)), respectively. Availability and locations of additional satellite data may be found in the supplementary materials of (1). Offline simulations from five global terrestrial ecosystem models (i.e., CLM4.5, JULES, OCN, ORCHIDEE, and VISIT) were obtained from the TRENDY project (<http://dgvm.ceh.ac.uk/node/9>).

28 September 2017; accepted 20 April 2018  
10.1126/science.aap9664

# RESEARCH

## Tau binding to tubulin monomers

Kellogg et al., p. 1242



## IN SCIENCE JOURNALS

Edited by **Stella Hurtley**



Human activities are driving many animals, including coyotes (*Canis latrans*) like this one, to become more nocturnal.

### HUMAN IMPACTS

## Nocturnal refuge

**A**s the human population grows, there are fewer places for animals to live out their lives independently of our influence. Given our mostly diurnal tendencies, one domain that remains less affected by humans is the night. Gaynor *et al.* found that across the globe and across mammalian species—from deer to coyotes and from tigers to wild boar—animals are becoming more nocturnal (see the Perspective by Benítez-López). Human activities of all kinds, including nonlethal pastimes such as hiking, seem to drive animals to make use of hours when we are not around. Such changes may provide some relief, but they may also have ecosystem-level consequences. —SNV  
*Science*, this issue p. 1232; see also p. 1185

is consistent with adjacent layers naturally having opposite magnetizations, which align parallel to each other in high magnetic fields. —JS  
*Science*, this issue p. 1218, p. 1214

### PHOTOSYNTHESIS

## Lower-energy photons do the work, too

Plants and cyanobacteria use chlorophyll-rich photosystem complexes to convert light energy into chemical energy. Deep-water organisms do not receive the full spectrum of light and have adaptations to take advantage of longer-wavelength photons. Nürnberg *et al.* studied photosystem complexes from cyanobacteria grown in the presence of far-red light. The authors identified the primary donor chlorophyll as one of a few chlorophyll molecules in the red light-adapted enzymes that were chemically altered to shift their absorption spectrum. Kinetic measurements demonstrated that far-red light is capable of directly driving water oxidation, despite having less energy than the red light used by most photosynthetic organisms. —MAF  
*Science*, this issue p. 1210

### MACHINE LEARNING

## A scene-internalizing computer program

To train a computer to “recognize” elements of a scene supplied by its visual sensors, computer scientists typically use millions of images painstakingly labeled by humans. Eslami *et al.* developed an artificial vision system, dubbed the Generative Query Network (GQN), that has no need for such labeled data. Instead, the GQN first uses images taken from different viewpoints and creates an

abstract description of the scene, learning its essentials. Next, on the basis of this representation, the network predicts what the scene would look like from a new, arbitrary viewpoint. —JS  
*Science*, this issue p. 1204

### MAGNETISM

## An intrinsic magnetic tunnel junction

An electrical current running through two stacked magnetic layers is larger if their magnetizations point in the same

direction than if they point in opposite directions. These so-called magnetic tunnel junctions, used in electronics, must be carefully engineered. Two groups now show that high magnetoresistance intrinsically occurs in samples of the layered material  $\text{CrI}_3$  sandwiched between graphite contacts. By varying the number of layers in the samples, Klein *et al.* and Song *et al.* found that the electrical current running perpendicular to the layers was largest in high magnetic fields and smallest near zero field. This observation

### CYTOMETRY

## Seeing ghosts

In fluorescence-activated cell sorting, characteristic target features are labeled with a specific fluorophore, and cells displaying different fluorophores are sorted. Ota *et al.* describe a technique called ghost cytometry that allows cell sorting based on the



morphology of the cytoplasm, labeled with a single-color fluorophore. The motion of cells relative to a patterned optical structure provides spatial information that is compressed into temporal signals, which are sequentially measured by a single-pixel detector. Images can be reconstructed from this spatial and temporal information, but this is computationally costly. Instead, using machine learning, cells are classified directly from the compressed signals, without reconstructing an image. The method was able to separate morphologically similar cell types in an ultrahigh-speed fluorescence imaging-activated cell sorter. —VV

*Science*, this issue p. 1246

## ANTHROPOLOGY

### Mesoamerican turquoise locally sourced

Scholars have long assumed that Aztec and Mixtec turquoise artifacts uncovered in Mesoamerica were imported from the American Southwest, which is home to abundant turquoise deposits. Thibodeau *et al.* analyzed the lead and strontium isotopic ratios in 38 Mesoamerican turquoise mosaic tiles (tesserae) from the Sacred Precinct of Tenochtitlan and in five tesserae from Mixtec turquoise mosaics. The isotopic compositions of most of the samples matched copper

deposits and crustal rocks of Mesoamerica, suggesting at least one Mesoamerican region as the source of this turquoise. —PJB

*Sci. Adv.* 10.1126/sciadv.aas9370 (2018).

## NEUROSCIENCE

### Disentangling specific memories

Each memory is stored in a distinct memory trace in the brain, in a specific population of neurons called engram cells. How does the brain store and define the identity of a specific memory when two memories interact and are encoded in a shared engram? Abdou *et al.* used optogenetic reactivation coupled with manipulations of long-term potentiation to analyze engrams that share neurons in the lateral amygdala (see the Perspective by Ramirez). Synapse-specific plasticity guaranteed the storage and the identity of individual memories in a shared engram. Moreover, synaptic plasticity between specific engram assemblies was necessary and sufficient for memory engram formation. —PRS

*Science*, this issue p. 1227; see also p. 1182

## INFECTIOUS DISEASE

### Designer bugs as drugs

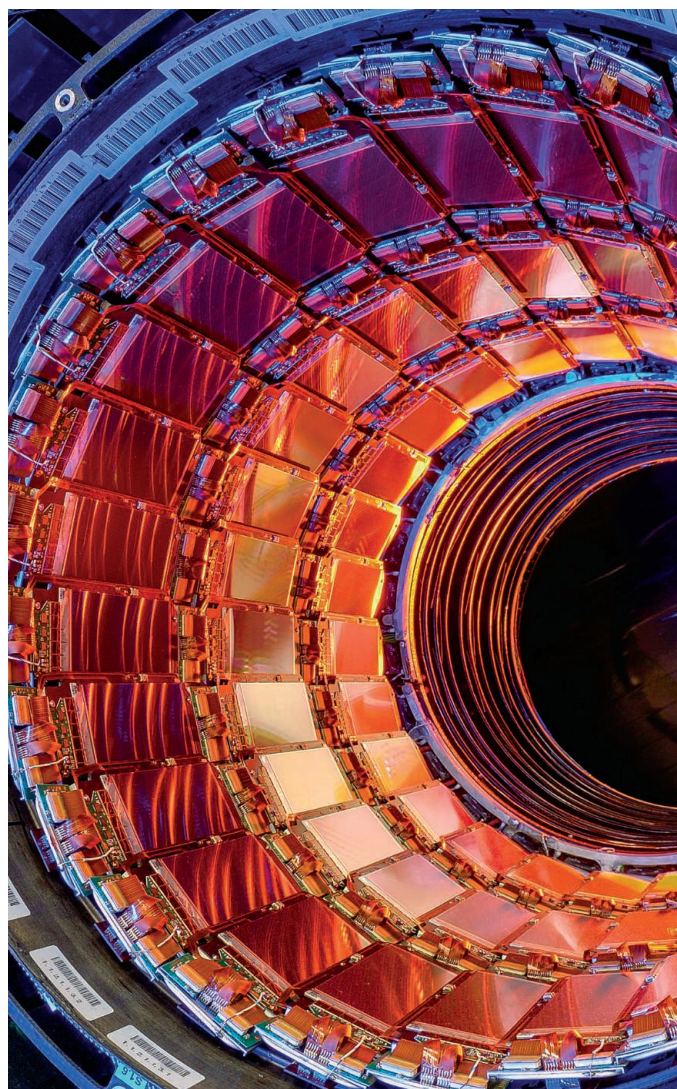
The endemic persistence and outbreaks of *Vibrio cholerae* indicate a need for new methods of control. Mao *et al.* discovered that lactic acid production by the

probiotic *Lactococcus lactis* rendered the infant mouse gut hostile to *V. cholerae* and engineered *L. lactis* to detect breakthrough infection. Hubbard *et al.* extensively modified a contemporary *V. cholerae* strain for a live oral vaccine, which resulted in an attenuated strain that could protect infant rabbits from *V. cholerae* challenge within 24 hours of vaccine administration. —LP

*Sci. Transl. Med.* 10, eaao2586, eaap8423 (2018).

## IN OTHER JOURNALS

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



Part of the Compact Muon Solenoid at CERN's Large Hadron Collider

## PARTICLE PHYSICS

### In the company of top quarks

The coupling of elementary particles to the Higgs boson is correlated with their masses. For some particles, it can be measured by observing the decay of the Higgs into those particles. The top quark, however, is too heavy for the Higgs to decay into. Therefore, Sirunyan *et al.* from the Compact Muon Solenoid Collaboration at the Large Hadron Collider (LHC) looked at the processes in which the Higgs boson is produced in conjunction with the top quark and its antiparticle. By combining the results from the runs at different center-of-mass energies of proton-proton collisions at the LHC, they found that such processes indeed occurred and that the signal strength was consistent with the predictions of the Standard Model of particle physics. —JS

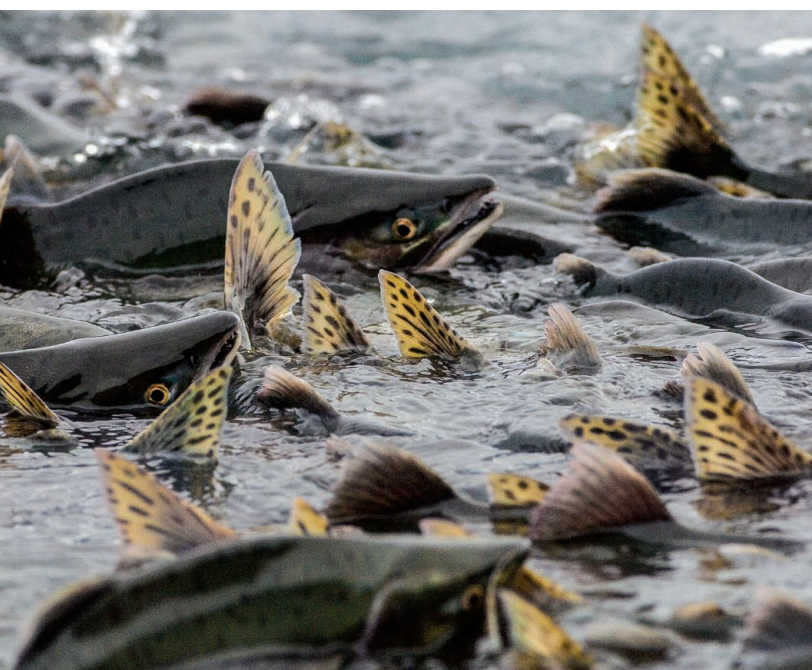
*Phys. Rev. Lett.* 120, 231801 (2018).



Isotope analysis indicates a Central American source for Aztec turquoise.

PHOTOS: (LEFT TO RIGHT) NATIONAL GEOGRAPHIC CREATIVE/ALAMY STOCK PHOTO; NG IMAGES/ALAMY STOCK PHOTO





## ECOSYSTEMS

## Salmon teleconnection disservice

**P**ink salmon (*Oncorhynchus gorbuscha*) populations in the North Pacific have grown during the past four decades. Fish numbers have expanded partly in response to climate change and partly because of hatchery release from Russia and the United States. Perhaps surprisingly, this is not entirely good news. Among other problems of competition, Springer *et al.* have discovered a remarkable 15,000-kilometer link between the pink salmon and the migration of a shearwater (*Ardenna tenuirostris*), which nests in Australia and New Zealand and winters in the North Pacific. Using data collected from ship-towed continuous plankton recorders, Batten *et al.* show that the pink salmon's consumption of zooplankton could be the cause of starvation and mass mortality of the birds. This in turn affects Pacific aboriginal societies that harvest the returning birds for food. —CA

*Proc. Natl. Acad. Sci. U.S.A.* 10.1073/pnas.1720577115 (2018);  
*Fisher. Oceanogr.* 10.1111/fog.12276 (2018).

Pink salmon spawning in an Alaskan river

## MICROBIOME

## Gut bacteria relieve epilepsy

A high-fat and low-carbohydrate, or ketogenic, diet is used to treat children with refractory epilepsy. Olson *et al.* show that the ketogenic diet enriches populations of the gut commensal bacteria *Akkermansia muciniphila* and *Parabacteroides* spp. The presence of these species appears to protect against induced epileptic seizures in mice, even in those on a nonketogenic diet. Together, these bacteria alter the profile of circulating metabolites, which subsequently modulates metabolism in the brain. Specifically, reduction in  $\gamma$ -glutamylolation of circulating amino acids and elevated hippocampal GABA ( $\gamma$ -aminobutyric acid)/glutamate are associated with seizure protection in mice. —GKA

*Cell* 10.1016/j.cell.2018.04.027 (2018).

## METABOLISM

## RNA processing for metabolic control

Autophagy is a cellular response to nutrient starvation that allows cells to recycle existing molecules and organelles. Tang *et al.* investigated the mechanisms by which TORC1 (target of

rapamycin complex 1) regulates autophagy. They discovered that regulation occurs through control of RNA processing. Using a genetic screen in fruitflies, they pinpointed the CPA (cleavage and polyadenylation) complex as the regulator of autophagy. CPA in turn regulates polyadenylation and alternative splicing of RNAs encoding the key proteins that regulate autophagy. This control appears to be conserved in mammalian cells, too: Loss of a component of the CPA complex disrupted regulation of metabolism and autophagy in cultured human cells. —LBR

*Cell Metab.* 10.1016/j.cmet.2018.02.023 (2018).

## HUMAN GENETICS

## Variants may affect disease heritability

Huntington's disease (HD) is a late-onset neurodegenerative disease caused by the inheritance of an expansion of a trinucleotide repeat in the *huntingtin* (*HTT*) gene. A Venezuelan population exhibits an earlier onset of HD than is observed in individuals of European ancestry. Chao *et al.* sequenced alleles from ~400 Venezuelan individuals carrying a pathogenic *HTT* allele. Although Venezuelan

sequence-specific changes were evident in the allele, they did not explain early onset of HD. A genome-wide investigation comparing European and Venezuelan individuals revealed other Mendelian modifying genes. This study demonstrates the importance of population-specific investigations of genetic diseases. —LMZ

*PLOS Genet.* 10.1371/journal.pgen.1007274 (2018).

## QUANTUM SENSING

## A quantum enhancement for plasmonic sensors

Laser light can induce electronic excitations on metal surfaces. These excitations, or plasmons, are sensitive to the surface environment and are now routinely used in chemical and biological sensing applications. Dowran *et al.* show that the sensitivity of these plasmonic sensors can be enhanced by using quantum states of light. They replace the laser light (used in the classical configuration) with quantum light, in this case a twin beam of entangled photons; one beam is used as a reference and the other to excite the surface plasmons. They find an appreciable sensitivity enhancement in the quantum configuration.

Moreover, the reduced light intensity could also prove useful for the detection of light-sensitive chemical or biological samples. —ISO

*Optica* 5, 628 (2018).

## STRUCTURAL BIOLOGY

## Edge cases cause trouble

Single-particle cryo-electron microscopy detects macromolecules trapped in a thin layer of vitreous ice. High-resolution, accurate reconstructions benefit from randomly arranged and well-dispersed particles, but biological samples do not always cooperate. Noble *et al.* used cryo-electron tomography to investigate particle distributions in cryo-electron microscopy samples. Troublingly, most macromolecules localized to the air-water interface, a phenomenon associated with denaturation of proteins and that may result in preferred orientations and uneven distribution of particles. Assessment of samples with cryo-electron tomography before further data collection would allow for tailored optimization of sample conditions, which should lead to higher-resolution structures and enable early detection of artifacts. —MAF

*eLife* 10.7554/eLife.34257 (2018).



ALSO IN *SCIENCE* JOURNALSEdited by **Stella Hurtley**

## IMMUNOLOGY

**Modeling memory differentiation in T cells**

The balance between effector and central memory T cells shifts toward the latter as the number of T cells participating in immune responses increases. Polonsky *et al.* determined the mechanisms by which T cell quorum sensing affects memory differentiation by using live-cell imaging to track cell proliferation and differentiation. They found that the rate of memory CD4<sup>+</sup> T cell differentiation is determined by cell number. This rate substantially increases above a threshold number of locally interacting cells. Mathematical modeling suggests that the number of initially seeded cells and the number of cell divisions are not critical. Instead, the instantaneous number of interacting cells continuously modulates the differentiation rate. This is partly fueled by increased sensitivity to the cytokines interleukin-2 (IL-2) and IL-6, independent of any effects on cell proliferation. —STS

*Science*, this issue p. 1201

## CANCER

**Chronic stress as a survival tactic**

Most patients with pancreatic ductal adenocarcinoma (PDA) develop liver metastases after surgical removal of their primary tumor. These metastases are thought to potentially arise from quiescent disseminated cancer cells, likely present at the time of surgery, which evade elimination by the immune system. Pommier *et al.* explored how these quiescent cells survive by analyzing mouse models and tissue samples from patients with PDA. They found that disseminated cancer cells do not express a cell surface molecule that triggers killing by T cells. This phenotypic feature is linked to their inability to resolve endoplasmic

reticulum stress. When this stress is resolved, the disseminated cells begin proliferating and form metastases. —PAK

*Science*, this issue p. 1202

## NANOTECHNOLOGY

**Reconsidering resonator sensing**

Changes in the frequency of a nanoscale mechanical resonator can be used for many sensing applications, provided that there is an adequate signal-to-noise ratio. Normally, this ratio is improved by creating resonators with higher quality factors that “ring” for longer times. Taking a cue from the approaches used in atomic force microscopy, Roy *et al.* show that if the thermomechanical noise of the resonator is well defined, the signal-to-noise ratio of the frequency shift can improve by lowering the quality factor. They used this approach to demonstrate temperature sensing with a double-clamped silicon beam resonator, which performed better at ambient pressures than in a vacuum. —PDS

*Science*, this issue p. 1203

## PLANT SCIENCE

**How to make bioactive alkaloids**

Vinblastine and vincristine are important, expensive anticancer agents that are produced by dimerization of the plant-derived alkaloids catharanthine and vindoline. The enzymes that transform tabersonine into vindoline are known; however, the mechanism by which the scaffolds of catharanthine and tabersonine are generated has been a mystery. Caputi *et al.* now describe the biosynthetic genes and corresponding enzymes responsible. This resolves a long-standing question of how plant alkaloid scaffolds are synthesized, which is important

not only for vinblastine and vincristine biosynthesis, but also for understanding the many other biologically active alkaloids found throughout nature. —PJH

*Science*, this issue p. 1235

## NEUROSCIENCE

**The mechanisms of fear attenuation**

Surprisingly little is known about how remote fearful memories are stored and attenuated. Khalaf *et al.* used independent fear memory attenuation paradigms, engram-based tagging techniques, and chemogenetic tools to alter neuronal activity (see the Perspective by Frankland and Josselyn). They found that a discrete subset of neurons within an ensemble is engaged during recall after memory attenuation, which correlated with fear reduction. Memory updating and extinction mechanisms thus likely coexist to make this happen. These findings support the notion that effective memory attenuation is mediated by a rewriting of the original memory trace of fear toward one of safety. —PRS

*Science*, this issue p. 1239; see also p. 1186

## STRUCTURAL BIOLOGY

**Tackling microtubule-tau interactions**

Alzheimer's disease is a major cause of death in the elderly. Disease progression is associated with the accumulation of neurofibrillary tangles composed of tau, a protein important for neuronal development and function. Tangle formation is preceded by phosphorylation events that cause tau to dissociate from its native binding partner, microtubules. Microtubule-tau interactions have been mysterious. Kellogg *et al.* used cryo-electron microscopy and molecular

modeling to show how tau interacts with the outer surface of the microtubule, stapling together tubulin subunits and thus stabilizing the polymer. A key tau amino acid within the tightly bound segment between tubulin subunits corresponds to a clinically relevant site of tau phosphorylation, explaining the competition between microtubule interaction and tau aggregation. —SMH

*Science*, this issue p. 1242

## NEUROSCIENCE

**Shifts in brain regions with brain size**

Brain size among normal humans varies as much as twofold. Reardon *et al.* surveyed the cortical and subcortical structure of more than 3000 human brains by noninvasive imaging (see the Perspective by Van Essen). They found that the scaling of different regions across the range of brain sizes is not consistent: Some brain regions are metabolically costly and are favored in larger brains. This shifts the balance between associative and sensorimotor brain systems in a brain size-dependent way. —PJH

*Science*, this issue p. 1222; see also p. 1184

## OCEANS

**Environmental DNA tracks rare species**

Marine animals are often elusive, making it difficult to establish their presence in a region or to estimate population sizes. Many of these animals are also rare, making them even more difficult to study. In a Perspective, Pikitch highlights the advantages of environmental DNA studies for detecting rare marine animals, particularly to aid conservation efforts. The method is noninvasive and highly sensitive, allowing detection of the recent presence of animals such as killer whales.

However, challenges remain, particularly for estimating population sizes, where the approach should prove useful in combination with more established methods. —JFU

*Science*, this issue p. 1180

## IMMUNOLOGY

### Letting thymocytes go

During the process of T cell development, thymocytes must travel from the cortex of the thymus to the medulla, where potentially autoreactive cells are removed by negative selection. Duke-Cohan *et al.* found that mouse thymocytes lacking the guanosine triphosphatase-activating protein Tagap failed to detach from the cortex. Given that in humans, single-nucleotide polymorphisms in the gene encoding TAGAP are associated with autoimmune disorders, the data suggest that this protein facilitates the trafficking required for the efficient negative selection of autoreactive cells. —JFF

*Sci. Signal.* **11**, eaan8799 (2018).

## HIV

### Spontaneous HIV controllers

A small number of HIV-infected individuals (<1%) can spontaneously control HIV in the absence of antiretroviral therapy. Because CD4<sup>+</sup> and CD8<sup>+</sup> T cell responses are thought to contribute to protection, HIV-responsive T cell receptors (TCRs) from these individuals are of considerable interest. Galperin *et al.* examined how three class II-restricted TCRs observed in spontaneous controllers are capable of binding a Gag peptide in the context of multiple HLA-DR molecules (HLA, human leukocyte antigen). The authors solved the structures of several TCR-peptide-HLA-DR complexes. The findings suggest that the ability of these TCRs to recognize the Gag peptide in the context of multiple HLA-DR allomorphs is shaped by extensive contacts between the TCRs and the peptide itself. —AB

*Sci. Immunol.* **3**, eaat0687 (2018).



## RESEARCH ARTICLE SUMMARY

## IMMUNOLOGY

# Induction of CD4 T cell memory by local cellular collectivity

Michal Polonsky, Jacob Rimer, Amos Kern-Perets, Irina Zaretsky, Stav Miller, Chamutal Bornstein, Eyal David, Naama Meira Kopelman, Gil Stelzer, Ziv Porat, Benjamin Chain\*, Nir Friedman\*

**INTRODUCTION:** Fate decisions play a central role in the operation of the vertebrate immune system. The establishment of efficient acquired immune responses depends on the differentiation of naïve T cells into various effector and memory cell types upon recognition of a cognate antigen, and on the appropriate balance between these populations. A number of studies have shown that the balance between effector and central memory T cells is shifted in favor of the latter when more T cells participate in the response. This observation has the hallmarks of quorum sensing, the ability of cells to respond to their population density. However, the mechanisms driving this behavior in T cells remain elusive.

**RATIONALE:** We observed increased differentiation of progenitor central memory T cells (pTCMs) at high cell densities both in vivo and in vitro. However, activated T cells rapidly form dense dynamic clusters, precluding the distinction between the effects of local interactions within cell clusters from global, long-range interactions

through soluble factors. To overcome these difficulties, we used live-cell imaging to track the proliferation and differentiation of cells cultured in microwell arrays. This microculture system provides precise control and monitoring of the number of interacting T cells and their state after T cell activation. Continuously tracking differentiation and proliferation enabled us to investigate the mechanisms of cellular collectivity and its influence on memory differentiation.

**RESULTS:** We first validated that the pTCM cells formed early in our cultures show the markings of established central memory T cells using RNA sequencing and in vivo experiments. Then, with our microwell system, we showed that the rate of differentiation of pTCM cells is determined by the number of cells within individual microwells and sharply increases above a threshold number of locally interacting cells. Further analysis showed that cells follow a universal differentiation trajectory, whereby their differentiation rate is con-

tinuously modulated by the instantaneous number of interacting cells, rather than simply by the number of cells present initially within each microwell, or by the number of cell divisions. A combination of experimental manipulation and computational simulations showed that the observed collectivity involved increased sensitivity of clustered T cells to the cytokines interleukin-2 (IL-2) and IL-6, orthogonal to their effect on cell proliferation.

**CONCLUSION:** By systematically analyzing the role of intercellular interactions in a synthetic microenvironment, we showed that local T cell density could modulate the balance between central memory and effector cells independent of further potential influence by antigen-presenting cells or T cell receptor

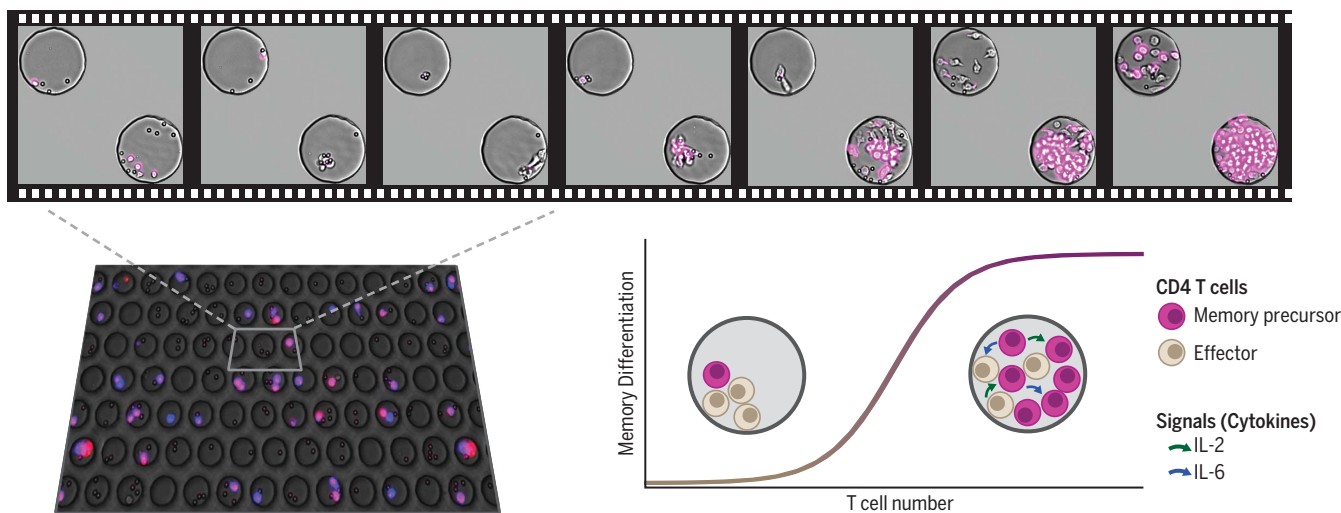
signaling strength. This cellular collectivity is a continuous process and is not determined by the number of cell divisions, but rather by the number of locally interacting cells

at any given time. Local collectivity can influence the diversity and magnitude of immune memory, by modulating interactions between T cell clones during their priming in response to antigens. Understanding the rules of T cell social behavior will be important to learn how to manipulate the immune system for therapeutic or prophylactic goals. ■

The list of author affiliations is available in the full article online.  
\*Corresponding author. Email: b.chain@ucl.ac.uk (B.C.); nir.friedman@weizmann.ac.il (N.F.)

Cite this article as M. Polonsky *et al.*, *Science* **360**, eaaj1853 (2018). DOI: 10.1126/science.aaj1853

**Collective local interactions enhance CD4<sup>+</sup> memory T cell differentiation.** The effects of intercellular interactions on T cell memory formation were studied in microwell arrays, each well holding a different number of locally interacting cells. Proliferation and differentiation were evaluated by using time-lapse movies. Differentiation into memory precursors sharply increased above a threshold number of interacting cells. This was modulated by increased sensitivity of the interacting cells to the cytokines IL-2 and IL-6.



## RESEARCH ARTICLE

## IMMUNOLOGY

# Induction of CD4 T cell memory by local cellular collectivity

Michal Polonsky<sup>1</sup>, Jacob Rimer<sup>1</sup>, Amos Kern-Perets<sup>1</sup>, Irina Zaretsky<sup>1</sup>, Stav Miller<sup>1</sup>, Chamutal Bornstein<sup>1</sup>, Eyal David<sup>1</sup>, Naama Meira Kopelman<sup>2\*</sup>, Gil Stelzer<sup>2</sup>, Ziv Porat<sup>2</sup>, Benjamin Chain<sup>3,†</sup>, Nir Friedman<sup>1,†</sup>

Cell differentiation is directed by signals driving progenitors into specialized cell types. This process can involve collective decision-making, when differentiating cells determine their lineage choice by interacting with each other. We used live-cell imaging in microwell arrays to study collective processes affecting differentiation of naïve CD4<sup>+</sup> T cells into memory precursors. We found that differentiation of precursor memory T cells sharply increases above a threshold number of locally interacting cells. These homotypic interactions involve the cytokines interleukin-2 (IL-2) and IL-6, which affect memory differentiation orthogonal to their effect on proliferation and survival. Mathematical modeling suggests that the differentiation rate is continuously modulated by the instantaneous number of locally interacting cells. This cellular collectivity can prioritize allocation of immune memory to stronger responses.

Upon recognition of a cognate antigen, naïve T cells expand and differentiate into various effector and memory cell types. The establishment of efficient acquired immune responses depends on an adequate balance between these cellular populations. Various models have been proposed to describe the mechanisms that drive T cell specialization, including cell-autonomous stochastic processes (1–3), deterministic differentiation in response to external signals (4, 5), and asymmetric cell division (5–7).

## T cell number influences memory CD4<sup>+</sup> T formation in vivo and in vitro

A number of studies have shown that the function and phenotype of CD8<sup>+</sup> T cells that expand in vivo in response to antigen stimulation depends on the number of responding T cells (8–10). Specifically, central memory T cell (TCM) differentiation is enhanced when a larger number of T cells participate in the response. We observed a similar dependency on cell number in early CD4<sup>+</sup> T cell differentiation into CD44<sup>+</sup>CD62L<sup>+</sup> cells (hereafter referred to as progenitor central memory T cells, pTCMs) in vivo, as soon as 5 days after the vaccination of mice with a cognate antigen (Fig. 1, A and B).

T cells differentiate in a complex environment in vivo, interacting with several cell types over time. Thus, we asked whether a dependence on precursor numbers can be observed in a minimal

ex vivo system, in which cellular composition and concentration, and cell-cell interactions, can be manipulated and monitored more easily. We isolated naïve splenic CD4<sup>+</sup> T cells and cultured them at increasing concentrations in vitro. These T cells were activated either by ovalbumin (OVA) peptide presented by dendritic cells (Fig. 1C, top), by microbeads coated with antibodies against CD3 and CD28 (anti-CD3 + anti-CD28) (Fig. 1C, bottom, and fig. S1, A to C), or by phorbol myristate acetate (PMA) + ionomycin (fig. S1D). Cell state was evaluated by using flow cytometry at different time points. Regardless of cell density and the mode of stimulation, the expression of CD62L decreased to its lowest level 24 hours after activation and then increased in a density-dependent manner (fig. S1B). Increasing the cell density resulted in an increased fraction of pTCMs in response to all activation regimes (Fig. 1C, fig. S1, and table S1). The maximal fraction of pTCMs was different for the various stimulations used, potentially reflecting differences in strength of activation, which have been shown to influence T cell differentiation (11). A dependency on cell density was not observed for the activation markers CD69 and interleukin-2 receptor alpha (IL2Ra) (fig. S1, B and C, and table S1). Differences in pTCM frequencies were apparent 48 hours after activation and lasted for at least 96 hours (experimental end point, Fig. 1C; fig. S1, B and C).

## pTCMs isolated from 72-hour cultures exhibit a gene expression pattern that is characteristic of established central memory T cells and persist over long time periods in vivo

To further characterize the phenotype of early differentiated pTCMs, we sorted CD62L<sup>+</sup> and CD62L<sup>−</sup> cells after 72 hours of culture. Each

sorted group was subjected to genome-wide gene expression analysis by RNA sequencing (RNA-seq). Together with CD62L (*Sell*), the expression of other T cell central memory-related genes such as *Cd27*, *Il7r*, *Ccr7*, and *Il2rb* (7, 12–16) was elevated in CD62L<sup>+</sup> cells. The expression of transcription factors (TFs) implicated in memory differentiation such as *Klf2*, *Tcf7*, *Bcl6*, *Foxo1*, and *Eomes* (17–22) was also increased (Fig. 1D). Thus, pTCMs in culture were associated with a transcriptional program that resembles that of mature memory T cells. The early expression of a transcriptional program resembling that of mature TCM cells was recently observed also in CD8<sup>+</sup> T cells, 2 to 4 days after in vivo infection (7, 23). CD62L<sup>−</sup> cells expressed high levels of the TF *Id2*, which has been associated with the inhibition of memory differentiation (24, 25). These cells also expressed higher levels of cell cycle and apoptosis-related genes (Fig. 1D) such as *Cdkn1a*, *Myc*, and *Casp3* (26, 27), whereas CD62L<sup>+</sup> cells expressed higher levels of homeostatic and self-renewal genes such as *Grp2* and *Cd4* (28). This differential gene expression is consistent with our observation that CD62L<sup>+</sup> cells exhibited a lower rate of proliferation compared with CD62L<sup>−</sup> cells in our culture system (fig. S2). Gene ontology (GO) enrichment analysis (Fig. 1E) revealed that CD62L<sup>−</sup> cells are enriched for genes related to apoptosis, ribosomal activity, and nucleotide metabolic processes, suggesting increased cell growth as well as cell death for these cells. CD62L<sup>+</sup> cells were enriched for genes related to cytokine responses, leukocyte activation, and proliferation. CD62L<sup>+</sup> cells were also enriched for genes associated with cell communication and adhesion, consistent with a role for intercellular interactions in modulating pTCM differentiation.

Initial culture densities also influenced the long-term in vivo persistence of adoptively transferred T cells. Cells activated at high density in vitro persisted in vivo and were fourfold more abundant at late time points (>35 days after in vivo transfer) compared with cells pre-cultured at low density (Fig. 1F). Transferred cells expressed large amounts of the memory markers CD62L and CD27 and low amounts of the activation marker KLRG1 (which is high on T effector cells) relative to host cells (fig. S3B). These results suggest the acquisition of an established central memory phenotype after short-term in vitro stimulation at high cell densities.

## pTCM formation is induced by local cellular collectivity

The increased differentiation of pTCMs at high cell densities can arise either from global changes in the composition of the culture medium through cytokine secretion by T cells, or by local interactions between activated cells. T cells in culture rapidly form dense dynamic clusters. Cells join and leave the clusters over time, and clusters can join to form larger clusters, or break into smaller ones (29). Thus, in conventional cell cultures as well as in vivo, it is difficult to distinguish the effect of local interactions within cell clusters

<sup>1</sup>Department of Immunology, Weizmann Institute of Science, Rehovot, Israel. <sup>2</sup>Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel. <sup>3</sup>Division of Infection and Immunity, University College London, London, UK.

\*Present address: Department of Computer Science, Holon Institute of Technology, Holon, Israel.

†Corresponding author. Email: b.chain@ucl.ac.uk (B.C.); nir.friedman@weizmann.ac.il (N.F.)



from global long-range interactions. To overcome these difficulties, we used a culture system that employs microwell arrays at the bottom of the culture plate (30). Naïve T cells were seeded within small deep microwells (diameter = 80  $\mu$ m,

depth = 120  $\mu$ m) together with microbeads coated with anti-CD3 + anti-CD28 for activation. Cell seeding was random, such that each microwell held a different number of cells at the beginning of the experiment (Fig. 2A and figs. S4 and S5A).

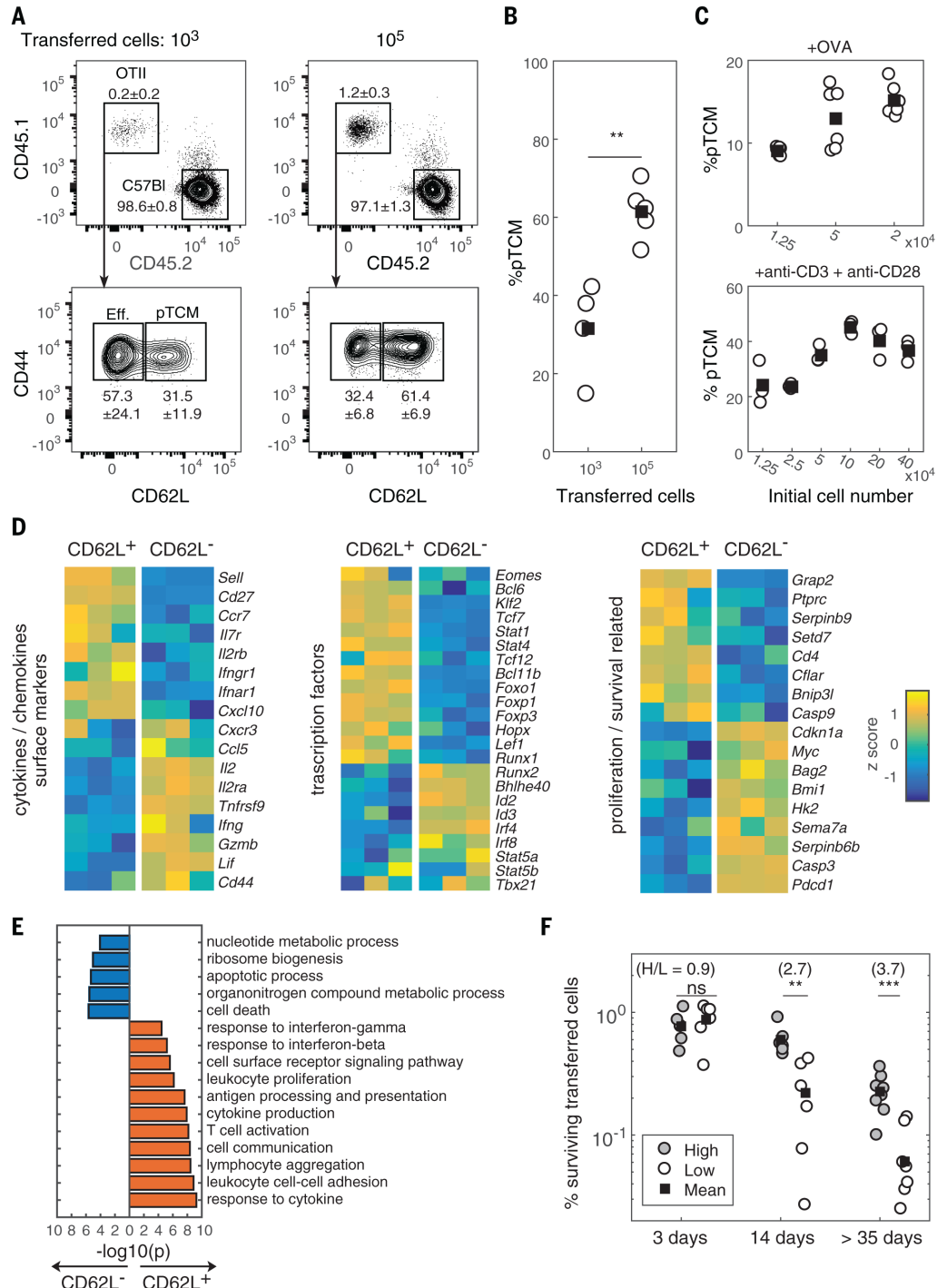
We followed the expansion and differentiation of cells within the microwells by live-cell imaging (Fig. 2, A and B). Cells were imaged for 96 hours, and the levels of expression of CD44 and CD62L were measured with live antibody stain (31)

### Fig. 1. Differentiation of pTCMs is modulated by T cell number in vivo and in vitro.

**(A)** Either  $10^3$  or  $10^5$  OT-II TCR transgenic CD4 T cells (CD45.1) were transferred into C57BL/6 recipients (CD45.2), which were then immunized with albumin protein together with an adjuvant (alum). Five days after immunization, the fractions of pTCMs (CD44<sup>+</sup>CD62L<sup>+</sup>) and effector (CD44<sup>+</sup>CD62L<sup>-</sup>) cells were evaluated in the population of transferred cells. Results are from two representative mice out of nine in one experiment. Mean  $\pm$  SD values are indicated for each population. **(B)** The percentage of pTCMs in mice injected with either  $10^3$  ( $n = 4$ ) or  $10^5$  ( $n = 5$ ) OT-II cells, 5 days after immunization. Filled squares: mean values.  $P$ -value was calculated by a two-tailed Student's  $t$  test (\*\* $P < 0.01$ ).

**(C)** The percentage of pTCMs measured at 72 hours in cultures of CD4<sup>+</sup> naïve T cells cultured ex vivo at the indicated densities and activated either with OVA-presenting dendritic cells (top,  $n = 6$  samples in one experiment) or with activation microbeads coated with anti-CD3 and anti-CD28 (bottom,  $n = 3$  samples from one representative experiment out of three). Filled squares: mean values.  $P$ -values were calculated by using 1-way analysis of variance with  $P < 0.01$  in both cases (see table S1).

**(D)** The expression levels, measured by RNA-seq, of selected gene transcripts in CD4<sup>+</sup> T cells sorted into CD62L<sup>+</sup> and CD62L<sup>-</sup> populations after 72 hours of culture. Cells were cultured at an initial cell number of  $2.5 \times 10^5$  cells per well and activated with anti-CD3- and anti-CD28-coated microbeads. **(E)** Gene ontology (GO) enrichment analysis of differentially expressed genes in the CD62L<sup>+</sup> and CD62L<sup>-</sup> subpopulations. Differentially expressed genes ( $P \leq 0.05$ , Benjamini-Hochberg correction), with  $\log_2$  fold change of  $\geq 0.5$  between CD62L<sup>+</sup> and CD62L<sup>-</sup> samples were chosen for the analysis. Results are from  $n = 3$  repeats. **(F)** Naïve CD45.1 CD4<sup>+</sup> T cells were cultured ex vivo for 72 hours in either high ( $2 \times 10^5$  cells/ml) or low ( $6.25 \times 10^4$  cells/ml) initial densities, activated with anti-CD3- and anti-CD28-coated microbeads, and then transferred into CD45.2 recipients (total of 40 mice in two experiments). Spleens were harvested from recipient mice after 3, 14, 35, or



(Fig. 2, A and B, and movie S1). For each microwell, we extracted time traces of CD44 and CD62L levels, and the area covered by cells, which linearly correlated with cell number (fig. S5B).

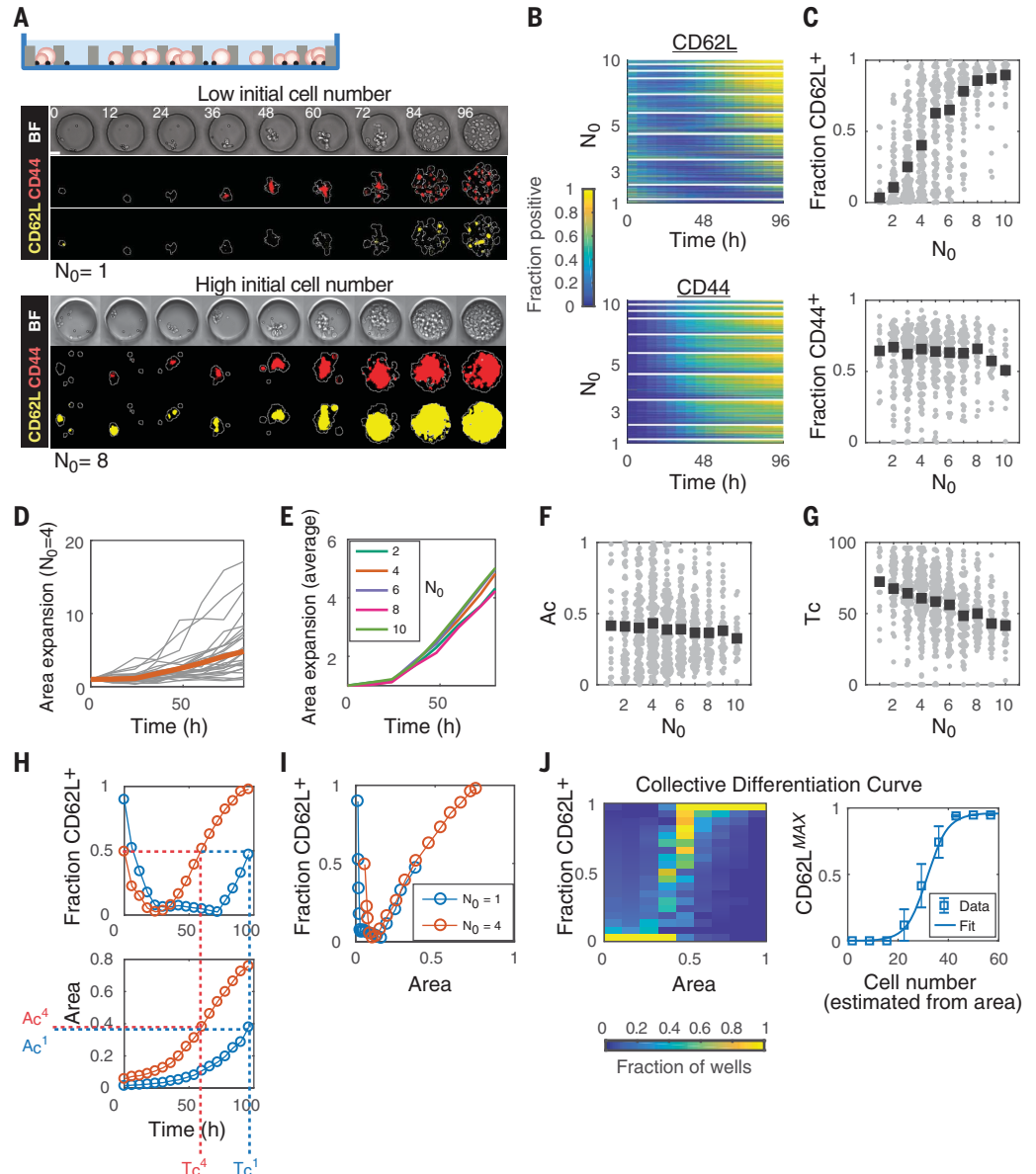
The average CD62L and CD44 expression dynamics in the microwells closely resembled those observed by flow cytometry of bulk cultures (fig. S5C). We found that the expression of CD62L,

but not of CD44, was dependent on the initial number of T cells in the microwell,  $N_0$ . In wells initially containing one or two cells, CD62L was expressed only in a small fraction of the cells at late time points ( $t = 72$  to 96 hours), whereas in microwells initially containing more than seven cells, CD62L was expressed by most cells at these times (Fig. 2, B and C). Moreover, the reexpres-

sion of CD62L occurred earlier in microwells with a high initial cell number (Fig. 2B). The level and timing of CD44 expression however, did not depend on the initial cell number (Fig. 2, B and C). Higher cell densities enhanced differentiation toward pTCMs in both microwells and conventional culture. However, the microwell culture established that this effect

## Fig. 2. Induction of CD62L expression depends on local interactions between differentiating T cells.

An array of microwells was placed on the bottom of an optical 96-well plate. Naïve CD4 T cells and activation microbeads were seeded such that the microwells randomly received a different initial number of cells ( $N_0$ ). Cells in different microwells shared the same culture medium. (A) Top: Schematic of the experiment, showing the microwell array in side view. T cells and activation microbeads (black dots) are shown. Representative image strips show microwells with  $N_0 = 1$  (top three rows) and  $N_0 = 8$  (bottom three rows). Cells were imaged for 96 hours with bright-field illumination (BF) and two fluorescent channels (CD44, red; CD62L, yellow). Numbers on top show time in hours. Scale bar, 20  $\mu\text{m}$ . (B) Time traces of all microwells in one representative experiment out of three ( $n = 674$  microwells) showing the fraction of CD62L<sup>+</sup> (top) and CD44<sup>+</sup> (bottom) in each microwell over 96 hours. Microwells are grouped by  $N_0$  ( $1 \leq N_0 \leq 10$  initial cells). The color code represents the fraction of positive cells in each microwell. (C) Fraction of CD62L<sup>+</sup> (top) and CD44<sup>+</sup> (bottom) cells after 72 hours of culture, plotted versus  $N_0$ . Each gray dot represents one microwell; black squares represent median values. Data are from the same experiment as in (B). (D) Representative traces of the area covered by cells, from microwells starting with  $N_0 = 4$  cells. The average trace is depicted in color. All traces are normalized by their initial area. (E) Average area traces for microwells starting with different  $N_0$  values. Each trace is normalized by its initial cell area. Data are combined from three experiments ( $n = 1734$  microwells). (F) The critical area ( $A_c$ ), defined as the area in which 50% of the cells are differentiated (CD62L<sup>+</sup>), plotted versus  $N_0$  for all microwells in three experiments as in (E). (G) Critical time ( $T_c$ ), defined as the time in which 50% of the cells have differentiated, plotted versus  $N_0$  as in (F). (H) Traces of two representative microwells starting with one (blue) or four (red) initial cells.



Top: fraction of CD62L<sup>+</sup> over time; bottom: cell area over time.  $T_c$  and  $A_c$  are indicated by the dotted lines. (I) Fraction of CD62L<sup>+</sup> plotted as a function of cell area for the two microwells shown in (H). (J) Derivation of the collective differentiation curve (CDC). Left: The frequency distribution of all microwells from all time points in one representative experiment plotted on the area-CD62L<sup>+</sup> plane. Right: The CDC is defined by the maxima of the heatmap on the left, binned by cell area. Data for CDC derivation were averaged over eight individual experiments (blue squares). Error bars show SEM. The obtained CDC was fitted with a logistic curve (line) of the form  $F(N) = M / (1 + e^{-(N - N_c)})$ ;  $R^2 = 0.99$ .



was local and was restricted to cells within individual microwells. Indeed, cells in neighboring microwells, which share culture medium, can have different differentiation outcomes based on their initial cell number, despite their proximity (fig. S4). CD62L expression did not depend on the number of activating microbeads (fig. S5D and table S1), precluding competition for a limited amount of stimulatory signals as a cause for enhanced pTCM formation. Thus, the observed collective behavior was driven by short-range interactions that modulated the differentiation of cells within the same microwell.

### Collective pTCM differentiation is a continuous process depending on the instantaneous number of interacting T cells

On the basis of these results, we then sought to assess the mechanisms that drive collective differentiation. In particular, we asked whether differentiation depends on the number of cell divisions, on the initial number of interacting cells, or on the varying number of cells that interact (which grows with time because of cell proliferation). This analysis is typically complicated by the fact that the acquisition of a pTCM cell state occurs in parallel with cell proliferation. Thus, it is difficult to distinguish direct mechanisms that affect the differentiation process itself and are not mediated indirectly by their effect on cell proliferation. We found that the number of cell divisions was not a major factor in regulating

pTCM formation, as cells cultured in high cell density had higher CD62L expression compared with cells that had the same number of divisions but came from a culture of low cell density (fig. S6). The microwell assay further demonstrated that proliferation did not display the collective characteristic that we observed for CD62L induction. Although the expansion dynamics of cells in individual microwells were highly variable even when starting from the same initial cell number (Fig. 2D), the average proliferation rate was independent of  $N_0$  (Fig. 2E and table S3).

Further insight was gained by examining the relationship between cell area and differentiation within individual microwells. We defined a critical time ( $T_c$ ) at which 50% of the cells in a microwell were differentiated (CD62L<sup>+</sup>), and a critical area ( $A_c$ ), which is the total cell area at  $T_c$  (Fig. 2, F to H). When comparing traces of two representative microwells starting with one (blue) or four (red) initial cells, we found that  $A_c$  was almost identical in both cases, whereas  $T_c$  was much higher in microwells that started with one cell (Fig. 2H). Examining all microwells ( $1 \leq N_0 \leq 10$ ), we found that the average  $A_c$  was independent of initial cell number (Fig. 2F), whereas  $T_c$  decreased as initial cell numbers increased (Fig. 2G). Distinctive behavior was observed for CD44, for which  $T_c$  was constant, and  $A_c$  increased with  $N_0$  (fig. S11A).

When the fraction of CD62L<sup>+</sup> cells was plotted as a function of total cell area in individual microwells, we found that both traces collapsed onto the same trajectory (Fig. 2I). This suggests

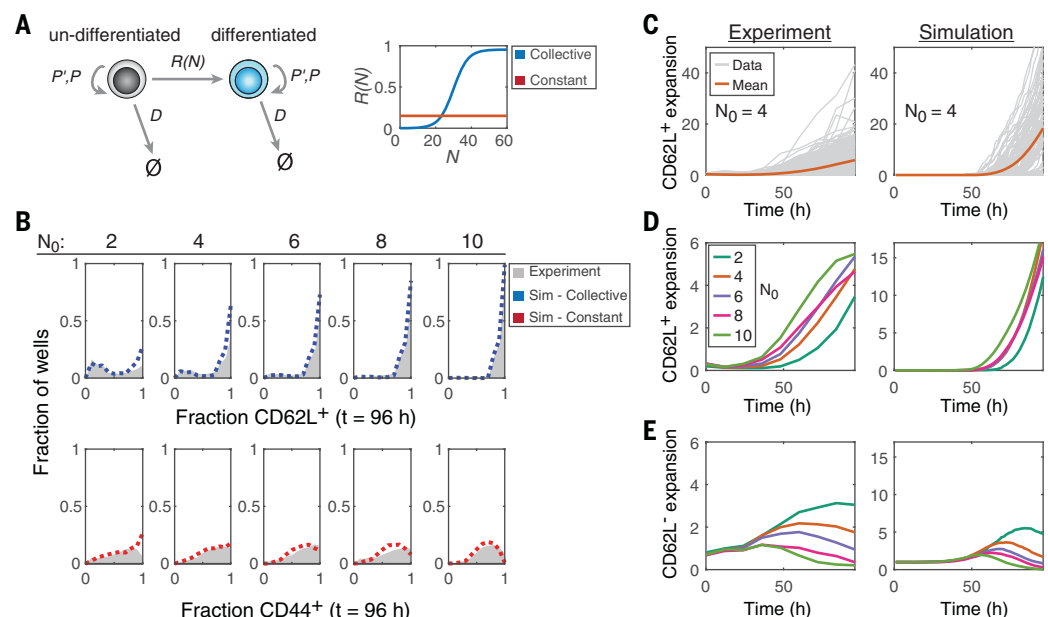
that the level of differentiation did not depend on time since activation or on the extent of cell proliferation, but rather on the instantaneous number of interacting cells. We generalized this observation by projecting all data points of all microwells (regardless of the number of initial cells, time from cell activation, or amount of cell proliferation) onto the area-CD62L plane (Fig. 2J, left). The normalized heatmap showed that cells in individual microwells tended to follow a universal trajectory in the area-CD62L plane. We represented this trajectory with a collective differentiation curve (CDC), which we defined by the local maxima of the heatmap. Thus, the CDC describes the most commonly observed fraction of differentiated cells for a given number of cells in a microwell. The CDC, averaged over eight individual experiments, is plotted in Fig. 2J, right. For further analysis, the CDC was fitted by a logistic function of the form  $F(N) = M / (1 + e^{(-r(N-N_c))})$ , where  $N$  is the number of cells in the microwell (which can change with time),  $N_c$  is a critical cell number above which differentiation is more efficient,  $M$  is the maximal differentiation fraction, and  $r$  defines the steepness of the curve.

The existence of a universal differentiation trajectory, which describes all microwells regardless of  $N_0$ , was consistent with the above observation that the mean  $A_c$  does not depend on  $N_0$ , whereas  $T_c$  decreases with  $N_0$  (Fig. 2, F and G, and table S1). Additional support for a universal differentiation trajectory came from global gene expression analyses performed at different time

**Fig. 3. CD62L expression dynamics can be simulated by assuming collective differentiation. (A)** A stochastic agent-based model of the differentiation process. Cells can proliferate, differentiate, and die.

The proliferation rates ( $P'$ , first division;  $P$ , subsequent divisions) and death rate ( $D$ ) were experimentally obtained from single-cell data, as detailed in the supplementary text. We assume in the model identical rates for undifferentiated and differentiated cells. The differentiation rate,  $R$ , can be either constant ( $R_{\text{constant}}$ , red line) or have a logistic dependence on the number of cells in a microwell ( $R_{\text{collective}}$ , blue curve),  $R(N) = M / (1 + e^{(-r(N-N_c))})$ .

**(B)** Distributions of the fraction of CD62L<sup>+</sup> (top) and CD44<sup>+</sup> (bottom) cells after 96 hours for microwells starting with varying  $N_0$  values. Experimental data (gray) are compared to simulations assuming  $R_{\text{constant}}$  (red) or  $R_{\text{collective}}$  (blue). Data are combined from three experiments as in Fig. 2. The parameters used for  $R_{\text{collective}}$  were obtained from fitting the CDC of Fig. 2J ( $M = 0.95$ ,  $r = 0.25$ , and  $N_c = 30$ ).  $R_{\text{constant}} = 0.15$  gave the best fit to the CD44<sup>+</sup> data. **(C)** The relative expansion of CD62L<sup>+</sup> cells over time, showing experimentally obtained (left) and simulated (right) traces,



of microwells starting with  $N_0 = 4$  cells. The colored trace shows mean values. Data are from the same experiments as in (B). **(D)** The average expansion of CD62L<sup>+</sup> cells for experimentally obtained (left) or simulated (right) data, for microwells with different values of  $N_0$ . **(E)** Same as in (D), showing average expansion of CD62L<sup>+</sup> cells. Traces in (C) to (E) are normalized by their initial value.

points. Cells that grow at low densities followed similar changes in their gene expression pattern but with a delay compared with cells that grow at high cell densities (fig. S7). Together, these findings provide a strong indication for a collective process in which the acquisition of a pTCM state depends on the instantaneous number of interacting cells.

### Stochastic simulation recapitulates experimentally observed collective differentiation

To gain a better understanding of the cellular processes giving rise to the observed CDC, we constructed a stochastic computational model simulating a transition from an undifferentiated to a differentiated state (Fig. 3A). The model simulates microwells that initially start with  $N_0$  undifferentiated cells. Cells stochastically proliferate and die at experimentally obtained rates (32) (see figs. S8 and S9 and supplementary text for a detailed explanation of model construction). We assumed in the model that division and death rates were identical for differentiated and undifferentiated cells, as delaying the division time of differentiated cells (consistent with the slower division observed for CD62L<sup>+</sup> cells; fig. S2) (33) did not have a pronounced effect on the simulation outcome (fig. S10). Cells in the model differentiate by some rate  $R$ , which can be either constant ( $R = \text{constant}$ , red line in Fig. 3A) or collective [i.e., dependent on  $N$ , the instantaneous number of cells in the microwell ( $R = \text{collective}$ , blue line in Fig. 3A)]. As we showed that differentiation in our system depended on cell number, we assumed that  $R_{\text{collective}}(N)$  had the same functional form as the CDC and used a logistic curve to describe it in the model. The three parameters that describe  $R_{\text{collective}}(N)$  (namely  $M$ ,  $r$ , and  $N_c$ ) were thus extracted from the fit to the experimental CDC (Fig. 2J). We scanned the parameters of the logistic curve and found that the experimentally derived parameters that we used for the simulations were inside a broad optimal region in parameter space (fig. S9, A and B), justifying their use without fitting. This allowed us to run simulations with all model parameters obtained directly from experimental observations.

We then calculated the simulated fraction of differentiated cells at 96 hours for different values of  $N_0$ , for both differentiation regimes (constant and collective), and compared them with the measured distributions of CD62L<sup>+</sup> and CD44<sup>+</sup> cells (Fig. 3B). Simulations assuming collective differentiation fit the CD62L<sup>+</sup> data well over all values of  $N_0$  (Fig. 3B, top), much better than simulations of a model based on a constant differentiation rate (fig. S9C). A model assuming a constant differentiation rate, however, fit the CD44<sup>+</sup> data (Fig. 3B, bottom). Of note,  $T_c$  and  $A_c$  obtained by the simulation resembled the experimentally derived values, with the collective and constant differentiation rates showing behavior similar to that of CD62L and CD44, respectively (fig. S11). The stochastic simulation with collective differentiation also captured the well-to-well variability of the experimental

data (Fig. 3C), as well as the experimentally observed dynamic changes in the average numbers of both CD62L<sup>+</sup> and CD62L<sup>-</sup> cells, and their dependence on  $N_0$  (Fig. 3, D and E, and table S3). A model in which cell differentiation depends on the number of neighbors only at the beginning of the experiment, rather than continuously changing with cell number, is less consistent with our data (fig. S11, D and E). We verified experimentally that differentiation remains plastic by transferring T cells into microwells at varying times after their activation in bulk culture and showed that the differentiation outcome depended on the new number of neighbors (fig. S12). Thus, a stochastic model for cell differentiation can describe the experimentally observed collectivity if the differentiation rate  $R$  depends on the instantaneous number of interacting cells as described by the logistic CDC, with differentiation markedly increasing above  $N_c \sim 30$  interacting T cells.

### The cytokines IL-2 and IL-6 and the transmembrane protein SLAMF6 modulate collective differentiation

Finally, we sought to identify candidate signaling molecules and pathways that facilitated the observed local collectivity. Short-range interactions between T cells within the same microwell can be mediated by cell-surface ligands and their receptors, as well as by secreted cytokines, which accumulate at high concentrations in the vicinity of the cells and sharply decline with distance (34, 35). Thus, we repeated our microwell array experiments, while adding antibodies to block specific cytokines or surface molecules, or using cells from knockout mouse strains that lacked relevant genes. Out of the several candidate cytokines and genes that we tested, three showed a significant effect on CD62L expression (fig. S13 and table S2). The inhibition of the cytokines interleukin-2 (IL-2) and IL-6 with blocking antibodies reduced CD62L expression after 96 hours of culture (Fig. 4A), whereas the absence of the cell-surface molecule SLAMF6 (*Slamf6*<sup>-/-</sup> cells) enhanced its expression (Fig. 4B). We note that blocking IL-2 strongly reduced cell numbers in the microwells at late time points (fig. S14A). To overcome this difficulty, we added IL-6 to increase cell viability (Fig. 4C and fig. S14A) (29), without altering CD62L expression (Fig. 4, A and C). Adding back external human IL-2 increased CD62L expression in a concentration-dependent manner (figs. S13 and S14).

Further evidence for the involvement of IL-2 and IL-6 in driving collective differentiation was revealed from investigation of their signaling pathways. We found that in clustering cells, the receptor subunits IL-2Ra and IL-6Rst were non-uniformly distributed on the cell surface and displayed patches that were typically directed toward neighboring cells within the cluster (Fig. 4D and fig. S15, B and C). IL-2 is expressed and secreted by activated T cells at early time points (fig. S15A) (36) and is directed toward T-T synapses (37). We verified that IL-6 was also produced by T cells in our cultures early after activation

(fig. S15A), although we could not infer its localization because of low signal. The polarization of cytokine receptors toward neighboring cells, together with accumulation of cytokines within T cell clusters (37), may lead to increased signaling capacity. Supporting this hypothesis, we observed that the phosphorylation of signal transducer and activator of transcription 5 (STAT5) and STAT3 in response to IL-2 or IL-6, respectively, was significantly higher when cells were cultured at high density, thus forming more clusters (Fig. 4E and fig. S16). Finally, we observed that Janus kinase (JAK) inhibitors, which block STAT signaling downstream of IL-2 and IL-6, reduced CD62L expression and pTCM formation, whereas phosphatidylinositol 3-kinase (PI3K) pathway inhibition, which is also activated by IL-2 signaling, did not (Fig. 4F).

### IL-2 and IL-6 modulate collective pTCM formation distinctly and orthogonal to their effect on cell proliferation

The CDC derived for each of these perturbations provided a faithful and compact description of the differentiation trajectory, allowing us to assess the net effect on collective cell differentiation and survival. The results recapitulated our previous observations, as blocking IL-2 or IL-6 resulted in reduced CD62L<sup>MAX</sup> values, whereas in the absence of SLAMF6, CD62L<sup>MAX</sup> was increased around  $N_c$  (Fig. 5A and table S2). The CDC further showed that when IL-2 was blocked but the culture was supplemented with IL-6, CD62L expression remained low even in microwells in which a substantial number of cells had accumulated ( $N > N_c$ ) (Fig. 4C and Fig. 5A, left). Addition of human IL-2 restored the CDC to its unperturbed form in a concentration-dependent manner (fig. S13B and table S2).

The different perturbations can be described by their effect on the parameters of the logistic curve that fits the perturbed CDC. Blocking IL-2 reduced the maximum responsiveness (given by the parameter  $M$ ), which is consistent with IL-2 regulating the probability of a cell to differentiate at a given number of neighbors (Fig. 5A, left). By contrast, blocking IL-6 only marginally decreased  $M$  but shifted the CDC curve to higher cell numbers (Fig. 5A, middle). This is consistent with IL-6 playing a role in decreasing  $N_c$ , the critical number of interacting cells that promotes differentiation. The surface molecule SLAMF6 had the opposite effect, as its absence somewhat reduced  $N_c$  but did not change  $M$  (Fig. 5A, right). We cannot preclude an additional effect of IL-2 also on  $N_c$  with the current data. The effects of these perturbations on the CDC can be captured by the stochastic computational model. Changing only the parameter  $M$  in  $R_{\text{collective}}$  resulted in simulated trajectories that resemble those obtained by blocking IL-2 and also captured the gradual recovery of differentiation that is observed when adding back external IL-2 (Fig. 5B, left, and Fig. 5C, bottom). Simulating reduced proliferation (which is caused by IL-2 blockade) cannot describe the experimentally observed distributions



or the CDC if  $M$  is unchanged (Fig. 5, C and D). The experimentally observed behavior for anti-IL-6 and SLAMP6 knockout conditions can be described by changing the parameter  $N_c$  in the model (Fig. 5B, middle and right panels). These results demonstrate that the three factors affect collective CD62L expression, orthogonal to their effect on cell proliferation and survival: IL-2 regulates the maximal differentiation rate, whereas IL-6 and SLAMP6 tune the critical number of cells required for differentiation.

## Discussion

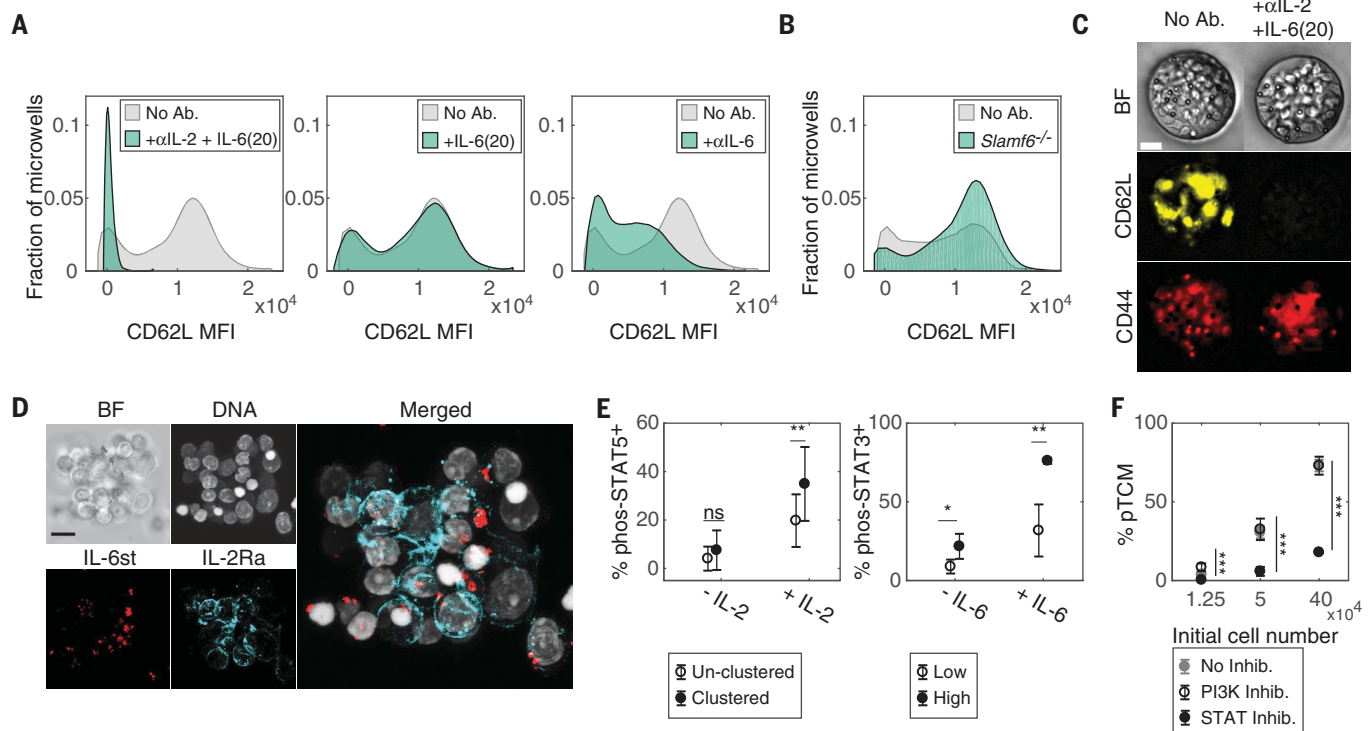
In this work, we systematically analyzed the role that intercellular interactions between CD4 T cells play in central T cell memory formation. By activating T cells in a synthetic microenvironment,

we showed that local cell density can modulate the balance between CD62L<sup>+</sup> and CD62L<sup>-</sup> cells, independent of further potential influence by antigen-presenting cells or T cell receptor (TCR) signaling strength. Using this system, we could determine that differentiation into memory precursors is most efficient at  $N > \sim 30$  interacting cells and that this collective property depends on the instantaneous number of interacting cells, rather than on the number of division cycles the cells undergo.

The signals that direct the development of T cell memory have been studied extensively, and several models of T cell diversification have been suggested (31, 38). Hence, TCR stimulation strength (39) and duration (40), as well as signaling by various cytokines (11), have been shown

to modulate the generation and maintenance of memory T cells. Our results support the notion that local, short-range interactions between T cells early after TCR stimulation serve as another potent modulator of memory induction. Depending on the experimental model, local collectivity may influence differentiation to memory or effector phenotypes, depending on the number of precursor cells that participate in the response and their extent of proliferation.

Our findings further suggest that the cytokines IL-2 and IL-6, which are expressed by T cells just a few hours after TCR stimulation, are mediators of local collectivity. The surface protein SLAMP6 also affected collective memory formation, although to a lower extent. Both IL-2 (41) and IL-6 (42) have been previously shown



**Fig. 4. IL-2, IL-6, and SLAMP6 modulate CD62L expression.** (A and B) Histograms showing the mean expression levels of CD62L at  $t = 96$  hours [given as mean fluorescent intensity (MFI)] in all microwells from two independent experiments, comparing different perturbations (shaded in green) to the control ("No Ab.," shaded in gray). (A) Left: microwells supplemented with 10 μg/ml of anti-IL-2, together with 20 μg/ml of IL-6, which was added to increase cell viability ["+αIL-2 + IL-6(20),"  $n = 789$ ]. Middle: microwells supplemented with 20 ng/ml of IL-6 ["+αIL-6(20),"  $n = 863$ ]. Right: microwells supplemented with 10 μg/ml anti-IL-6 ["+αIL-6,"  $n = 1605$ ]. All three plots show the control sample from the same experiment ( $n = 812$ ). (B) MFI of microwells harboring *Slamp6*<sup>-/-</sup> cells ( $n = 686$ ) compared to control cells ( $n = 645$ ). (C) Representative images of control microwells ("No Ab.") or microwells treated with anti-IL-2 (10 μg/ml) supplemented with IL-6 (20 ng/ml). Images were taken at  $t = 54$  hours of culture. Images show bright-field illumination (BF, top), CD62L (middle), and CD44 (bottom). Scale bar, 20 μm. (D) Confocal images of a T cell cluster after 24 hours of activation by using PMA + ionomycin, showing IL-6 receptor (IL-6st), IL-2Ra (CD25), and nuclei (DNA). Images are maximum projection over the total z stack. Scale bar,

20 μm. (E) IL-2 and IL-6 downstream signaling of clustered (filled circles) and unclustered (open circles) cells (for phos-STAT5, see also fig. S16), or for cells cultured in dense (filled circles) or sparse (empty circles) cultures (for phos-STAT3). Percentage of phos-pSTAT5<sup>+</sup> cells (left) and phos-STAT3<sup>+</sup> cells (right) was measured by flow cytometry after 24 hours of culture followed by a 10-min pulse of 5 ng/ml of IL-2 (left) or 10 ng/ml of IL-6 (right). IL-2 data are pooled from three individual experiments with a total of  $n = 13$  samples for each condition. IL-6 data are from one experiment with  $n = 4$  repeats for each condition. (F) Small-molecule inhibition of the JAK-STAT and the PI3K pathways. Cells were cultured at the indicated initial cell numbers with either PI3K (empty circles) or JAK-STAT (black circles) small-molecule inhibitors, or without inhibition (gray circles). The percentage of pTCMs was measured by flow cytometry 72 hours after activation. Data were averaged from  $n = 3$  to 4 wells for each condition, from one representative experiment out of three. In (E) and (F), mean ± SD are shown.  $P$ -values were calculated by using a two-tailed Student's  $t$  test (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ).  $P$ -values in (F) compare STAT inhibition to the control. ns, not significant.

to promote T cell memory differentiation. We found that IL-2 is required for collective pTCM generation at the early stage of the response (< 3 days). IL-2 may have other effects at later time points—for example, at the peak of in vivo responses (43, 44), or during the contraction phase and maintenance of memory cells (43), which,

together with other cytokines such as IL-7 and IL-15, can further modulate the long-term magnitude of the memory response. We also provided data that suggest that IL-2 and IL-6 contribute to collectivity in the generation of pTCMs at least in part by increased sensitivity of clustered T cells to IL-2 and IL-6. Notably,

IL-6 (42), as well as members of the SLAMF family of surface receptors (45), have been shown to enhance the IL-2 sensitivity of CD4 T cells. Antibodies blocking the adhesion molecule LFA-1 disrupt T cell clustering and affect memory formation in vivo (46), but we did not observe a significant effect in our cultures (fig. S13 and table S2). This may stem from different physical constraints or from interactions of T cells with other cells in vivo, which were lacking in our ex vivo cultures. The formation of T cell memory is a highly complex process, and we expect that other molecular components, not yet identified, may contribute to the phenomenon of density-dependent cellular cooperation that we have described.

We hypothesize that the property of increased memory formation above a threshold number of locally interacting cells can have a functional role, as it may prioritize the allocation of immune memory to insults that result in large responses, while preventing aberrant memory of potentially less relevant small events. Local collectivity can also affect the diversity of immune memory, by tuning interactions between T cell clones of different TCR specificities that interact during priming—for example, through clustering on the same antigen-presenting cell. Understanding the rules of T cell social behavior will be important to learn how to manipulate the immune system for therapeutic or prophylactic goals.

## Materials and methods

### Mice

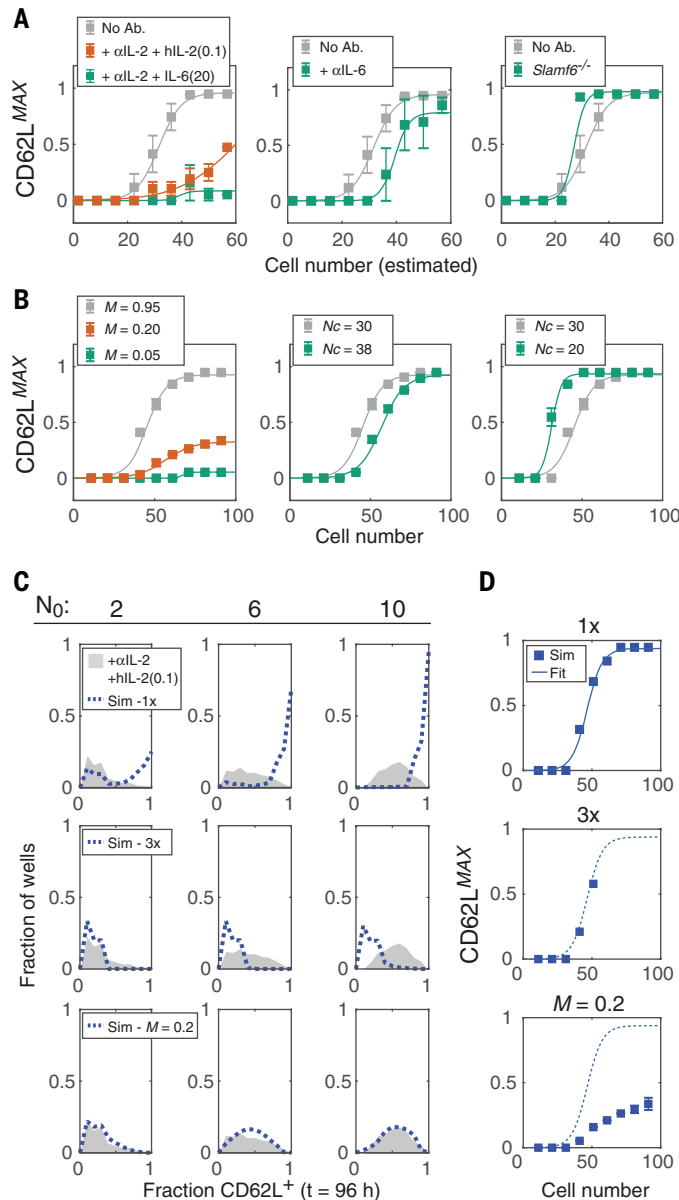
C57BL/6, B6SJL, and TCR-transgenic OT-II mice [harboring ovalbumin (OVA)-specific CD4<sup>+</sup> T cells] were housed under specific pathogen-free conditions at the animal facility of Weizmann Institute and were used at 6 to 8 weeks of age. SLAMF6-deficient mice (*Slamf6*<sup>-/-</sup>) (47) were a donation from the laboratory of I. Shachar from the Weizmann Institute of Science. All animal experiments were performed under protocols approved by the Animal Care and Use Committee of the Weizmann Institute.

### Adoptive transfer

For the experiment detailed in Fig. 1A: At day -1, naïve OT-II cells (expressing anti-CD45.1 on a C57BL/6 background) were injected into C57BL/6 recipients. Either 10<sup>3</sup> or 10<sup>5</sup> cells per mouse were injected intravenously (i.v.). On day zero, mice were immunized intraperitoneally (i.p.) with 100 µl of phosphate-buffered saline (PBS) containing 50 µg of albumin protein (Sigma-Aldrich) together with alum as an adjuvant (diluted 1:3, ThermoFisher Scientific). On day 5, total cells were isolated from recipient spleens and stained with APC-Cy7-labeled anti-CD45.1, Alexa Fluor 488-labeled anti-CD45.2, PE-labeled anti-CD3, and PerCP-Cy5.5-labeled anti-CD4 to assess the fraction of donor cells out of recipient CD4<sup>+</sup> T cells. Cells were also stained with APC anti-CD62L and PE-Cy7-labeled anti-CD44 to analyze their differentiated state.

For the experiment detailed in Fig. 1F: Naïve CD4<sup>+</sup> T cells were isolated from B6SJL mice

**Fig. 5. CDC analysis and stochastic simulations show distinct effects of IL-2 and IL-6 on collective differentiation, orthogonal to their effect on T cell proliferation (A)** CDC plots for the experimental perturbations shown in Fig. 4, compared to control ("No Ab," gray, averaged over eight individual experiments). The experimental CDCs (squares) were obtained by averaging data from all experiments for each condition (as given below) and were fitted with a logistic function (line). Left: addition of 10 µg/ml of anti-IL-2 supplemented with 0.1 ng/ml human-IL-2 ["+αIL-2 + hIL-2(0.1)," orange, *n* = 3 experiments]. Middle: addition of anti-IL-6 [10 µg/ml, "αIL-6," green, *n* = 4 experiments]. Right: *Slamf6*<sup>-/-</sup> cells [green, *n* = 2 experiments]. (B) CDCs obtained by model simulations using modified *R*<sub>collective</sub> parameters. Control CDC (gray) was constructed with a simulation assuming



collective differentiation with nominal values as in Fig. 3 (*M* = 0.95, *N<sub>c</sub>* = 30). Modified parameters that were used are: left: changing the maximal differentiation rate: *M* = 0.2 (orange) or 0.05 (green); middle: changing the critical cell number: *N<sub>c</sub>* = 38; right: *N<sub>c</sub>* = 20. (C) Distributions of the fraction of CD62L<sup>+</sup> cells after 96 hours, for microwells starting with varying *N<sub>0</sub>* values and supplemented with 10 µg/ml of anti-IL-2 and 0.1 ng/ml of hIL-2, as in (A), left. Experimental data (gray) were compared to simulations assuming collective differentiation as in Fig. 3B. As inhibition of IL-2 also inhibits proliferation, the data were compared to a simulation in which the division time of the cells was extended. Top: Division time is unchanged ("1x"); middle: division is three times slower ("3x"); bottom: division time is unchanged and *M* is reduced (*M* = 0.2). (D) CDC curves of simulation results (squares) for the corresponding conditions in (C). Simulation results with the nominal parameter values (1x, as in Fig. 3) were fitted with a logistic curve, which is shown in all three graphs as a reference.



(expressing CD45.1) and cultured in low and high initial concentrations ( $6.25 \times 10^4$  and  $2 \times 10^6$  cells/ml, respectively). Cells were activated using anti-CD3- and anti-CD28-coated microbeads and cultured in six-well plates for 72 hours. At the end of the culture period, dead cells were removed on a Ficoll gradient and either  $1 \times 10^6$  or  $2 \times 10^6$  donor cells (depending on the experiment) were injected i.v. into C57BL/6 recipients. At three time points (3, 4, and >35 days), total cells were isolated from recipient spleens. Cells were stained with APC-Cy7-labeled anti-CD45.1, Alexa Fluor 488-labeled anti-CD45.2, PE-labeled anti-CD3, and Pacific Blue-labeled anti-CD4 to assess the fraction of donor cells out of recipient CD4<sup>+</sup> T cells. Cells were also stained with APC-labeled anti-CD62L, PE-Cy7-labeled anti-CD44, PerCP-Cy5.5-labeled anti-KLRG1, and Brilliant Violet 510-labeled anti-CD27 to analyze their differentiation state. For antibody specifications, see table S4.

### Bulk cell culture

Naïve CD4<sup>+</sup> T cells were purified from mice spleens using magnetic microbeads separation (CD4<sup>+</sup>CD62L<sup>+</sup> MACS T-cell Isolation Kit, Miltenyi Biotec). Unless stated otherwise, cells were cultured in 96-well plates in 200 µl of RPMI. Cells were seeded at varying concentrations indicated in the main text, and activated either with anti-CD3 and anti-CD28-coated microbeads at a 1:1 bead:cell ratio (Dynabeads, Invitrogen) or in the presence of 2 µl/ml cell activation cocktail (PMA + ionomycin, Biologend). For the experiment presented in Fig. 1C (top), naïve OT-II cells were used and activated using 10 µg/ml OVA peptide (ISQAVHAAHAEINEAGR, InvivoGen) presented by preloaded dendritic cells (1:1 OTII: DC ratio). Cells were cultured in complete RPMI 1640 medium with phenol red, supplemented with 10% (vol/vol) fetal calf serum (FCS), 100 U/ml of penicillin, 100 mg/ml of streptomycin, 2 mM glutamine, 10 mM HEPES, 1 mM sodium pyruvate, and 50 mM β-mercaptoethanol, all from Biological Industries (Beit Haemek, Israel). When cultured for flow cytometry measurements, cells were stained with the proliferation dye eFluor 450 (ebioscience) before the start of culture; after isolation, cells were supplemented with 1 ml of PBS + 1 µl of eFluor-450, incubated for 10 min at 37°C and then washed three times with RPMI.

### Extracellular markers flow cytometry sample preparation

Cells were harvested at the different time points indicated in the text (typically at  $t = 0, 24, 48$ , and 72 hours), placed in a U-shaped 96-well plate, and washed twice with PBS. The supernatant was aspirated and cells were supplemented with live-dead blue reagent (Invitrogen) diluted 1:1000 in PBS, to a final volume of 20 µl. Samples were then supplemented with 5 µl of antibody mix: PE-Cy7-labeled anti-CD62L, APC-Cy7-labeled anti-CD44, FITC-labeled anti-CD4, PE-labeled anti-CD69, and APC-labeled anti-IL2Ra (0.25 µl/sample for each antibody), to a final volume of 25 µl (for antibody specifications, see table S4). Samples

were then incubated in the dark for 30 min at room temperature, washed twice with PBS, and measured using a flow cytometer (LSRII; BD).

### Microwell array design and fabrication

Photolithography masks were designed using autoCAD (Autodesk, San Rafael, CA). Molds were designed as an array of several hundred hexagonally spaced microwells, each being 80 µm in diameter and 120 µm in depth. This depth is 12 to 20 times the cell diameter, thus reducing the chance of escape by activated cells. Microwell preparation is described in detail in Zaretsky *et al.* (30). Briefly, molds were fabricated using photolithography of negative photoresist on silicon wafers. In a slight alteration from the cited protocol, photoresist spinning was performed twice yielding feature heights of 120 µm. Wafers were exposed to ultraviolet (UV) irradiation on a contact mask aligner using a dark-field mask, hardened, and treated to remove unbound photoresist. PDMS was mixed in 10:1 weight ratio (base:curing agent). One milliliter was then poured onto the template wafer and centrifuged (WS-650S spin processor, Laurell Technologies) for 30 s at 300 rpm followed by 2 min at 1000 rpm. Wafers were then left to stand for 10 min to even the PDMS surface and then baked for 1 hour at 80°C until PDMS was fully cured. After curing, the thin PDMS layer was cut into stripes, peeled, and gently placed on strips of a thick PDMS slice. This thick PDMS was used to help place the thin layer containing the array into a 96-well plate. Small squares (~5 mm by 5 mm) containing the microwell pattern were cut and punched into an optical bottom 96-well plate (Thermo Fisher Scientific, Rochester, NY) using tweezers. As the thin layer adhered strongly to the bottom of the plate, it detached from the thick layer and remained inside the 96-well plate.

### Cell loading and culture in microwell arrays

For all microscopy experiments, naïve CD4<sup>+</sup> T cells were purified from mice spleens using magnetic microbead negative separation (Naïve CD4<sup>+</sup> isolation, StemCell Technologies). Cells were cultured in medium identical to that used for bulk culture, but without phenol red. For all microscopy experiments, activating microbeads were used at a 2:1 bead:cell ratio. To facilitate cell loading into the small-volume microwells and eliminate trapped air bubbles that remained in the microwells due to the hydrophobicity of PDMS, wells were filled with 200 µl of culture medium, and the plate was placed in vacuum for 1 hour followed by 1 min centrifugation at 300g to remove residual bubbles. The plate was then left at 4°C overnight. Cells were loaded into the microwell array, followed by loading of the activation microbeads. First, the medium was removed and replaced with  $12.5 \times 10^3$  primary naïve T cells in 100 µl of culture medium. The plate was centrifuged at 300g for 1 min to allow cells to settle. Residual cells were aspirated, and 100 µl of medium containing activation microbeads was loaded into the wells. Beads were

left to settle for 5 min after which the medium was aspirated and replaced with 100 µl of fresh tissue culture medium without phenol red as indicated. This seeding procedure gave an average cell number of  $4.7 \pm 2.6$  and average bead number of  $8.5 \pm 4.5$  (fig. S5A). Then, wells were loaded with 100 µl of culture medium supplemented with a combination of FITC-labeled anti-CD44, PE-labeled anti-CD62L, and APC-labeled anti-CD45.2 antibodies in dilutions of  $1:5 \times 10^3$ ,  $1:10^4$ , and  $1:10^4$  respectively. This gave a final culture volume of 200 µl with twice the listed antibody dilution. For antibody specifications, see table S4.

### Perturbation in microwells

Whenever antibodies were used, microwells were supplemented with 100 µl of culture medium containing fluorescent antibodies as indicated above, and 100 µl of culture medium containing the blocking antibodies at double the final concentration. Final antibody concentrations were as follows: anti-IL-2 (10 µg/ml), anti-IL-6 (10 µg/ml), anti-IL-15/15R (5 µg/ml), and anti-IL-6Ra (5 µg/ml). IL-6 and recombinant human-IL-2 (both from R&D) were given at final concentrations of 20 ng/ml for IL-6 and 0.1 or 10 ng/ml for human-IL-2.

When anti-LFA-1 and ICAM-1 were used, microwells were precoated with the reagent to interfere with cell-cell adhesion. For anti-LFA-1, microwells were precoated with 20 µg/ml of the antibody in PBS overnight at 4°C. ICAM-1 coating was performed in two steps. First, microwells were coated with protein A to enable correct positioning of the ICAM-1 molecules, and then coated with ICAM-1/Fc chimera protein (R&D): 20 µg/ml of Protein A was diluted in PBS (+calcium +magnesium) supplemented with 1 mM NaHCO<sub>3</sub>. Seventy microliters of the mix was added to the microwells and incubated for 2 hours at 37°C. Protein A solution was aspirated and replaced with 70 µl of blocking buffer (PBS supplemented with 2% human serum albumin, Calbiochem) followed by a 10 min incubation at room temperature. Finally, ICAM-1 was diluted in blocking buffer to 20 µg/ml, and 70 µl was added to the microwells and incubated overnight at 4°C. Wells were washed with culture medium before cell seeding.

### Live cell imaging acquisition

For time-lapse experiments, a Ti-eclipse microscope (Nikon Instruments) was used equipped with an automated stage, incubator, and a closed chamber that allowed for CO<sub>2</sub> flow over the 96-well plate. Cells were imaged using 20×/NA = 0.17 objective (sFluor, Nikon) and monitored using bright-field illumination and three fluorescence channels: FITC, Cy3, and Cy5. Time-lapse movies were collected using the Andor software. Cells were imaged every 2 to 6 hours, depending on the experiment, using an Andor iXon-888 EMCCD camera (1024 × 1024 pixels, 13-µm pixel size).

### Confocal microscopy imaging

Naïve CD4 cells were cultured at a concentration of  $10^6$  cells/ml in 24-well plates, and in the

presence of 2  $\mu\text{M}$  Cell Activation Cocktail (PMA + ionomycin, Biolegend). After 24 hours, paraformaldehyde (PFA, Biolegend) was added directly to the cells to a final concentration of 1.6%, and cells were incubated for 15 min at room temperature in the dark. To keep clusters intact, washing of the cells was performed gently on the culture dish by adding staining buffer (PBS + 4% FCS) and removing it several times. Cells were supplemented with 500  $\mu\text{l}$  of staining buffer with 5  $\mu\text{l}$  of PE-labeled anti-IL-6st and 5  $\mu\text{l}$  of Alexa Fluor 488-labeled anti-IL-2Ra and incubated for 30 min at room temperature in the dark. Cells were washed with PBS and seeded on a six-well plate with a glass bottom. Cells were imaged using Fv3000 laser scanning confocal microscope (Olympus) using PLAPON 60 $\times$  OSC2 super-corrected objective with 1.4 NA, at a sampling speed of 2  $\mu\text{s}/\text{pixel}$ .

### PI3K and JAK-STAT inhibition

Naïve CD4<sup>+</sup> T cells were cultured in three initial cell numbers as indicated in the text for 72 hours. At the culture start time, small-molecule inhibitors for PI3K (LY-294002, Sigma) and JAK-STAT (AZ-1480, Sellechem) were added to separate cultures at a concentration of 5  $\mu\text{M}$  each. After 72 hours, cells were labeled as described above and measured using flow cytometry.

### Pospho-STAT sample preparation

The protocol for phos-STAT measurements was adopted from Feinerman *et al.* (48). For phos-STAT5 measurements, naïve CD4 T cells were cultured in 24-well plates at a cell concentration of  $1 \times 10^6$  cells/ml with activation microbeads (1:1 bead:cell ratio) and in the presence of anti-IL-2 (10  $\mu\text{g}/\text{ml}$ ) to prevent endogenous IL-2 binding. After 24 hours, unclustered and clustered cells were separated using a 10  $\mu\text{m}$  mesh (PlutiSelect) by performing the following steps: The mesh was washed with culture medium from both sides. Cells were washed and supplemented with 500  $\mu\text{l}$  of fresh medium and (without mixing) passed through the mesh into one well on a 24-well plate; this contained the unclustered cells. The mesh was washed three times with 500  $\mu\text{l}$  of medium, flipped onto a second well, and washed from its other side with 500  $\mu\text{l}$  of medium. This well contained the clustered cells (see fig. S16A for images of clustered and unclustered cells). Each fraction was supplemented with 500  $\mu\text{l}$  of culture medium with 1  $\mu\text{l}$  of live-dead blue reagent and either with or without 5 ng/ml human-IL-2. The culture plate was incubated for 10 min at 37°C and immediately supplemented with PFA (Biolegend) in a final concentration of 1.6%. Cells were incubated in the dark for 15 min at room temperature and then centrifuged at 600g. The supernatant was aspirated and replaced with 1 ml of 90% ice-cold methanol followed by a 30 min incubation on ice in the dark. After incubation, cells were washed with staining buffer (PBS + 4% FCS) and supplemented with 90  $\mu\text{l}$  of stain buffer and 10  $\mu\text{l}$  of stain mix containing 5  $\mu\text{l}$  of FITC-labeled anti-pSTAT5 (ebioscience), 1  $\mu\text{l}$  PerCP-Cy5.5-labeled anti-IL-2Ra, 1  $\mu\text{l}$  of PE-labeled

anti-CD4, and 3  $\mu\text{l}$  stain buffer. Cells were incubated at room temperature for 30 min in the dark, washed twice with stain buffer, and measured on the flow cytometer (LSRII; BD). For phos-STAT3 measurements, cells were cultured in two initial concentrations: either  $1 \times 10^6$  or  $6.25 \times 10^4$  cells/ml. After 24 hours, cells were washed once and supplemented with 200  $\mu\text{l}$  RPMI either with or without 10 ng/ml of IL-6. Cell fixation and staining were performed as described above for phos-STAT5.

### Imaging flow cytometry (ImageStream) sample preparation and analysis

Naïve cells were cultured in 24-well plates at a concentration of  $10^6$  cells/ml with activation microbeads (1:1 bead:cell ratio). After 72 hours of culture, cells were collected into 5 ml tubes and washed with PBS. Supernatant was aspirated and replaced with 100  $\mu\text{l}$  PBS with 10  $\mu\text{M}$  Hoechst 33342 (life technologies) and with Alexa Fluor 488-labeled anti-CD45.2, APC-labeled anti-IL2Ra, and PE-labeled anti-CD3 (1  $\mu\text{l}$  each). Cells were incubated at room temperature in the dark for 30 min, washed once with PBS, transferred to 1.5-ml tubes (Eppendorf), and centrifuged at 400g. Supernatant was gently aspirated using a 10-ml syringe, and samples were supplemented with PBS to a final volume of 50  $\mu\text{l}$ . Samples were measured using multispectral Imaging Flow Cytometry (ImageStreamX mark II; Amnis Corp, part of EMD millipore, Seattle, WA). Imaging was performed using 60 $\times$ /NA = 0.9 lens; the lasers used were 405 nm (120 mW) for Hoechst, 488 nm (100 mW), 561 nm (200 mW), 642 nm (150 mW), and 785 nm (5 mW) for side scatter (SSC) channel imaging. At least  $5 \times 10^4$  cells were collected from each sample, and data were analyzed using image analysis software (IDEAS 6.2; Amnis Corp). Cells were gated for single cells or doublets using the area and aspect ratio features, and for focused cells using the Gradient RMS feature. CD3<sup>+</sup>CD25<sup>+</sup> cells were gated, and the relative concentration of the cell-cell synapse was calculated. First, a mask was created to delineate the cell synapse, using the VALLEY mask (rectangular mask that sits between two bright regions, such as between two nuclei. This minimum intensity identifies the intersection between the two objects) dilated for 3 pixels, based on the nuclear Hoechst staining. We calculated the intensity concentration ratio feature (the ratio of the intensity inside the first input mask to the intensity of the union of the two masks); the higher the score, the greater the concentration of intensity inside the first mask. The ratio is mapped to a log scale based on the Valley mask.

### RNA sequencing

For CD62L<sup>+/−</sup> cell sorting (described in Fig. 1, D and E), naïve CD4<sup>+</sup> cells were cultured for 48 and 72 hours at a concentration of  $2.5 \times 10^5$  cells/ml and activated using microbeads as described above. This cell concentration was selected to yield a relatively similar fraction of CD62L<sup>+</sup> and CD62L<sup>−</sup> cells at the indicated time points. At each time point, CD4<sup>+</sup>CD62L<sup>+/−</sup> cells were sorted

into 100  $\mu\text{l}$  of Lysis/Binding buffer (Invitrogen). For comparison between the transcriptome of cells cultured in high and low concentrations (fig. S7), naïve CD4 cells were cultured in 24-well plates (1 ml/well) for 18, 30, 48, and 72 hours at a starting concentration of either  $10^6$  or  $6.25 \times 10^4$  cells/ml. At the indicated time points, live CD4<sup>+</sup> cells were sorted into 100  $\mu\text{l}$  of Lysis/Binding buffer.

Total RNA was extracted with poly-dT beads (Dynabeads, Invitrogen). We used a variation of the MARS-seq protocol (49) developed to produce single-cell RNA-seq libraries. In brief, the protocol consists of special designed primers with unique molecular identifiers (UMI) for accurate molecule counting and a step of linear amplification of the initial mRNA pool, followed by a library construction step. In this way, the diversity of the original pool of messenger RNAs is preserved even if the amount of input RNA is low. Three replicate libraries were prepared for each of the different populations. First, the samples were incubated at 72°C for 3 min and immediately transferred to 4°C. Then, 2  $\mu\text{l}$  of an RT reaction mix [10 mM DTT, 4 mM dNTP, 2.5 U/ $\mu\text{l}$  Superscript III RT enzyme in 50 mM Tris-HCl (pH 8.3), 75 mM KCl, and 3 mM MgCl<sub>2</sub>] were added into each well. The samples were then centrifuged and incubated as follows: 2 min at 42°C, 50 min at 50°C, and 5 min at 85°C. Indexed samples with equivalent amount of cDNA were pooled. The pooled cDNA was converted to double-stranded DNA with a second strand synthesis kit (NEB) in a 20- $\mu\text{l}$  reaction, incubating for 2.5 hours at 16°C. The product was purified with 1.4 $\times$  volumes of SPRI beads, eluted in 8  $\mu\text{l}$  and in vitro-transcribed (with the beads) at 37°C overnight for linear amplification using the T7 High Yield RNA polymerase IVT kit (NEB). Following IVT, the DNA template was removed with Turbo DNase I (Ambion) 15 min at 37°C and the amplified RNA (aRNA) purified with 1.2 $\times$  volumes of SPRI beads. RNA was chemically fragmented (median size ~200 nucleotides) by incubating for 3 min at 70°C in Zn<sup>2+</sup> RNA fragmentation solution (Ambion) and purified with two volumes of SPRI beads. Next, a partial Illumina Read1 sequencing adapter that includes a pool barcode was single strand ligated to the fragmented RNA using a T4 RNA ligase I (New England Biolabs): The RNA (5  $\mu\text{l}$ ) was preincubated for 3 min at 70°C with 1  $\mu\text{l}$  of 100  $\mu\text{M}$  ligation adapter. Then, 14  $\mu\text{l}$  of a mix containing 9.5% DMSO, 1 mM ATP, 20% PEG8000, and 1 U/ $\mu\text{l}$  T4 ligase in 50 mM Tris HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, and 1mM DTT was added. The reaction was incubated at 22°C for 2 hours. The ligated product was reverse transcribed using Affinity Script RT enzyme (Agilent) and a primer complementary to the ligated adapter: The reaction was incubated for 2 min at 42°C, 45 min at 50°C and 5 min at 85°C. cDNA was purified with 1.5 $\times$  volumes of SPRI beads. The library was completed and amplified through a nested polymerase chain reaction (PCR) with 0.5  $\mu\text{M}$  of P5\_Rd1 and P7\_Rd2 primers and PCR ready mix (Kapa Biosystems): The forward primer contains



the Illumina P5-Read1 sequences and the reverse primer contains the P7-Read2 sequences. The amplified pooled library was purified with 0.7× volumes of SPRI beads to remove primer leftovers. Library concentration was measured by a Qubit fluorometer (Life Technologies), and mean molecule size was determined by TapeStation (Agilent). DNA libraries were sequenced on an Illumina NextSeq 500 with an average of  $\sim 1 \times 10^6$  aligned reads per sample.

### Quantitative PCR

Naïve CD4<sup>+</sup> T cells were isolated, and  $3 \times 10^5$  naïve cells were transferred to 1.5 ml tubes (Eppendorf), washed once with PBS, supplemented with 1 ml TRI-reagent (Sigma), and immediately frozen in  $-80^\circ\text{C}$ , for subsequent RNA extraction (naïve sample). The remaining cells were cultured in 24-well plates at a concentration of  $10^6$  cells/ml and activated using microbeads. At 8 and 24 hours, cells were detached from the activation microbeads by 5-min incubation in culture medium containing 5 mM EDTA. Then, cells were placed on a magnet for 3 min, and the cell fraction was transferred to 1.5-ml tubes. Cells were washed (450g for 10 min) supplemented with 1 ml of tri-reagent and frozen. Total RNA was isolated according to the manufacturer's protocol (Life Technologies). The total RNA was reverse-transcribed into cDNA using M-MLV RT primed with oligo(dT) primers (Promega). Real-time PCR was performed using the Fast SYBR Green master mix in the Quant Studio 5 machine (Applied Biosystems). See table S5 for primer specifications. Primer amplification efficiency and specificity were verified for each set of primers at a final concentration of 250 nM. The determined amounts of cDNA template were 4 ng for *Hprt* and *Il2* and 32 ng for *Il6*. mRNA expression levels of the tested genes relative to *Hprt* were calculated using the  $\Delta\Delta\text{Ct}$  method, with the naïve sample as a reference.

### REFERENCES AND NOTES

1. V. R. Buchholz *et al.*, Disparate individual fates compose robust CD8<sup>+</sup> T cell immunity. *Science* **340**, 630–635 (2013). doi: [10.1126/science.1235454](https://doi.org/10.1126/science.1235454); pmid: [23493420](https://pubmed.ncbi.nlm.nih.gov/23493420/)
2. P. D. Hodgkin, M. R. Dowling, K. R. Duffy, Why the immune system takes its chances with randomness. *Nat. Rev. Immunol.* **14**, 711 (2014). doi: [10.1038/nri3734-cl](https://doi.org/10.1038/nri3734-cl); pmid: [25212742](https://pubmed.ncbi.nlm.nih.gov/25212742/)
3. J. M. Marchingo *et al.*, T cell signaling. Antigen affinity, costimulation, and cytokine inputs sum linearly to amplify T cell expansion. *Science* **346**, 1123–1127 (2014). doi: [10.1126/science.1260044](https://doi.org/10.1126/science.1260044); pmid: [25430770](https://pubmed.ncbi.nlm.nih.gov/25430770/)
4. C. R. Plumlee, B. S. Sheridan, B. B. Cicek, L. Lefrançois, Environmental cues dictate the fate of individual CD8<sup>+</sup> T cells responding to infection. *Immunity* **39**, 347–356 (2013). doi: [10.1016/j.immuni.2013.07.014](https://doi.org/10.1016/j.immuni.2013.07.014); pmid: [23932571](https://pubmed.ncbi.nlm.nih.gov/23932571/)
5. S. L. Reiner, W. C. Adams, Lymphocyte fate specification as a deterministic but highly plastic process. *Nat. Rev. Immunol.* **14**, 699–704 (2014). doi: [10.1038/nri3734](https://doi.org/10.1038/nri3734); pmid: [25190286](https://pubmed.ncbi.nlm.nih.gov/25190286/)
6. J. T. Chang *et al.*, Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science* **315**, 1687–1691 (2007). doi: [10.1126/science.1139393](https://doi.org/10.1126/science.1139393); pmid: [17323376](https://pubmed.ncbi.nlm.nih.gov/17323376/)
7. J. Arsenio *et al.*, Early specification of CD8<sup>+</sup> T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. *Nat. Immunol.* **15**, 365–372 (2014). doi: [10.1038/ni.2842](https://doi.org/10.1038/ni.2842); pmid: [24584088](https://pubmed.ncbi.nlm.nih.gov/24584088/)
8. A. L. Marzo *et al.*, Initial T cell frequency dictates memory CD8<sup>+</sup> T cell lineage commitment. *Nat. Immunol.* **6**, 793–799 (2005). doi: [10.1038/ni1227](https://doi.org/10.1038/ni1227); pmid: [16025119](https://pubmed.ncbi.nlm.nih.gov/16025119/)
9. V. P. Badovinac, J. S. Haring, J. T. Hartly, Initial T cell receptor transgenic cell precursor frequency dictates critical aspects of

- the CD8<sup>+</sup> T cell response to infection. *Immunity* **26**, 827–841 (2007). doi: [10.1016/j.immuni.2007.04.013](https://doi.org/10.1016/j.immuni.2007.04.013); pmid: [17555991](https://pubmed.ncbi.nlm.nih.gov/17555991/)
10. J. J. Obar, K. M. Khanna, L. Lefrançois, Endogenous naïve CD8<sup>+</sup> T cell precursor frequency regulates primary and memory responses to infection. *Immunity* **28**, 859–869 (2008). doi: [10.1016/j.immuni.2008.04.010](https://doi.org/10.1016/j.immuni.2008.04.010); pmid: [18499487](https://pubmed.ncbi.nlm.nih.gov/18499487/)
11. A. Lanzavecchia, F. Sallusto, Understanding the generation and function of memory T cell subsets. *Curr. Opin. Immunol.* **17**, 326–332 (2005). doi: [10.1016/j.coi.2005.04.010](https://doi.org/10.1016/j.coi.2005.04.010); pmid: [15886125](https://pubmed.ncbi.nlm.nih.gov/15886125/)
12. F. Sallusto, D. Lenig, R. Förster, M. Lipp, A. Lanzavecchia, Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712 (1999). doi: [10.1038/44385](https://doi.org/10.1038/44385); pmid: [10537110](https://pubmed.ncbi.nlm.nih.gov/10537110/)
13. S. M. Kaech *et al.*, Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. *Nat. Immunol.* **4**, 1191–1198 (2003). doi: [10.1038/ni1009](https://doi.org/10.1038/ni1009); pmid: [14625547](https://pubmed.ncbi.nlm.nih.gov/14625547/)
14. R. M. Kondrack *et al.*, Interleukin 7 regulates the survival and generation of memory CD4 cells. *J. Exp. Med.* **198**, 1797–1806 (2003). doi: [10.1084/jem.20030735](https://doi.org/10.1084/jem.20030735); pmid: [14662907](https://pubmed.ncbi.nlm.nih.gov/14662907/)
15. J. Li, G. Huston, S. L. Swain, IL-7 promotes the transition of CD4 effectors to persistent memory cells. *J. Exp. Med.* **198**, 1807–1815 (2003). doi: [10.1084/jem.20030725](https://doi.org/10.1084/jem.20030725); pmid: [14676295](https://pubmed.ncbi.nlm.nih.gov/14676295/)
16. J. J. Obar, L. Lefrançois, Memory CD8<sup>+</sup> T cell differentiation. *Ann. N. Y. Acad. Sci.* **1183**, 251–266 (2010). doi: [10.1111/j.1749-6632.2009.05126.x](https://doi.org/10.1111/j.1749-6632.2009.05126.x); pmid: [20146720](https://pubmed.ncbi.nlm.nih.gov/20146720/)
17. J. A. Best *et al.*, Transcriptional insights into the CD8<sup>+</sup> T cell response to infection and memory T cell formation. *Nat. Immunol.* **14**, 404–412 (2013). doi: [10.1038/ni.2536](https://doi.org/10.1038/ni.2536); pmid: [23396170](https://pubmed.ncbi.nlm.nih.gov/23396170/)
18. B. Youngblood, J. S. Hale, R. Ahmed, Memory CD8 T cell transcriptional plasticity. *FASEB J.* **29**, 38 (2015). doi: [10.12703/P7-38](https://doi.org/10.12703/P7-38); pmid: [26097712](https://pubmed.ncbi.nlm.nih.gov/26097712/)
19. C. M. Carlson *et al.*, Kruppel-like factor 2 regulates thymocyte and T-cell migration. *Nature* **442**, 299–302 (2006). doi: [10.1038/nature04882](https://doi.org/10.1038/nature04882); pmid: [16855590](https://pubmed.ncbi.nlm.nih.gov/16855590/)
20. H. Ichii *et al.*, Bcl6 is essential for the generation of long-term memory CD4<sup>+</sup> T cells. *Int. Immunol.* **19**, 427–433 (2007). doi: [10.1093/intimm/dxm007](https://doi.org/10.1093/intimm/dxm007); pmid: [17307796](https://pubmed.ncbi.nlm.nih.gov/17307796/)
21. A. Banerjee *et al.*, Cutting edge: The transcription factor eomesodermin enables CD8<sup>+</sup> T cells to compete for the memory cell niche. *J. Immunol.* **185**, 4988–4992 (2010). doi: [10.4049/jimmunol.1002042](https://doi.org/10.4049/jimmunol.1002042); pmid: [20935204](https://pubmed.ncbi.nlm.nih.gov/20935204/)
22. Y. Lou, X. Lu, X. Dang, FOXO1 Up-Regulates Human L-selectin Expression Through Binding to a Consensus FOXO1 Motif. *Gene Regul. Syst. Bio.* **6**, 139–149 (2012). pmid: [23133314](https://pubmed.ncbi.nlm.nih.gov/23133314/)
23. B. Kakaradov *et al.*, Early transcriptional and epigenetic regulation of CD8<sup>+</sup> T cell differentiation revealed by single-cell RNA sequencing. *Nat. Immunol.* **18**, 422–432 (2017). doi: [10.1038/ni.3688](https://doi.org/10.1038/ni.3688); pmid: [28218746](https://pubmed.ncbi.nlm.nih.gov/28218746/)
24. M. A. Cannarile *et al.*, Transcriptional regulator Id2 mediates CD8<sup>+</sup> T cell immunity. *Nat. Immunol.* **7**, 1317–1325 (2006). doi: [10.1038/ni1403](https://doi.org/10.1038/ni1403); pmid: [17086188](https://pubmed.ncbi.nlm.nih.gov/17086188/)
25. F. Masson *et al.*, Id2-mediated inhibition of E2A represses memory CD8<sup>+</sup> T cell differentiation. *J. Immunol.* **190**, 4585–4594 (2013). doi: [10.4049/jimmunol.1300099](https://doi.org/10.4049/jimmunol.1300099); pmid: [23536629](https://pubmed.ncbi.nlm.nih.gov/23536629/)
26. B. Hoffman, D. A. Liebermann, Apoptotic signaling by c-MYC. *Oncogene* **27**, 6462–6472 (2008). doi: [10.1038/onc.2008.312](https://doi.org/10.1038/onc.2008.312); pmid: [18955973](https://pubmed.ncbi.nlm.nih.gov/18955973/)
27. O. Cazzalini, A. I. Scovassi, M. Savio, L. A. Stivala, E. Prosperi, Multiple roles of the cell cycle inhibitor p21(CDKN1A) in the DNA damage response. *Mutat. Res.* **704**, 12–20 (2010). doi: [10.1016/j.mrrev.2010.01.009](https://doi.org/10.1016/j.mrrev.2010.01.009); pmid: [20096807](https://pubmed.ncbi.nlm.nih.gov/20096807/)
28. S. M. Kaech, W. Cui, Transcriptional control of effector and memory CD8<sup>+</sup> T cell differentiation. *Nat. Rev. Immunol.* **12**, 749–761 (2012). doi: [10.1038/nri3307](https://doi.org/10.1038/nri3307); pmid: [23080391](https://pubmed.ncbi.nlm.nih.gov/23080391/)
29. S. Aduttler-Lieber *et al.*, Substrate-bound CCL21 and ICAM1 combined with soluble IL-6 collectively augment the expansion of antigen-specific murine CD4<sup>+</sup> T cells. *Blood Adv.* **1**, 1016–1030 (2017). doi: [10.1182/bloodadvances.2016001545](https://doi.org/10.1182/bloodadvances.2016001545); pmid: [29296744](https://pubmed.ncbi.nlm.nih.gov/29296744/)
30. I. Zaretsky *et al.*, Monitoring the dynamics of primary T cell activation and differentiation using long term live cell imaging in microwell arrays. *Lab Chip* **12**, 5007–5015 (2012). doi: [10.1039/c2lc40808b](https://doi.org/10.1039/c2lc40808b); pmid: [23072772](https://pubmed.ncbi.nlm.nih.gov/23072772/)
31. M. Polonsky, B. Chain, N. Friedman, Clonal expansion under the microscope: Studying lymphocyte activation and differentiation using live-cell imaging. *Immunol. Cell Biol.* **94**, 242–249 (2016). doi: [10.1038/icb.2015.104](https://doi.org/10.1038/icb.2015.104); pmid: [26606974](https://pubmed.ncbi.nlm.nih.gov/26606974/)
32. E. D. Hawkins, M. L. Turner, M. R. Dowling, C. van Gend, P. D. Hodgkin, A model of immune regulation as a consequence of randomized lymphocyte division and death

- times. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5032–5037 (2007). doi: [10.1073/pnas.0700026104](https://doi.org/10.1073/pnas.0700026104); pmid: [17360353](https://pubmed.ncbi.nlm.nih.gov/17360353/)
33. I. Kinyio *et al.*, Real-time tracking of cell cycle progression during CD8<sup>+</sup> effector and memory T-cell differentiation. *Nat. Commun.* **6**, 6301 (2015). doi: [10.1038/ncomms7301](https://doi.org/10.1038/ncomms7301); pmid: [25709008](https://pubmed.ncbi.nlm.nih.gov/25709008/)
34. K. Thurler, D. Gerecht, E. Friedmann, T. Höfer, Three-dimensional gradients of cytokine signaling between T cells. *PLoS Comput. Biol.* **11**, e1004206 (2015). doi: [10.1371/journal.pcbi.1004206](https://doi.org/10.1371/journal.pcbi.1004206); pmid: [25923703](https://pubmed.ncbi.nlm.nih.gov/25923703/)
35. Z. Liu *et al.*, Immune homeostasis enforced by co-localized effector and regulatory T cells. *Nature* **528**, 225–230 (2015). doi: [10.1038/nature16169](https://doi.org/10.1038/nature16169); pmid: [26605524](https://pubmed.ncbi.nlm.nih.gov/26605524/)
36. N. Waysbord, D. Russ, B. M. Chain, N. Friedman, Coupled IL-2-dependent extracellular feedback governs two distinct consecutive phases of CD4 T cell activation. *J. Immunol.* **191**, 5822–5830 (2013). doi: [10.4049/jimmunol.1301575](https://doi.org/10.4049/jimmunol.1301575); pmid: [24244020](https://pubmed.ncbi.nlm.nih.gov/24244020/)
37. C. A. Sabatos *et al.*, A synaptic basis for paracrine interleukin-2 signaling during homotypic T cell interaction. *Immunity* **29**, 238–248 (2008). doi: [10.1016/j.immuni.2008.05.017](https://doi.org/10.1016/j.immuni.2008.05.017); pmid: [18674934](https://pubmed.ncbi.nlm.nih.gov/18674934/)
38. R. Ahmed, M. J. Bevan, S. L. Reiner, D. T. Fearon, The precursors of memory: Models and controversies. *Nat. Rev. Immunol.* **9**, 662–668 (2009). doi: [10.1038/nri2619](https://doi.org/10.1038/nri2619); pmid: [19680250](https://pubmed.ncbi.nlm.nih.gov/19680250/)
39. A. V. Gett, F. Sallusto, A. Lanzavecchia, J. Geginat, T cell fitness determined by signal strength. *Nat. Immunol.* **4**, 355–360 (2003). doi: [10.1038/ni908](https://doi.org/10.1038/ni908); pmid: [12640450](https://pubmed.ncbi.nlm.nih.gov/12640450/)
40. M. A. Williams, M. J. Bevan, Shortening the infectious period does not alter expansion of CD8 T cells but diminishes their capacity to differentiate into memory cells. *J. Immunol.* **173**, 6694–6702 (2004). doi: [10.4049/jimmunol.173.11.6694](https://doi.org/10.4049/jimmunol.173.11.6694); pmid: [15557161](https://pubmed.ncbi.nlm.nih.gov/15557161/)
41. H. Dooms, K. Wolslegel, P. Lin, A. K. Abbas, Interleukin-2 enhances CD4<sup>+</sup> T cell memory by promoting the generation of IL-7R  $\alpha$ -expressing cells. *J. Exp. Med.* **204**, 547–557 (2007). doi: [10.1084/jem.20062381](https://doi.org/10.1084/jem.20062381); pmid: [17312008](https://pubmed.ncbi.nlm.nih.gov/17312008/)
42. S. A. Nish *et al.*, T cell-intrinsic role of IL-6 signaling in primary and memory responses. *eLife* **3**, e01949 (2014). pmid: [24842874](https://pubmed.ncbi.nlm.nih.gov/24842874/)
43. D. M. Mitchell, E. V. Ravkov, M. A. Williams, Distinct roles for IL-2 and IL-15 in the differentiation and survival of CD8<sup>+</sup> effector and memory T cells. *J. Immunol.* **184**, 6719–6730 (2010). doi: [10.4049/jimmunol.0904089](https://doi.org/10.4049/jimmunol.0904089); pmid: [20483725](https://pubmed.ncbi.nlm.nih.gov/20483725/)
44. K. K. McKinstry *et al.*, Effector CD4 T-cell transition to memory requires late cognate interactions that induce autocrine IL-2. *Nat. Commun.* **5**, 5377 (2014). doi: [10.1038/ncomms6377](https://doi.org/10.1038/ncomms6377); pmid: [25369785](https://pubmed.ncbi.nlm.nih.gov/25369785/)
45. D. Comte *et al.*, Engagement of SLAMF3 enhances CD4<sup>+</sup> T-cell sensitivity to IL-2 and favors regulatory T-cell polarization in systemic lupus erythematosus. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9321–9326 (2016). doi: [10.1073/pnas.1605081113](https://doi.org/10.1073/pnas.1605081113); pmid: [27482100](https://pubmed.ncbi.nlm.nih.gov/27482100/)
46. A. Gérard *et al.*, Secondary T cell-T cell synaptic interactions drive the differentiation of protective CD8<sup>+</sup> T cells. *Nat. Immunol.* **14**, 356–363 (2013). doi: [10.1038/ni.2547](https://doi.org/10.1038/ni.2547); pmid: [23475183](https://pubmed.ncbi.nlm.nih.gov/23475183/)
47. F. Zhao, J. L. Cannons, M. Dutta, G. M. Griffiths, P. L. Schwartzberg, Positive and negative signaling through SLAM receptors regulate synapse organization and thresholds of cytotoxicity. *Immunity* **36**, 1003–1016 (2012). doi: [10.1016/j.immuni.2012.05.017](https://doi.org/10.1016/j.immuni.2012.05.017); pmid: [22683123](https://pubmed.ncbi.nlm.nih.gov/22683123/)
48. O. Feinerman *et al.*, Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Mol. Syst. Biol.* **6**, 437 (2010). doi: [10.1038/msb.2010.90](https://doi.org/10.1038/msb.2010.90); pmid: [21119631](https://pubmed.ncbi.nlm.nih.gov/21119631/)
49. D. A. Jain *et al.*, Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014). doi: [10.1126/science.1247651](https://doi.org/10.1126/science.1247651); pmid: [24531970](https://pubmed.ncbi.nlm.nih.gov/24531970/)

### ACKNOWLEDGMENTS

We thank I. Shachar and L. Radomir for help with *Slamf6*<sup>−/−</sup> mice and useful discussions, and R. Alon and S. Feigelson for help with ICAM-1 coating protocol and productive conversations. We also thank T. Bigdary for the design of the graphical abstract and N. David for help with figure design. **Funding:** N.F. was supported by the Israel Science Foundation (grant no. 1184/15) and by a grant from the Israeli Ministry of Science, Technology and Space and the German Cancer Research Center (DKFZ). N.F. is Incumbent of The Eugene and Marcia Appelbaum Professorial Chair.

N.F. and B.C. were supported by a grant from the Weizmann UK foundation. B.C. is supported by a Weston Visiting Professorship from the Weizmann Institute of Science.

**Author contributions:** M.P. was involved in the design and implementation of all experiments and computer simulations, performed downstream and statistical analysis, generated the figures, and wrote the manuscript. J.R. was involved in implementation of image analysis algorithms and provided initial insights for this study. A.K.P. constructed the stochastic simulations. I.Z. conducted adoptive transfer experiments. S.M. conducted quantitative PCR experiments. C.B. constructed cDNA libraries and conducted RNA-seq. E.D., N.K., and G.S.

were involved in downstream analysis of RNA-seq. data. Z.P. conducted image-stream analysis. B.C. and N.F. designed experiments, supervised all stages of the project, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** MATLAB code used for single-cell stochastic simulations, including additional parameters that are required for the run, can be downloaded through GitHub: <https://github.com/innate2adaptive/Single-cell-imaging>. RNA-seq data have been deposited to Gene Expression Omnibus (GEO) with the accession no. GSE112347. All other data needed to evaluate the conclusions in this paper are present either in the main text or the supplementary materials.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/eaaj1853/suppl/DC1](http://www.sciencemag.org/content/360/6394/eaaj1853/suppl/DC1)  
Supplementary Text  
Figs. S1 to S16  
Tables S1 to S5  
Movie S1  
References (50–52)

9 September 2016; resubmitted 19 December 2017  
Accepted 17 April 2018  
10.1126/science.aaj1853



## RESEARCH ARTICLE SUMMARY

## CANCER

# Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases

Arnaud Pommier, Naishitha Anaparthi, Nicoletta Memos, Z. Larkin Kelley, Alizée Gouronnec, Ran Yan, Cédric Auffray, Jean Albregues, Mikala Egeblad, Christine A. Iacobuzio-Donahue, Scott K. Lyons, Douglas T. Fearon\*

**INTRODUCTION:** Pancreatic ductal adenocarcinoma (PDA) is the fourth most common cause of death from cancer worldwide and has a 5-year survival rate of 6%. Patients who have had their primary PDA surgically resected often develop metastatic disease, despite intra-operative examination of the liver confirming the absence of macrometastatic lesions. These observations lead to the conclusion that latent metastases, detectable only microscopically, were present in these patients and were responsible for the postoperative development of metastatic disease.

**RATIONALE:** Latent metastases with the potential for outgrowth had been considered to be lesions in which cancer cell proliferation is balanced by immune-mediated cancer cell death, but a more recent explanation invokes quiescent, single disseminated cancer cells

(DCCs). Single, nonreplicating DCCs have been observed in several cancer types, but whether quiescence is enforced by the microenvironment or is cancer cell-autonomous is not known. Immunity, both innate and adaptive, also is likely to have a role in the selection and/or maintenance of latent DCCs. This has long been suspected on the basis of the clinical observation of donor-derived cancer in immune-suppressed recipients of allografts. However, there is an unexplained paradox of immunity preventing the outgrowth of latent metastases but not eliminating them.

**RESULTS:** We studied the metastatic process in the context of an ongoing adaptive immune response because of the occurrence of cancer cell-specific immunity in human and mouse PDA. Livers from patients and mice with PDA contained single DCCs with

an unusual phenotype of being negative for cytokeratin 19-negative (CK19) and major histocompatibility complex class I (MHCI). The absence of the expression of MHCI in DCCs and the occurrence of cancer-specific CD8<sup>+</sup> T cells in the genetically engineered mouse model of PDA, and possibly in patients with PDA, suggested that DCCs may be selected by an anticancer immune response during the metastatic process.

To investigate this hypothesis, we created a mouse model that would allow us to determine how DCCs develop, their relationship to metastatic latency, and the role of immunity. Intraportal injection of immunogenic PDA cells into preimmunized mice seeded livers only with single, nonreplicating DCCs lacking MHCI and CK19, whereas naïve recipients of PDA cells had macrometastases. We found that a subpopulation of PDA cells with the phenotype of DCCs was present

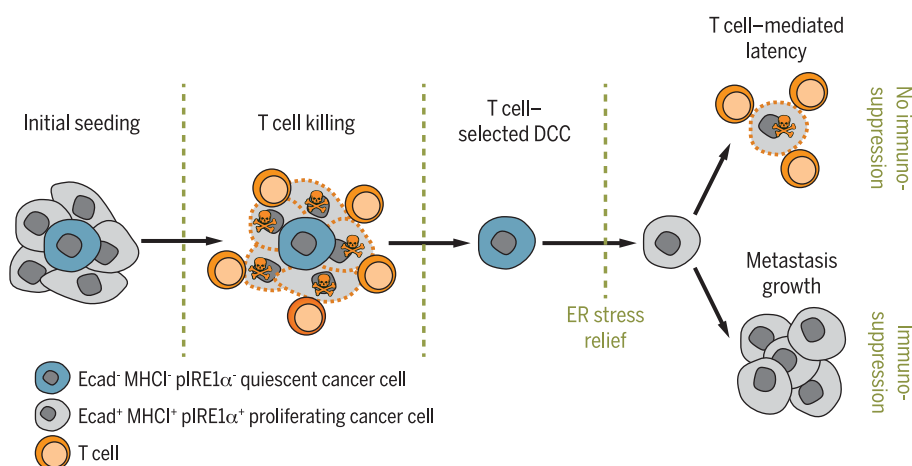
## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aao4908>

in vitro and that those cells are the precursors of DCCs in vivo. We found that T cells select DCCs by eliminating MHCI<sup>+</sup> proliferating cancer cells.

To identify the cell-autonomous "switch" regulating the developmental state of the metastases, we performed single-cell RNA sequencing of PDA cells with the DCC phenotype. This transcriptomic analysis demonstrated an endoplasmic reticulum (ER) stress response. Moreover, DCCs showed a lack of activation of the IRE1 $\alpha$  (inositol-requiring enzyme 1 $\alpha$ ) pathway of the unfolded protein response, whereas the PERK (protein kinase RNA-like ER kinase) pathway was activated, suggesting that DCCs cannot resolve ER stress. Relieving ER stress pharmacologically with a chemical chaperone or genetically by overexpression of spliced XBP1, in combination with T cell depletion, stimulated outgrowth of macrometastatic lesions containing PDA cells expressing MHCI and CK19.

**CONCLUSION:** We find that a PDA-specific adaptive immune response selects DCCs, in which the ER stress response accounts for both quiescence and resistance to immune elimination. Accordingly, outgrowth of DCCs to macrometastases requires not only relief from the cancer cell-autonomous ER stress response, but also suppression of systemic immunity. Thus, the ER stress response is a cell-autonomous reaction that enables DCCs to escape immunity and establish latent metastases. ■



## Unresolved ER stress allows disseminated cancer cells to escape the T cell response.

Quiescent cancer cells that exhibit unresolved ER stress lack expression of MHCI. They avoid killing by T cells and become latent disseminated cancer cells (DCCs). Resolution of ER stress allows DCCs to regain proliferative capacities and grow as overt metastases, only if the T cell response is disrupted, because they also regain the expression of MHCI. Ecad, E-cadherin.

The list of author affiliations is available in the full article online.

\*Corresponding author. Email: [dfearon@cshl.edu](mailto:dfearon@cshl.edu)

Cite this article as A. Pommier et al., *Science* **360**, eaao4908 (2018). DOI: [10.1126/science.aao4908](https://doi.org/10.1126/science.aao4908)

## RESEARCH ARTICLE

## CANCER

# Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases

Arnaud Pommier<sup>1</sup>, Naishitha Anaparthi<sup>1,2</sup>, Nicoletta Memos<sup>1</sup>, Z. Larkin Kelley<sup>3</sup>, Alizée Gouronnet<sup>1</sup>, Ran Yan<sup>1</sup>, Cédric Auffray<sup>4</sup>, Jean Albregues<sup>1</sup>, Mikala Egeblad<sup>1</sup>, Christine A. Iacobuzio-Donahue<sup>5</sup>, Scott K. Lyons<sup>1</sup>, Douglas T. Fearon<sup>1,3,6\*</sup>

The majority of patients with pancreatic ductal adenocarcinoma (PDA) develop metastatic disease after resection of their primary tumor. We found that livers from patients and mice with PDA harbor single disseminated cancer cells (DCCs) lacking expression of cytokeratin 19 (CK19) and major histocompatibility complex class I (MHC I). We created a mouse model to determine how these DCCs develop. Intraportal injection of immunogenic PDA cells into preimmunized mice seeded livers only with single, nonreplicating DCCs that were CK19<sup>−</sup> and MHC I<sup>−</sup>. The DCCs exhibited an endoplasmic reticulum (ER) stress response but paradoxically lacked both inositol-requiring enzyme 1 $\alpha$  activation and expression of the spliced form of transcription factor XBP1 (XBP1s). Inducible expression of XBP1s in DCCs, in combination with T cell depletion, stimulated the outgrowth of macrometastatic lesions that expressed CK19 and MHC I. Thus, unresolved ER stress enables DCCs to escape immunity and establish latent metastases.

**P**ancreatic ductal adenocarcinoma (PDA) is the fourth most common cause of death from cancer worldwide (1) and has a 5-year survival rate of 6% (2). The majority of patients are diagnosed after the disease has spread beyond the primary tumor site. Patients who show no clinical evidence of local invasion or distant metastasis are treated by surgery, but about 75% of these patients develop metastatic disease within 2 years after resection of their primary tumors (3, 4), despite intra-operative examination of the liver confirming the absence of macrometastatic lesions (5). These observations lead to the conclusion that latent metastases, detectable only microscopically, were present in these patients at the time of surgery and were responsible for the postoperative development of metastatic disease.

Latent metastases are thought to be lesions in which cancer cell proliferation is balanced by immune-mediated cancer cell death (6–8). A more recent hypothesis highlights the role of quiescent, single disseminated cancer cells (DCCs) (9–11). Single, nonreplicating DCCs have been observed in several cancer types, most often in the bone marrow (12, 13), but whether quiescence

is enforced by the microenvironment or is cancer cell-autonomous is not known (14). Immunity, both innate (15) and adaptive (7, 16, 17), also is likely to have a role in the selection and/or maintenance of latent DCCs. This has long been suspected based on the clinical observation that immunosuppressed recipients of allografts occasionally present with donor-derived cancer (18, 19). However, there is an unexplained paradox of immunity preventing the outgrowth of latent metastases but not eliminating them.

We examined the nature of latent metastases in PDA by developing a mouse model that replicates the characteristics of hepatic DCCs that are found in human PDA and in spontaneously arising PDA in mice. We studied the metastatic process in the context of an ongoing adaptive immune response because of the occurrence of cancer cell-specific immunity in human and mouse PDA (20–23).

## Quiescent, single DCCs in the livers of humans and mice with PDA

To determine whether hepatic DCCs occur in human PDA, we microscopically examined tissue sections from the primary tumors and livers of five patients with PDA who had no clinically detectable hepatic metastases. The clinicopathologic characteristics of the patients are shown in table S1. The tumors were genotyped as having p53 loss of heterozygosity, which permitted staining for mutant p53 accumulation as an identifier of cancer cells (24). p53<sup>+</sup> cancer cells were present in both the primary tumors and livers of all five patients. The p53<sup>+</sup> cancer cells resided in the livers

as single cells that were consistently negative for cytokeratin 19 (CK19), Ki67, and major histocompatibility complex class I (MHC I), in contrast to the cancer cells in the primary tumors, which exhibited all three markers (Fig. 1, A to C). We also examined the livers from mice bearing the autochthonous LSL-Kras<sup>G12D/+</sup>; LSL-Trp53<sup>R172H/+</sup>; Pdx-1-Cre; Rosa<sup>YFP</sup> (KPCY) model of PDA (G, glycine; D, aspartic acid; R, arginine; H, histidine; YFP, yellow fluorescent protein) (25–27), which recapitulates human PDA. In livers devoid of macrometastases, we found both YFP<sup>+</sup> micro-metastases and DCCs. Whereas the micrometastases always expressed CK19, Ki67, and MHC I, the single DCCs were mainly CK19<sup>−</sup> (32 of 40), Ki67<sup>−</sup> (22 of 22), and MHC I<sup>−</sup> (28 of 28) (Fig. 1, D to F). Thus, the livers of patients and mice with PDA contain DCCs that share an unusual phenotype linking the loss of epithelial gene expression and quiescence with a potential for escape from T cell recognition.

## A mouse model of hepatic metastasis in the context of an adaptive immune response

The absence of MHC I expression in DCCs and the presence of cancer-specific CD8<sup>+</sup> T cells in the genetically engineered mouse model of PDA (20), and possibly in patients with PDA (28), suggested that DCCs may be selected by an anti-cancer immune response during the metastatic process. Accordingly, we developed a mouse model of hepatic metastases that allowed us to assess the effect of a preexisting immune response. The mM1 (mouse metastasis 1) cell line was derived from a spontaneous liver metastasis of a mouse bearing an autochthonous PDA and was stably transfected with a transposon vector directing the expression of diphtheria toxin receptor (DTR), Herpes simplex thymidine kinase (HSV-TK), firefly luciferase, and mTagBFP2 to generate mM1DTLB cells. Syngeneic C57Bl/6 mice were injected subcutaneously with 10<sup>6</sup> mM1DTLB cells; tumors were grown for 14 days and then eliminated by treatment with diphtheria toxin (DTx) and ganciclovir (GcV). These “preimmunized” mice and naïve mice were challenged by intrasplenic injections of 10<sup>6</sup> mM1DTLB cells, followed by splenectomy, thereby seeding the liver via the portal vein, as in PDA (Fig. 2A).

We found that in naïve mice, whole-body bioluminescence increased from day 1 after injection, consistent with the growth of hepatic metastases. In preimmunized mice, however, whole-body bioluminescence decreased after day 1, and by day 7, luminescence was at background levels (Fig. 2, B and C). In additional cohorts of naïve and preimmunized mice, livers were removed at intervals after the intrasplenic injection of mM1DTLB cells and assessed for bioluminescence. Both photon flux (Fig. 2B) and visually detectable metastases increased in the livers of naïve mice between days 5 and 20, whereas metastatic foci were barely detectable in preimmunized mice at day 5 and were absent at later time points (Fig. 2D). A potential role for T cells in the elimination of mM1DTLB cancer cells was

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. <sup>2</sup>Department of Molecular and Cellular Biology, Stony Brook University, Stony Brook, NY 11794, USA. <sup>3</sup>Weill Cornell Medicine, New York, NY 10065, USA. <sup>4</sup>Institut Cochin, Paris Descartes Université, CNRS UMR8104, INSERM U1016, 75014 Paris, France. <sup>5</sup>Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. <sup>6</sup>Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge CB2 0RE, UK.

\*Corresponding author. Email: dfearon@cshl.edu



suggested by the finding that in the livers of preimmunized mice, tumor cells were frequently surrounded by both CD8<sup>+</sup> and CD8<sup>+</sup> CD3<sup>+</sup> T cells by 24 hours (fig. S2A). This possibility was confirmed by treating preimmunized mice with depleting antibodies against CD4 and CD8, alone or together, or with isotype control antibody (fig. S2B). Depleting either CD4<sup>+</sup> or CD8<sup>+</sup> T cells abrogated the ability of preimmunization to suppress the development of macrometastases (fig. S2B).

Microscopic examination of livers from naïve and preimmunized mice revealed the presence of macrometastatic lesions in the former, but only single DCCs in the latter. The DCCs differed from cancer cells in the macrometastases in the following ways: They did not express the epithelial markers CK19 and E-cadherin (Ecad), they did not express Ki67, they did not incorporate 5-ethynyl-2'-deoxyuridine (EdU), and they did not express MHC1 (Fig. 2, E to I). Importantly, the phenotype of the DCCs in this metastasis model is similar to that of DCCs in human PDA and the KPCY mouse. T cells in the vicinity of DCCs were infrequent, whereas they surrounded macrometastatic lesions in naïve mice (Fig. 2J). The absence of CK19 and Ecad expression did not indicate an epithelial-mesenchymal transition (EMT) because DCCs did not express the EMT markers desmin,  $\alpha$ SMA, Snail1, or Slug (fig. S3).

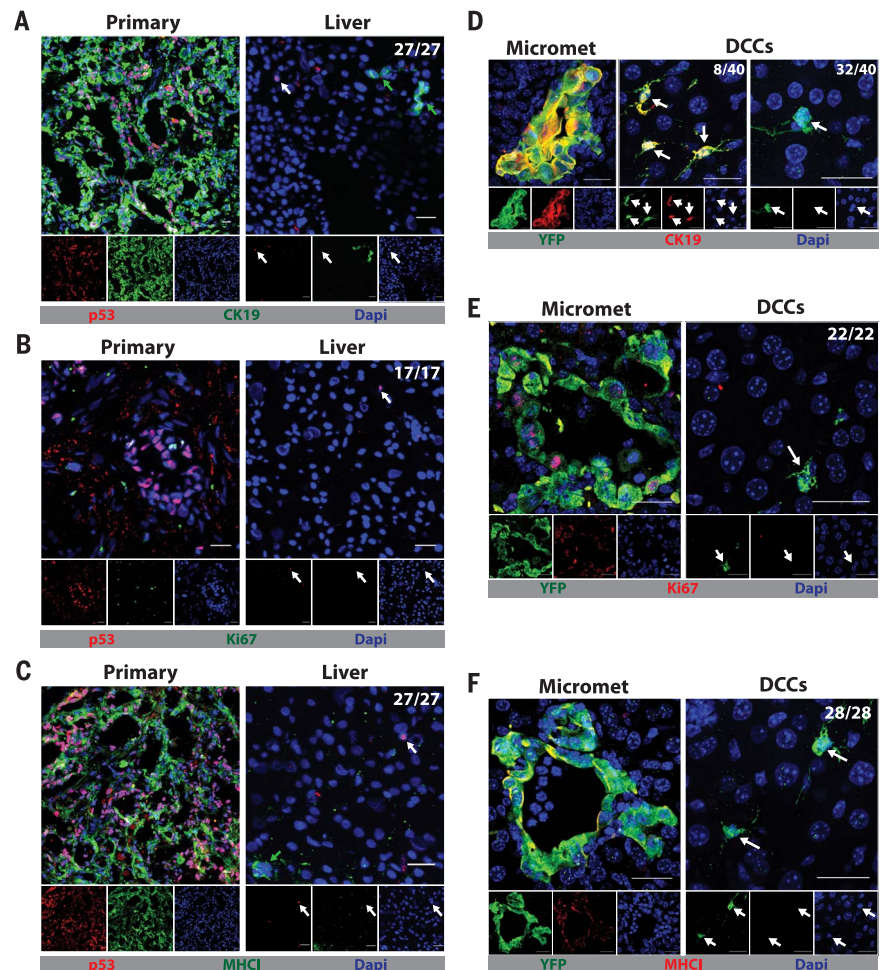
When naïve and preimmunized mice were challenged with mMIDTLB cells via the tail vein, lung macrometastases developed only in the naïve mice, whereas CK19<sup>+</sup> DCCs were observed in the lungs of preimmunized mice (fig. S4). We also assessed the 1242 cell line derived from a primary PDA tumor of a KPC mouse (29), which had been modified with the same transposon vector, for its ability to generate DCCs. When naïve and preimmunized mice were challenged by intrasplenic injection of 1242DTLB cells, hepatic macrometastases developed in the naïve mice, whereas only CK19<sup>+</sup> DCCs were observed in the livers of preimmunized mice (fig. S5). To assess more stringently the proliferation of DCCs, we labeled the mMIDTLB cells with carboxyfluorescein diacetate succinimidyl ester (CFSE) before intrasplenic injection into naïve or preimmunized mice. Whereas all cancer cells in the macrometastases of naïve mice became CFSE<sup>+</sup>, DCCs in pre-immunized mice retained CFSE (fig. S6A). Mice that had received intrasplenic injections of mMIDTLB cells were also given EdU in drinking water for 20 days. Whereas cancer cells in macrometastases in naïve mice incorporated EdU, almost all DCCs in preimmunized mice were EdU<sup>+</sup> (fig. S6B), indicating a lack of proliferation. Therefore, the occurrence of quiescent, MHC1<sup>+</sup> DCCs in the absence of macrometastases is a consequence of an ongoing cancer-specific immune response.

### A latent capacity for outgrowth of DCCs is controlled by T cells

We next examined whether a latent capacity of DCCs for outgrowth into macrometastatic lesions might be revealed by T cell depletion. When depleting antibodies against CD4 and CD8 were administered to preimmunized mice 3 weeks

after the establishment of DCCs, macrometastases appeared in 10 of 15 mice. When T cells were depleted at 9 weeks, macrometastases appeared in 2 of 15 mice (Fig. 3). The metastases were composed of cancer cells that had reexpressed CK19 and MHC1, suggesting that DCCs revert to an epithelial phenotype to initiate the formation of macrometastases. The lower frequency of macrometastases in mice in which T cells were depleted at 9 weeks suggests that DCCs with a capacity for reversion decreased between 3 and 9 weeks. Indeed, we found that livers of preimmunized mice had fewer DCCs at 9 weeks than at 3 weeks (fig. S7A). This loss of DCCs may reflect the killing by T cells of spontaneous revertants, as suggested by the occasional occurrence of a CK19<sup>+</sup> DCC surrounded by T cells in the preimmunized mice

(fig. S7B). The reexpression of MHC1 by these growing metastases also suggests the means by which T cells control the outgrowth of spontaneously reverting DCCs. To demonstrate that T cells alone are both necessary and sufficient for controlling the growth of MHC1<sup>+</sup> DCCs, we depleted natural killer (NK) cells by administering antibody against NK1.1 at the time of mMIDTLB cell challenge. Preimmunized mice lacking NK cells were indistinguishable, with respect to the occurrence of DCCs and absence of macrometastases, from control antibody-treated mice (fig. S8). This evidence for a dominant role of the T cell in controlling DCCs is supported by the finding that DCCs were never seen to be in contact with CD45<sup>+</sup> (fig. S9A), F4/80<sup>+</sup> (fig. S9B), CD19<sup>+</sup> (fig. S9C), CD31<sup>+</sup> (fig. S9D),  $\alpha$ SMA<sup>+</sup> (fig. S9E), or Ly-6G<sup>+</sup> (fig. S9F)



**Fig. 1. Single DCCs with a characteristic phenotype are present in the livers of humans and mice with PDA.** (A to C) Immunofluorescence (IF) of sections from the primary tumor and liver of a patient with PDA that were stained with anti-p53 to reveal cancer cells (red) and with (A) anti-CK19, (B) anti-Ki67, or (C) anti-MHC1 (all green). Photomicrographs are representative of five patients. (D to F) IF of sections from a liver of a KPCY mouse with spontaneous PDA and no hepatic macrometastases. Sections were stained with anti-YFP to reveal cancer cells (green) and with (D) anti-CK19, (E) anti-Ki67, and (F) anti-MHC1 (all red). Photomicrographs are representative of three mice. The ratios shown in the top right corners of the photomicrographs indicate the frequency of the observed DCC phenotype relative to the total number of DCCs that were assessed. All frequencies are compiled in table S2. White arrows designate DCCs and, in the sections of human livers, green arrows designate normal, liver-resident CK19<sup>+</sup> or MHC1<sup>+</sup> cells. Scale bars, 25  $\mu$ m. Micromet, micrometastasis; DAPI, 4',6-diamidino-2-phenylindole.

cells; the distribution of these cell types was similar to that in the normal liver (fig. S9, G to N).

### A rare subpopulation of PDA cells in vitro with the phenotype of DCCs

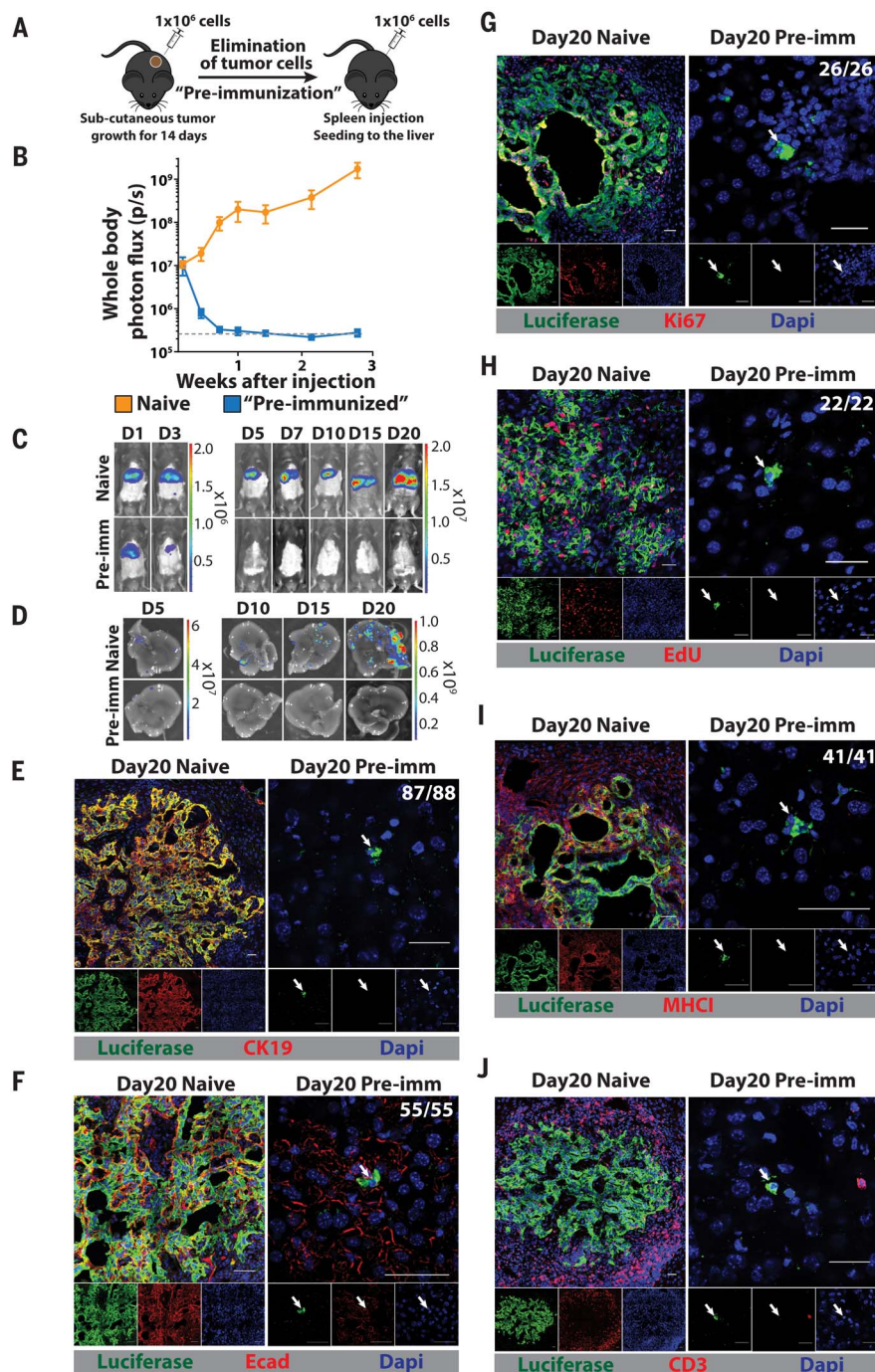
The absence of MHC1 expression by DCCs raised the possibility that these cells were present in the injected PDA population and were negatively selected by T cells. Indeed, ~1% of the mM1DTLB cells in tissue culture were Ecad<sup>+</sup> and CK19<sup>+</sup>, and all Ecad<sup>+</sup> cells were MHC1<sup>+</sup> (Fig. 4, A and B). The Ecad<sup>+</sup> cells resided in a nonproliferating subpopulation of cells, as indicated by the resistance of these HSV-TK-expressing cells to GcV (Fig. 4C).

### Fig. 2. A mouse model for hepatic DCCs.

(A) Mice were preimmunized by subcutaneous injection of  $10^6$  mM1DTLB PDA cells derived from a hepatic metastasis of a KPC mouse. After 2 weeks, tumors were eliminated by treating mice with DTx and GcV. For hepatic metastases,  $10^6$  mM1DTLB PDA cells were injected intrasplenically into naïve and preimmunized mice, followed immediately by splenectomy. (B and C) Tumor growth was measured by whole-body bioluminescence imaging (p/s, photons per second; D, day). The dashed gray line represents the background luminescence in tumor-free mice. (D) Ex vivo photon flux of whole livers was measured at days 5, 10, 15, and 20 after cancer cell injection. Results are representative of three experiments with at least five mice per group. (E to J) IF of sections from the livers of a naïve mouse (left panels) and a preimmunized mouse (right panels) that were stained with anti-luciferase (green) to identify cancer cells and with (E) anti-CK19, (F) anti-Ecad, (G) anti-Ki67, (H) EdU, (I) anti-MHC1, and (J) CD3 (all red). For EdU staining, mice were injected every 12 hours with EdU for 3 days. Photographs are representative of 20 mice from three independent experiments. The ratios shown in the top right corners of the photomicrographs indicate the frequency of the observed DCC phenotype relative to the total number of DCCs that were assessed. All frequencies are compiled in table S2. White arrows designate DCCs. Scale bars, 25  $\mu$ m.

A phenotypic plasticity of the mM1DTLB cells was shown when we cultured FACS (fluorescence-activated cell sorting)-purified Ecad<sup>+</sup> and Ecad<sup>+</sup> cells for 3 days and found that they generated Ecad<sup>+</sup> and Ecad<sup>+</sup> cells, respectively (Fig. 4D). The Ecad<sup>+</sup> MHC1<sup>+</sup> phenotype shared by PDA cells in vitro and DCCs in vivo suggested that the former may be the precursors of the latter. We assessed this possibility by intrasplenically injecting  $10^6$  Ecad<sup>+</sup> or  $10^4$  Ecad<sup>+</sup> cells into naïve and preimmunized mice. The growth of Ecad<sup>+</sup> macrometastases in the livers of the naïve mice receiving Ecad<sup>+</sup> cells was similar to that in naïve mice receiving unsorted mM1DTLB cells (Fig. 4E). Microscopic examination of the livers

of these mice, however, revealed no Ecad<sup>+</sup> DCCs (Fig. 4F). Injection of Ecad<sup>+</sup> cells into preimmunized mice resulted in a rapid decline in photon flux, with no specific signal by day 7 (Fig. 4E), and microscopic examination of these livers also demonstrated the absence of DCCs (Fig. 4F). Injection of Ecad<sup>+</sup> cells into naïve mice resulted in delayed development of hepatic macrometastases (Fig. 4E), and microscopy revealed the presence of DCCs (Fig. 4F). Injection of Ecad<sup>+</sup> cells into preimmunized mice led to DCCs but no macrometastases, confirming that an ongoing T cell response controls DCC outgrowth (Fig. 4, E and F). The macrometastases found in naïve mice injected with Ecad<sup>+</sup> cells were





CK19<sup>+</sup> (Fig. 4G) and MHCI<sup>+</sup> (Fig. 4H), indicating reversion to an epithelial phenotype.

In summary, the origin of the DCC is the Ecad<sup>-</sup> cell, because DCCs were not present in naïve or preimmunized mice after the injection of Ecad<sup>+</sup> cells. Reversion from the quiescent, Ecad<sup>-</sup> state to the proliferating, Ecad<sup>+</sup> phenotype was observed in naïve mice but was masked by the ongoing T cell response in preimmunized mice because reversion is associated with reexpression of MHCI. Thus, the two states of mMIDTLB cells that were observed in vivo—the proliferating MHCI<sup>+</sup>, Ecad<sup>+</sup>, CK19<sup>+</sup> macrometastasis and the quiescent MHCI<sup>-</sup>, Ecad<sup>-</sup>, CK19<sup>-</sup> DCC—occur in vitro, reflecting a developmental plasticity that may be controlled by a cell-autonomous process.

### Unresolved endoplasmic reticulum (ER) stress characterizes PDA cells with the DCC phenotype

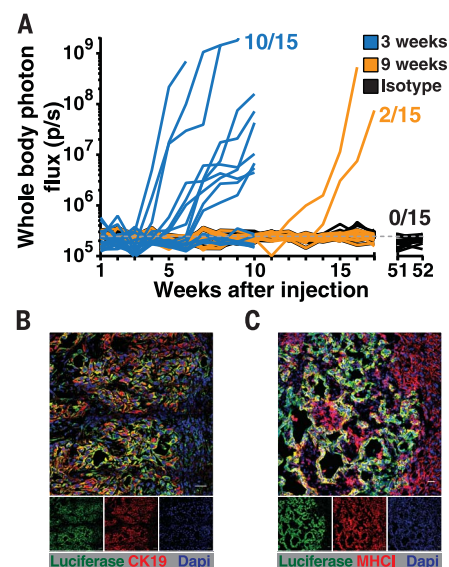
To identify the cell-autonomous “switch” regulating the developmental state of the metastases, we performed single-cell RNA sequencing (scRNA-seq) of in vitro sorted Ecad<sup>+</sup> and Ecad<sup>-</sup> cells. The most up-regulated pathway in Ecad<sup>-</sup> cells relative to Ecad<sup>+</sup> cells was “response to ER stress” (Fig. 5A and fig. S10C), and the most down-regulated pathway was “cell division” (Fig. 5B and fig. S10D). Network analysis of other up-regulated pathways that distinguish the Ecad<sup>-</sup> and Ecad<sup>+</sup> PDA populations, such as “autophagy,” showed that they were linked to “response to ER stress”. Similarly, network analysis of other down-regulated pathways demonstrated linkage to “cell division” (Fig. 5B). These two populations of the mMIDTLB PDA cells were also distinct in principal components analysis, in which Ecad<sup>-</sup> cells appeared to be more heterogeneous than Ecad<sup>+</sup> cells (fig. S10A). Indeed, analysis at the single-cell level identified four subpopulations of Ecad<sup>-</sup> cells, in three of which the dominant up-regulated pathway was related to “response to ER stress,” and the major down-regulated pathway was “cell division” (fig. S11). An EMT signature was not present among the 1639 genes that were differentially expressed between Ecad<sup>+</sup> and Ecad<sup>-</sup> cells (fig. S10B), confirming the immunofluorescence analysis of Ecad<sup>-</sup> cells (fig. S6E). In addition, neither Ecad<sup>+</sup> nor Ecad<sup>-</sup> cells expressed three of the four major NKG2D ligands—Ulbpl1, H60b, and H60c—and both strongly expressed the inhibitory ligand, Qa1, providing an explanation for the absence of a role for NK cells in controlling outgrowth of macrometastases (fig. S7C) (30). Genes that are involved in the processing of MHCI were not differentially expressed, which is consistent with reports that the ER stress response suppresses the expression of MHCI by a posttranscriptional mechanism (31, 32). Last, the conclusion that the ER stress response alters the expression of MHCI and Ecad is supported by the finding that treatment of mMIDTLB PDA cells with tunicamycin, an inducer of ER stress, increased the proportion of cells that were MHCI<sup>+</sup> and Ecad<sup>-</sup> (fig. S12).

The most differentially expressed gene was *Ddit3/CHOP*, the mRNA level of which was 18-fold as high in Ecad<sup>-</sup> cells as in Ecad<sup>+</sup> cells. This

gene encodes a transcription factor, C/EBP homologous protein (CHOP), that is part of the ER stress response. We examined the expression of CHOP protein by immunofluorescence in the Ecad<sup>-</sup> DCCs in human and mouse PDA. Anti-CHOP staining was demonstrated in hepatic DCCs in preimmunized mice (Fig. 5C) and in KPCy mice (Fig. 5D), but not in PDA cells of macrometastases from naïve mice or of micrometastases from KPCy mice (Fig. 5E). Importantly, DCCs in the livers of three of five patients with PDA were also stained by antibody against CHOP (Fig. 5E).

The higher expression level of CHOP in DCCs versus in growing PDA cells with an epithelial phenotype suggested that the response of DCCs to ER stress was not as effective as the response of growing CK19<sup>+</sup>/Ecad<sup>+</sup>/MHCI<sup>+</sup> PDA cells. To restore protein homeostasis, cells undergoing ER stress activate the unfolded protein response (UPR) (33, 34). CHOP is induced by the protein kinase RNA-like ER kinase (PERK) pathway of the UPR (33, 34), and, indeed, this pathway was activated in both Ecad<sup>-</sup> and Ecad<sup>+</sup> cells, as shown by phosphorylation of PERK and EIF2 $\alpha$  (fig. S14A). In contrast, the inositol-requiring enzyme 1 (IRE1 $\alpha$ ) pathway was activated in Ecad<sup>+</sup> cells but not in Ecad<sup>-</sup> cells, as shown by phosphorylation of IRE1 $\alpha$  and splicing of the X-box binding protein 1 (XBP1) transcription factor mRNA (fig. S14A). Impaired activation of the IRE1 $\alpha$  pathway in Ecad<sup>-</sup> cells was supported by scRNA-seq results showing decreased expression of XBP1 target genes relative to their expression levels in Ecad<sup>+</sup> cells (fig. S14B). The cancer cells in primary PDA tumors in patients and hepatic macrometastases in mice also demonstrated phosphorylation of both EIF2 $\alpha$  and IRE1 $\alpha$ , but hepatic DCCs in patients and mice were stained only by anti-pEIF2 $\alpha$  (Fig. 6, A to D). Thus, DCCs may not be able to resolve the ER stress response because they do not activate the IRE1 $\alpha$  pathway, which is required to generate the spliced, active form of XBP1 (XBPIs) that regulates the transcription of multiple proteins that promote protein folding in the ER.

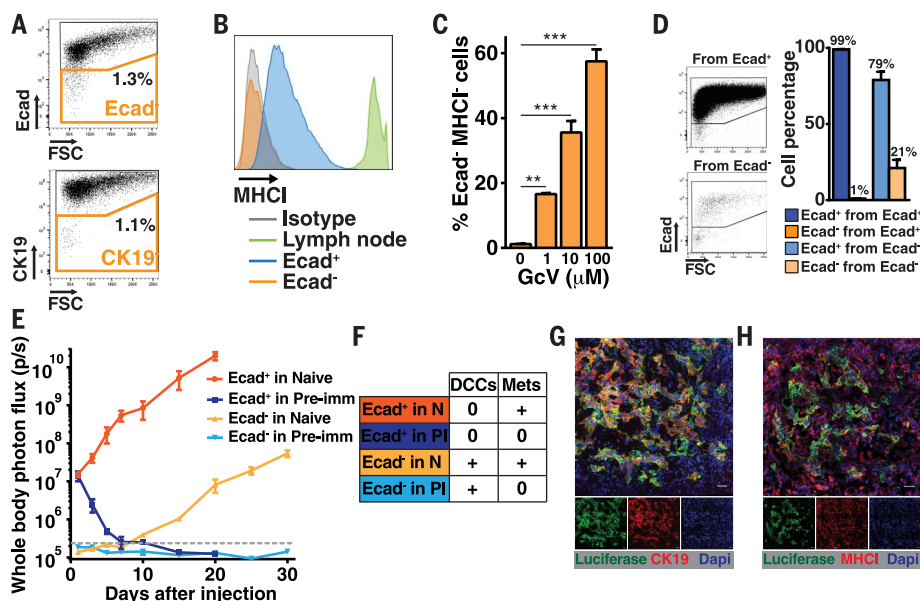
To determine whether this unresolved ER stress of DCCs contributes to their phenotype, we treated mMIDTLB PDA cells with the chemical chaperone 4-phenylbutyrate (4-PBA), which binds to solvent-exposed hydrophobic segments of unfolded or improperly folded proteins, thereby “protecting” them from aggregation and relieving ER stress (35). Treatment with 4-PBA decreased the proportion of cells lacking expression of Ecad (fig. S15A) and increased MHCI expression in both Ecad<sup>+</sup> and Ecad<sup>-</sup> cells (fig. S15B). We also assessed the effect of 4-PBA on the proliferative capability of mMIDTLB PDA cells by first eliminating proliferating mMIDTLB PDA cells through treatment with GeV and then pulsing the residual quiescent cells with EdU overnight in the presence or absence of 4-PBA. Relieving ER stress with 4-PBA increased the proportion of cells incorporating EdU by a factor of 10 (fig. S15C). We extended the analysis to DCCs by administering 4-PBA to preimmunized mice, beginning on the day of the mMIDTLB cell injection and continuing for 3 weeks. This treatment decreased the number of



**Fig. 3. T cells control outgrowth of latent DCCs.** (A) Using bioluminescence imaging, we assessed the growth of hepatic metastases in preimmunized mice that had been depleted of T cells by administration of antibodies against CD4 and CD8 beginning 3 or 9 weeks after splenic injection of mMIDTLB PDA cells. One group of mice was also treated with isotype control antibody. (B and C) IF of sections containing macrometastases from the liver of a preimmunized mouse that had been depleted of T cells 3 weeks after splenic injection of mMIDTLB PDA cells. Anti-luciferase identifies cancer cells. Scale bar, 25  $\mu$ m.

hepatic DCCs by a factor of 4 relative to the number in control mice (fig. S15D), consistent with the possibility that promoting the conversion of DCCs to replicating MHCI<sup>+</sup> cells leads to their T cell-mediated immune elimination. To confirm this interpretation, we began 4-PBA treatment of preimmunized mice 3 weeks after the intrasplenic injection of mMIDTLB cells and depleted the mice of T cells. Within 2 weeks of 4-PBA treatment, all mice had developed MHCI<sup>+</sup> macrometastases (fig. S15, E and F), and the number of macrometastatic lesions in the livers of 4-PBA-treated, T cell-depleted mice was six times the number in the livers of mice subjected only to depletion of T cells (fig. S15E).

We also specifically assessed the role of the IRE1 $\alpha$  pathway in generating PDA cells with the DCC phenotype. In a loss-of-function experiment, mMIDTLB cells were treated with the IRE1 $\alpha$  inhibitor Kira6 (36), which converted the mMIDTLB cells from an Ecad<sup>+</sup> MHCI<sup>+</sup> state to an Ecad<sup>-</sup> MHCI<sup>-</sup> phenotype in a dose-responsive manner (fig. S14C). We then performed gain-of-function experiments with the mM1TetXBPIs cell line expressing a doxycycline (Dox)-inducible form of XBPIs. This allowed us to test the hypothesis that circumventing the defect in IRE1 $\alpha$  activation by expression of the spliced form of XBP1 would resolve ER stress in a more specific and physiological manner than was achieved with 4-PBA



**Fig. 4. A subpopulation of PDA cells in vitro shares phenotypic features with DCCs.** (A) Flow cytometry analysis of mM1DTLB PDA cells that were stained with anti-CK19 or anti-Ecad. Results are representative of five independent experiments. FSC, forward scatter. (B) Flow cytometry measurement of anti-MHCI staining of Ecad<sup>+</sup> and Ecad<sup>-</sup> mM1DTLB PDA cells and of lymph node cells as a comparator. Results are representative of five independent experiments. (C) mM1DTLB PDA cells were treated in vitro for 48 hours with increasing doses of GcV to kill proliferating cells, and the proportion of viable cells that were Ecad<sup>-</sup> and MHC1<sup>-</sup> was measured by flow cytometry. Results are representative of two independent experiments. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . (D) FACS analysis of purified Ecad<sup>+</sup> and Ecad<sup>-</sup> mM1DTLB PDA cells that were cultured for 3 days. Dot plots (left) and histogram (right) are representative of three independent experiments. (E) Growth of hepatic metastases after intrasplenic injection of  $10^6$  Ecad<sup>+</sup> or  $10^4$  Ecad<sup>-</sup> mM1DTLB PDA cells into naïve and preimmunized mice was assessed by whole-body bioluminescence imaging ( $n = 5$  mice per group). The dashed gray line represents the luminescence background in tumor-free mice. (F) Table summarizing the occurrence of DCCs and/or metastases in each group of mice that had been injected with Ecad<sup>+</sup> or Ecad<sup>-</sup> mM1DTLB cells. N, naïve; PI, preimmunized. (G and H) IF of sections from the liver of a naïve mouse that had received an intrasplenic injection of Ecad<sup>-</sup> mM1DTLB cells. Anti-luciferase (green) identifies cancer cells. Photomicrographs are representative of five mice. Scale bars, 25  $\mu$ m. Error bars, mean  $\pm$  SEM.

(fig. S13). Overnight treatment of mM1TetXBPIs cells with Dox induced the expression of XBPIs and almost eliminated the Ecad<sup>-</sup> MHC1<sup>-</sup> subpopulation (fig. S14D). To examine the role of the IRE1 $\alpha$  pathway in DCCs in vivo, we continuously treated preimmunized mice with Dox beginning on the day that mM1TetXBPIs cells were injected intrasplenically. After 3 weeks, the frequency of hepatic DCCs was significantly decreased relative to their frequency in control mice not treated with Dox (Fig. 6E). The possibility that this decrease in XBPIs-expressing hepatic DCCs indicated killing by immune T cells was confirmed by repeating the experiment with the additional intervention of T cell depletion. Within 3 weeks of Dox-induced XBPIs expression in the DCCs, 9 of 10 T cell-deficient mice had developed hepatic macrometastases, whereas only 1 of 8 of the T cell-depleted mice without Dox treatment developed these lesions (Fig. 6F). Furthermore, the number of macrometastatic lesions in the livers of Dox-treated, T cell-depleted mice was seven times that in the livers of mice subjected only to depletion of T cells (Fig. 6G). These results lead us to conclude that unresolved ER stress has a

nonredundant, cell-autonomous role in the maintenance of quiescent, immune-resistant DCCs.

## Discussion

The clinical observation that PDA metastases develop in the majority of patients after the surgical removal of their primary tumors, despite no evidence of metastases at the time of surgery, indicates that these patients had harbored latent metastatic lesions. The nature of these latent metastases was suggested by our finding of single DCCs in the livers of patients and KPCY mice with PDA that have a distinctive phenotype of absent CK19, Ecad, and MHC1. The absence of two typical markers of epithelial ductal adenocarcinoma cells, without the occurrence of characteristic markers of EMT, indicates that these DCCs are distinct from the PDA cells that make up growing macrometastases (37). Moreover, the absence of MHC1 implies an unusual relationship of DCCs to the adaptive immune system (38), as previously suggested in studies of breast, stomach, and colon carcinoma (12, 39). These descriptive findings provided the rationale for developing a mouse model that replicates DCCs with this dis-

tinctive phenotype and allows mechanistic studies defining both the cell-autonomous response responsible for the phenotype and the role of the immune system.

We hypothesized that metastases in patients with PDA may occur in the context of a cancer-specific adaptive immune response (20–23). The absence of MHC1 expression at the surface of DCCs in human and mouse PDA, and the absent or relatively infrequent occurrence of hepatic macrometastases, raised the possibility that immunity prevents the outgrowth of macrometastases while ignoring MHC1<sup>-</sup> DCCs. This prediction was also based on the observation that DCCs in human and mouse PDA, in contrast to cancer cells in macrometastases, were quiescent. The finding that preexisting immunity prevented the occurrence of hepatic macrometastases while permitting the seeding of nonreplicating, MHC1<sup>-</sup> hepatic DCCs verified this prediction and provided a potential explanation of how quiescent metastases can persist in the presence of an adaptive immune response that is capable of suppressing the growth of macrometastases.

This selective effect of adaptive immunity on macrometastases requires that the cell-autonomous mechanism that is responsible for the phenotype of DCCs invariably links the expression of MHC1 to a capacity for cellular replication. This association not only was demonstrated in additional cell lines from primary and metastatic PDA tumors from KPC mice, but also was supported by the finding that ER stress, which inhibits MHC1 expression (31, 32), and cell division (40) were the major up-regulated and down-regulated transcriptional signatures that distinguished Ecad<sup>-</sup> from Ecad<sup>+</sup> PDA cells. Thus, MHC1 expression is “off” in quiescent cells and “on” in replicating cells. In addition, if the markedly increased expression in Ecad<sup>-</sup> PDA cells of CHOP, a transcription factor that is induced by the PERK pathway of the UPR (33, 34), is taken as an indicator of the ER stress response, then the finding that in human and mouse PDA, hepatic DCCs are MHC1<sup>-</sup>, Ki67<sup>-</sup>, and CHOP<sup>+</sup> links quiescence and immune concealment to an ongoing, unresolved ER stress in vivo. Our results confirm the previous suggestion that sustained activation of the UPR allows persistence of DCCs in the bone marrow of breast cancer patients (41) and may be related to the recent observation that knockdown of IRE1 $\alpha$  inhibits tumor growth (42). We showed that resolution of ER stress by a chemical chaperone or, more definitively, by the induced expression of XBPIs causes reversion of DCCs to an epithelial, proliferating, MHC1<sup>+</sup> phenotype. Thus, the DCC phenotype is caused by unresolved ER stress secondary to a block in IRE1 $\alpha$  activation. Our study does not address how PDA cells avoid activating IRE1 $\alpha$ , but other studies suggest that under high ER stress, IRE1 $\alpha$  acquires endonucleolytic activity against RNA targets, in addition to the mRNA encoding XBPI, in a reaction termed regulated IRE1-dependent decay of mRNA (RIDD) (43). These additional endonucleolytic events may be associated with apoptosis, so that the ability of the PDA cell to adopt the



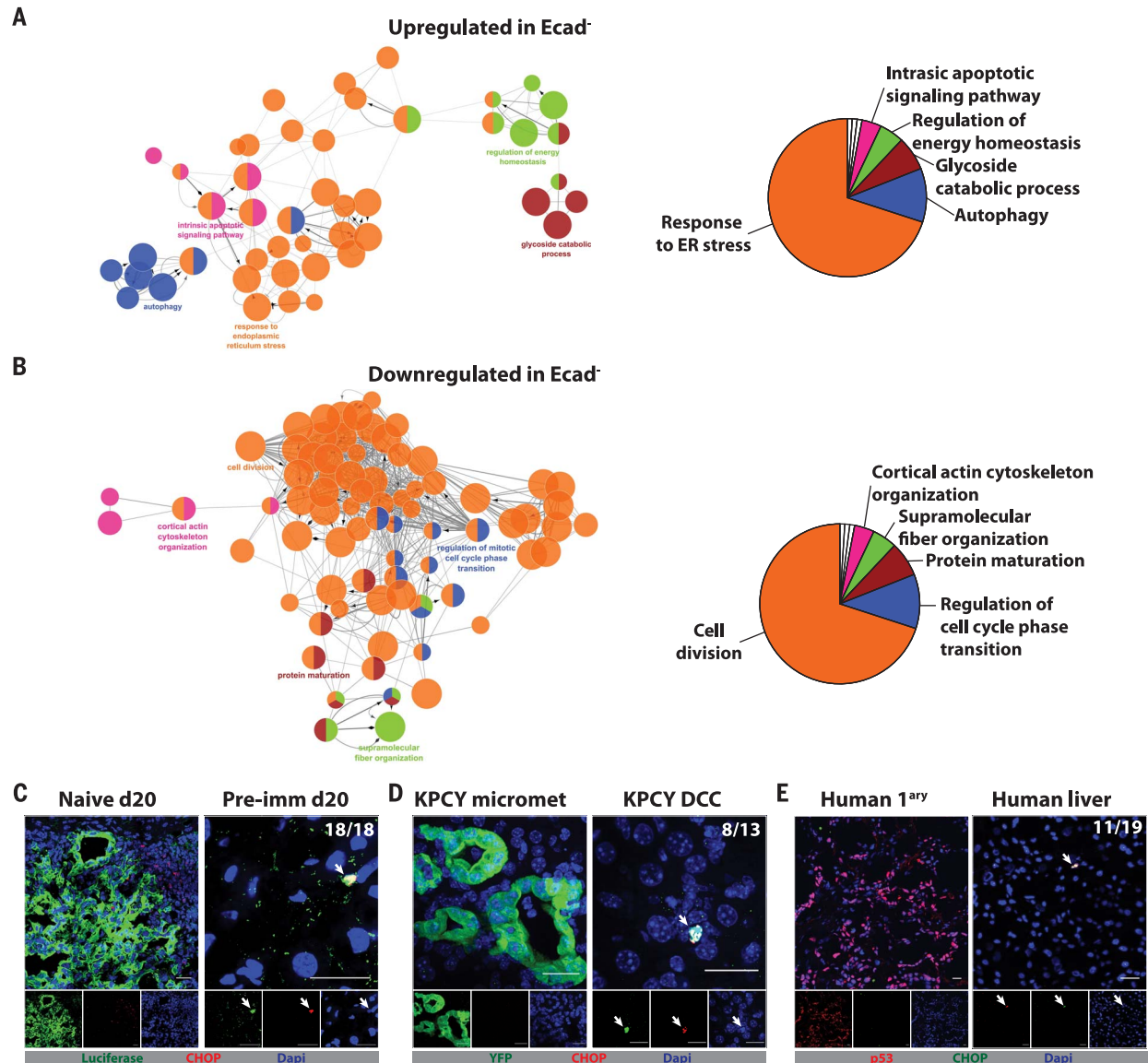
DCC phenotype may protect it not only from death caused by immunity, but also from death caused by the RIDD reaction.

A latent capacity of hepatic DCCs for reversion to growing cancer cells was revealed by the outgrowth of macrometastases after the depletion of T cells. This finding also implied that T cells eliminated those DCCs that spontaneously reverted to a replicating MHC1<sup>+</sup> epithelial phenotype. T cell killing of reverting DCCs must be efficient and occur before an immune-suppressive

microenvironment is established, given that no macrometastases were observed during a 12-month period in T cell-replete mice with hepatic DCCs. The causal link between reversal of ER stress in the DCC and a replicating MHC1<sup>+</sup> epithelial phenotype was established by the use of the chemical chaperone 4-PBA and by the induced expression of XBPIs in DCCs. Relief of ER stress by each intervention enhanced the expression of Ecad and MHC1 and the proliferation of DCCs in vitro and, more importantly,

caused the outgrowth of macrometastases from DCCs in T cell-depleted mice. This observation, coupled with the capacity of 4-PBA and XBPIs to decrease hepatic DCCs in T cell-replete mice, supports the essential role of the unresolved ER stress response in maintaining the DCC phenotype. Other factors, such as CXCR2 (C-X-C chemokine receptor type 2)-expressing neutrophils, may also have a role in this response (44).

The implications for therapy to prevent the occurrence of metastatic disease in patients after



**Fig. 5. The ER stress response in PDA cells that share phenotypic features of DCCs.** scRNA-seq was performed on 104 Ecad<sup>+</sup> and 98 Ecad<sup>-</sup> mM1DTLB PDA cells. **(A and B)** Network analysis, following pathway enrichment analysis, shows ontology relationships between the pathways (left panels). Their relative representation is depicted as a pie chart (right panels) for (A) up-regulated pathways and (B) down-regulated pathways in Ecad<sup>-</sup> cells relative to Ecad<sup>+</sup> cells. Pathways are significant with an adjusted  $P < 0.01$  after Benjamini-Hochberg false discovery rate correction. **(C)** Liver sections from a naïve mouse and a preimmunized mouse were stained with anti-luciferase (green) to identify PDA cells and

with anti-CHOP (red) to identify cells exhibiting an ER stress response. **(D)** Liver sections from a KPCY mouse were stained with anti-YFP (green) to identify PDA cells and with anti-CHOP (red). **(E)** Sections from the primary tumor and liver of a patient with PDA were stained with anti-p53 to identify PDA cells (red) and with anti-CHOP (green). Photomicrographs are representative of five patients who had no detectable liver metastases. The ratios shown in the top right corners of the photomicrographs indicate the frequency of the observed DCC phenotype relative to the total number of DCCs that were assessed. All frequencies are compiled in table S2. White arrows designate DCCs. Scale bars, 25µm.

the surgical removal of their primary PDA may be twofold. First, outgrowth of latent DCCs in the mouse model required suppression of T cell immunity. Elevations of plasma cortisol after pan-createctomy in patients with PDA (45) are in the

range that was found to be immune-suppressive in cachectic mice with PDA (46). This stimulation of the hypothalamic-pituitary-adrenal axis may also occur with the caloric deprivation that commonly occurs in patients after this surgical pro-

cedure (47). Therefore, postoperative parenteral hyperalimentation may be an effective means to decrease the occurrence of metastatic disease. Second, and more speculative, the administration of a chemical chaperone such as 4-PBA pre-operatively, when tumor immunity is intact, might purge organs of latent DCCs, thereby decreasing the likelihood of postoperative metastatic disease.

## Material and Methods

### Animals

Male C57Bl/6 mice 10 to 12 weeks of age purchased from The Jackson Laboratory were used. All procedures were approved by the Cold Spring Harbor Laboratory Institutional Animal Care and Use Committee (IACUC) and were conducted in accordance with the NIH "Guide for the Care and Use of Laboratory Animals."

### Cell culture

Cell lines used in that study were a gift of D. Tuveson (Cold Spring Harbor Laboratory). The metastatic cell line mM1 and the primary cells line 1242 were derived from a KPC mice liver metastasis and a primary tumor, respectively. Cells were cultured in DMEM medium (#10-013-CV, Cellgro) supplemented with 10% fetal calf serum (FCS) (#1500-500, Seradigm), 100 units/ml penicillin, and 100 µg/ml streptomycin.

### Human samples

Postmortem tissues were obtained following the Iacobuzio-Donahue laboratory rapid autopsy program previously described in detail (48). Patients were selected based on two criteria: p53 loss-of-heterozygosity, which allowed the staining for accumulation of mutant p53 protein, and absence of clinically detectable liver metastases.

### Cell culture reagents

Ganciclovir (GcV, Sigma-Aldrich #G2536), was used at concentration ranging from 1-100mM. Tunicamycin (#11445, Cayman chemicals) was used at 5µg/ml. Sodium 4-phenylbutyrate (4-PBA, #11323, Cayman chemicals) was used at 5 mM. 5-Ethynyl-2'-deoxyuridine (EdU, #sc-284628, Santa Cruz biotechnology) was used at 10 mM.

### EdU labeling

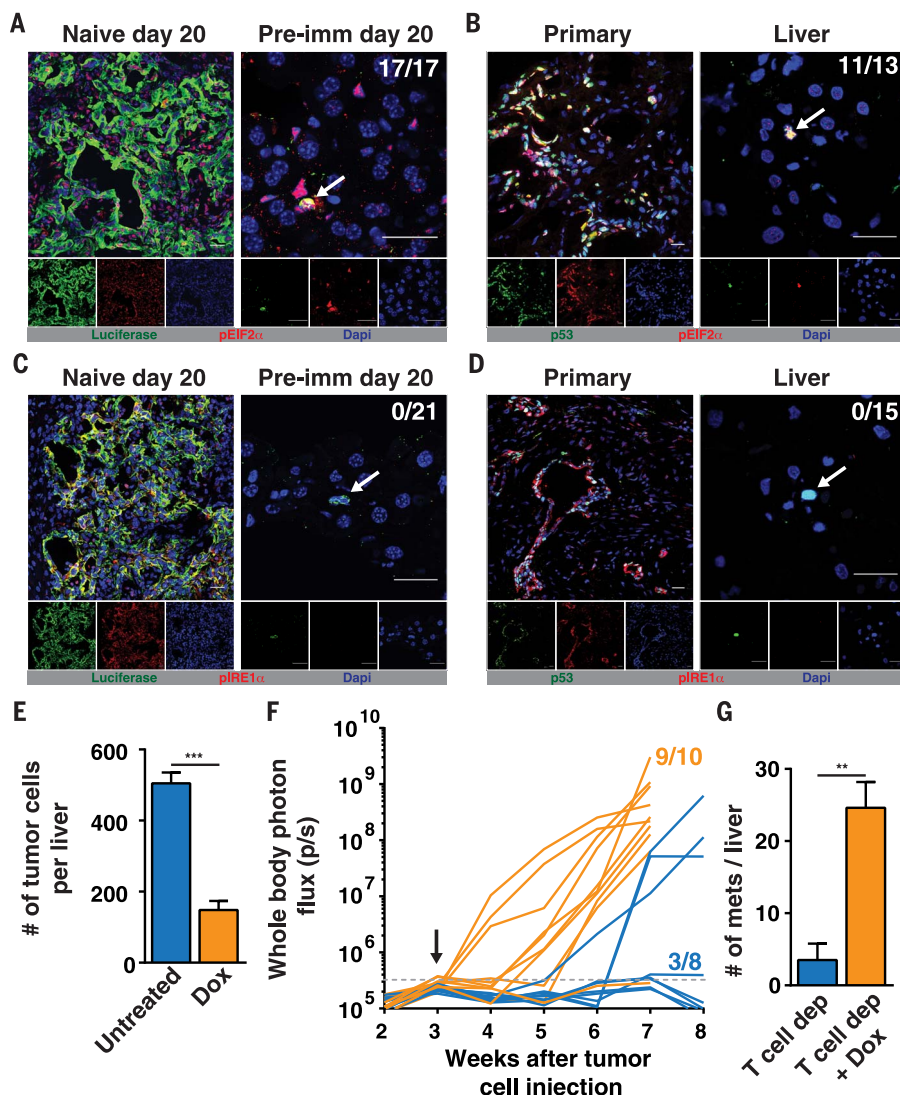
For short term EdU pulse, mice were injected every 12 hours with 1.4 mg of EdU over a 3-day period. For long term experiments, EdU was given in the drinking water at 0.82 mg/ml + 2.5% sucrose. Revelation of EdU containing cells was performed with a kit (Click-iT Plus EdU Imaging Kit, Molecular probes, #C10640) following the manufacturer protocol.

### CFSE labeling

Cells were harvested from the cell culture and CFSE (CellTrace CFSE Cell Proliferation Kit, Molecular probes, #C34554) was used according to the manufacturer protocol.

### Immune cell depletion

200µg of anti-CD4 (clone GK1.5, BioXcell) and/or anti-CD8 (clone 53.6-7, BioXcell) or anti-NK1.1



**Fig. 6. Absence of IRE1α pathway activation in DCCs.** (A) Liver sections from naïve and pre-immunized mice were stained with anti-luciferase (green) to identify PDA cells and with anti-pEIF2α (red) to assess activation of the PERK pathway. (B) Sections from the primary tumor and liver of a patient with PDA were stained with anti-p53 to identify PDA cells (green) and with anti-pEIF2α (red). (C) Liver sections from naïve and preimmunized mice were stained with anti-luciferase (green) to identify PDA cells and with anti-pIRE1α (red). (D) Sections from the primary tumor and liver of a patient with PDA were stained with anti-p53 to identify PDA cells (green) and with anti-pIRE1α (red). Photomicrographs are representative of five patients who had no detectable liver metastases. The ratios shown in the top right corners of the photomicrographs indicate the frequency of the observed DCC phenotype relative to the total number of DCCs that were assessed. All frequencies are compiled in table S2. White arrows designate DCCs. Scale bars, 25 µm. (E) Preimmunized mice ( $n = 5$ ) that had received splenic injections of mM1TetXBP1s PDA cells were treated with Dox starting on the day of injection. The number of hepatic DCCs was determined 3 weeks later and compared with that of mice not receiving Dox ( $n = 5$ ). (F) Preimmunized mice that had received splenic injections of mM1TetXBP1s PDA cells were treated ( $n = 10$ ) or not treated ( $n = 8$ ) with Dox beginning 3 weeks later. T cells were depleted by administering antibodies against CD4 and CD8. Growth of hepatic metastases was assessed by whole-body bioluminescence imaging. The dashed gray line represents the background luminescence in tumor-free mice. (G) The mean number of bioluminescent metastases formed with or without Dox treatment was determined in resected livers. \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . Error bars, mean  $\pm$  SEM.



(clone PK136, BioXcell) were injected intraperitoneally 2 times per week.

### Fluorochrome conjugation

For some applications (Ki67 staining in Fig. 1 and CHOP staining in Fig. 5), the anti-p53 antibody was conjugated using Zenon labeling reagent 555 (Life Sciences, # Z25005) following the manufacturer's protocol.

### Chemical chaperone treatment

4-PBA was given to mice in drinking water at 1 g/kg/day.

### Doxycycline treatment

Dox was given to mice in drinking water at 2mg/ml + 5% sucrose

### Plasmid preparation and transfection

For the generation of mMIDTLB PDA cells: Sequences of the genes of diphtheria toxin receptor (D), HSV-TK (T), firefly luciferase (L) and mTagBFP2 (B) were generated by gene synthesis (Invitrogen GeneArt) with appropriate restriction sites. For HSV-TK and firefly luciferase, sequences were digested and migrated with agarose gel electrophoresis. Appropriate size bands were purified with gel extraction kit (Qiagen, #28706) and cloned into the piggyback transposon vector (PB-EF1-MCS-IRES-Neo cDNA Cloning and Expression Vector, System Bioscience #PB533A-2). For mTagBFP2, the sequence was amplified by PCR and cloned into the linearized piggyback vector already containing HSV-TK and firefly luciferase with Gibson assembly master mix (New England Biolabs #E2611L). Plasmids were purified using Zippy plasmid miniprep kit (Zymo research, #D4037). Cells were transfected with the resulting vector and the plasmid of the super piggyback transposase (System Bioscience, #PB210PA-1) with Lipofectamine 3000 (ThermoFisher, #L3000015). Cells were selected for stable transfectant with G418 (Sigma-Aldrich #A1720) and sorted for the highest expression of mTagBFP2 (fig. S1A). The functionality of Luciferase, DTR and HSV-TK was verified in vitro (fig. S1, B and C).

For the generation of mM1TetXBPIs PDA cells: The sequence of spliced XBPI (XBPIs) and mCherry linked by a 2A peptide sequence was generated by gene synthesis (Genscript) with appropriate restriction sites. The sequence was digested and migrated with agarose gel electrophoresis. Appropriate size bands were purified with gel extraction kit (Qiagen, #28706) and cloned into an inducible TetON plasmid (gift of S. Lyons). Plasmids were purified using Zippy plasmid miniprep kit (Zymo research, #D4037). mMIDTLB PDA cells were transfected with the resulting vector with Lipofectamine 3000 (ThermoFisher, #L3000015). Cells were selected for stable transfectant with puromycin (Sigma-Aldrich #P8833) and sorted first for no mCherry expression without doxycycline (Dox) to remove cells with potentially leaking expression and then for the highest expression of mCherry upon Dox treatment to select inducible cells (fig. S13A). Tight inducible over-

expression of XBPIs was confirmed in vitro by Western blot analysis (fig. S13B).

### Generation of preimmunized mice

Mice were anesthetized by isoflurane and dorsal fur was removed with a clipper. A subcutaneous injection of  $10^6$  PDA cells was performed. Tumors were grown for 2 weeks. In order to eliminate the tumors, diphtheria toxin (DTx, List Biologicals, #150) was injected intraperitoneally on 2 consecutive days at a dose of 25 ng/g. GcV was injected intraperitoneally on 2 consecutive days at a dose of 5 mg/kg. In approximately 20% of the cases, one round of DTx and GcV injection was not sufficient to eliminate the tumors, and a second round of injection was performed 1 week later. On rare instances two rounds of injections were insufficient, and these animals were not used. The absence of remaining cancer cells was confirmed by bioluminescence (fig. S1D) and mice were used for experiments at least 3 weeks after the last injection of DTx and GcV.

### Pancreatic cancer metastasis model

For the liver metastasis model, animals were anesthetized by isoflurane, a 1 cm incision was made in the left subcostal region and the spleen was exposed. A suspension of  $10^6$  pancreatic cancer cells in 50  $\mu$ l of PBS was injected into the body of the spleen. Immediately following injection, mice were splenectomized to prevent growing of extra-hepatic tumors. The peritoneum was closed with a 5-0 absorbable surgical suture (Vicryl, Ethicon) and the skin with wound clips (Roboz surgical instruments). For the lung metastasis model  $10^6$  cells in 100  $\mu$ l were injected through the tail vein.

### Bioluminescence imaging

Mice were anesthetized by isoflurane and ventral fur was removed with a clipper. 150  $\mu$ l of a 30mg/ml D-Luciferin K<sup>+</sup> salt (Perkin-Elmer, #122799) solution was injected intraperitoneally. Mice were imaged with an IVIS Spectrum in Vivo Imaging System (Perkin-Elmer) 14 min after the injection.

### Immunofluorescence

Organs were harvested and fixed with PLP buffer (1% PFA, 80mM L-lysine, 10mM NaIO<sub>4</sub>); through portal vein perfusion in the case of livers. Organs were embedded in Tissue-Tek OCT compound (Sakura, #4583) and section of 10  $\mu$ m were cut on a Leica cryostat. Between each of the following steps, sections were washed three times for five min with PBS. Tissue sections were post-fixed for 15 min at room temperature with PLP buffer and permeabilized for 15 min at room temperature with 0.1% Triton X100 in PBS for mouse tissues or 30 min at room temperature with 0.2% Triton X100 in PBS for human tissues. Sections were surrounded with hydrophobic barrier pen (ImmEdge, Vector labs, #H-4000) and unspecific antibody labeling was blocked for 1 hour at room temperature with 10% donkey serum (Jackson immunoresearch, #017-000-121) in PBS for mouse tissues or 10% goat serum (Thermo Fisher, #16210-

064) in PBS for human tissues. Primary antibodies (table S3) were incubated overnight at 4°C in the dark. Secondary antibodies (table S3) were incubated for 2 hours at room temperature in the dark. Nuclei were counterstain with DAPI (Molecular probes, #R37606) for 10 min at room temperature in the dark and sections were mounted with ProLong diamond antifade mountant (Molecular probes, #P36965) for mouse tissues or ProLong gold antifade mountant (Molecular probes, #P36934). Images were acquired using a spinning disk confocal microscope (Perkin-Elmer UltraVIEW VoX, High speed spinning disk Yokogawa CSU-X1) with 20X or 60X objectives. Images were analyzed with ImageJ. CFSE (CellTrace CFSE Cell Proliferation Kit, Molecular probes, #C34554) and EdU (Click-iT Plus EdU Imaging Kit, Molecular probes, #C10640) staining were performed following the manufacturer protocol.

### Flow cytometry analysis and sorting

Cells were harvested with TrypLE (Gibco, #12605036), distributed in 96-well round-bottom plate and centrifuged at 2000 rpm for 1 min at 4°C. Supernatant was removed and the cells incubated in ice cold FACS Buffer (1% FBS and 0.02% Sodium Azide in PBS) with Fc Block (Biolegend clone 93) at 4°C for 15 min. After a wash in FACS Buffer, the cells were incubated with primary and secondary antibodies (table S3) for 15 min at 4°C in the dark. The cells were then washed twice in FACS Buffer and resuspended in 450  $\mu$ l PBS before analysis using a LSR Fortessa (BD Biosciences) operated by a FACSDIVA (BD Biosciences) software. Data analysis was performed on FlowJo 10 (FlowJo, LLC). For sorting, cells were submitted to the same procedure as for flow cytometry analysis and processed using BD FACSaria (BD Biosciences).

### Disseminated cancer cell count

Sections were taken every 100  $\mu$ m through the entire thickness of the liver. The thickness of the liver was measured at the same time. DCCs were identified by luciferase expression and counted in every section. The number of DCCs per whole liver was then calculated.

### scRNA-seq

scRNA libraries were generated from viable Ecad<sup>+</sup> and Ecad<sup>-</sup> tumor cells obtained by flow cytometry sorting. PDA cells were adjusted to 200 cells/ $\mu$ l and applied to the C1 system for single-cell capture with a 10-17 $\mu$ m IFC (Fluidigm). In the C1, whole-transcriptome amplification was performed with the SMARTer kit (Clontech), and the product was converted to Illumina sequencing libraries using Nextera XT (Illumina). RNA-seq was performed on a NextSeq instrument (Illumina) single read 75. Quality control was performed using FastQC and cells with low read numbers were eliminated. Reads were aligned against reference mouse genome (EnsMart72) and transcript expression values were determined after transcript normalization (transcript per million; TPM) with AltAnalyze. Cells with aberrant expression for

Gapdh, Actb and Hprt were eliminated leaving 104 Ecad<sup>+</sup> and 98 Ecad<sup>-</sup> cells for further analysis. Transcripts were considered significantly expressed if TPM  $\geq 1$ . Differential expression analysis was performed considering fold-change between Ecad<sup>+</sup> and Ecad<sup>-</sup> cells  $\geq 2$  fold as a cut-off. Pathway enrichment analysis and network maps were performed with Cytoscape v3.5.1 and the ClueGo v2.3.3 plugin. Pathways with  $P < 0.01$  after Benjamini-Hochberg procedure for false discovery rate were considered significant.

### Statistical analysis

Two-column comparisons were performed with an unpaired  $t$  test. Comparison of three or more columns was performed using a one-way ANOVA followed by Tukey's procedure. A  $P$  value less than 0.05 was considered significant: \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , and \* $P < 0.05$ . All statistical analyses were performed using GraphPad Prism software version 6.

### REFERENCES AND NOTES

- Q. Zhang *et al.*, Pancreatic Cancer Epidemiology, Detection, and Management. *Gastroenterol. Res. Pract.* **2016**, 8962321 (2016). doi: [10.1155/2016/8962321](#); pmid: [26941789](#)
- D. Hariharan, A. Saied, H. M. Kocher, Analysis of mortality rates for pancreatic cancer across the world. *HPB (Oxford)* **10**, 58–62 (2008). doi: [10.1080/13651820701883148](#); pmid: [18695761](#)
- V. P. Groot *et al.*, Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma. *Ann. Surg.* **267**, 936–945 (2018). doi: [10.1097/SLA.0000000000002234](#); pmid: [28338509](#)
- A. Van den Broeck *et al.*, Patterns of recurrence after curative resection of pancreatic ductal adenocarcinoma. *Eur. J. Surg. Oncol.* **35**, 600–604 (2009). doi: [10.1016/j.ejso.2008.12.006](#); pmid: [19131205](#)
- R. Kim *et al.*, PET/CT Fusion Scan Prevents Futile Laparotomy in Early Stage Pancreatic Cancer. *Clin. Nucl. Med.* **40**, e501–e505 (2015). doi: [10.1097/RLU.0000000000000837](#); pmid: [26053713](#)
- C. M. Koebel *et al.*, Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **450**, 903–907 (2007). doi: [10.1038/nature06309](#); pmid: [18026089](#)
- J. Eyles *et al.*, Tumor cells disseminate early, but immunosurveillance limits metastatic outgrowth, in a mouse model of melanoma. *J. Clin. Invest.* **120**, 2030–2039 (2010). doi: [10.1172/JCI42002](#); pmid: [20501944](#)
- M. Müller *et al.*, EblacZ tumor dormancy in bone marrow and lymph nodes: Active control of proliferating tumor cells by CD8+ immune T cells. *Cancer Res.* **58**, 5439–5446 (1998). pmid: [9850077](#)
- Y. Kang, K. Pantel, Tumor cell dissemination: Emerging biological insights from animal models and cancer patients. *Cancer Cell* **23**, 573–581 (2013). doi: [10.1016/j.ccr.2013.04.017](#); pmid: [23680145](#)
- C. A. Klein, Selection and adaptation during metastatic cancer progression. *Nature* **501**, 365–372 (2013). doi: [10.1038/nature12628](#); pmid: [24048069](#)
- M. S. Sosa, P. Bragado, J. A. Aguirre-Ghiso, Mechanisms of disseminated cancer cell dormancy: An awakening field. *Nat. Rev. Cancer* **14**, 611–622 (2014). doi: [10.1038/nrc3793](#); pmid: [25118602](#)
- K. Pantel *et al.*, Frequent down-regulation of major histocompatibility class I antigen expression on individual micrometastatic carcinoma cells. *Cancer Res.* **51**, 4712–4715 (1991). pmid: [1873815](#)
- M. Mohme, S. Riethdorf, K. Pantel, Circulating and disseminated tumour cells - mechanisms of immune surveillance and escape. *Nat. Rev. Clin. Oncol.* **14**, 155–167 (2017). doi: [10.1038/nrclinonc.2016.144](#); pmid: [27644321](#)
- J. Massagué, A. C. Obenauf, Metastatic colonization by circulating tumour cells. *Nature* **529**, 298–306 (2016). doi: [10.1038/nature17038](#); pmid: [26791720](#)
- S. Malladi *et al.*, Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. *Cell* **165**, 45–60 (2016). doi: [10.1016/j.cell.2016.02.025](#); pmid: [27015306](#)
- N. Müller-Hermelink *et al.*, TNFR1 signaling and IFN-gamma signaling determine whether T cells induce tumor dormancy or promote multistage carcinogenesis. *Cancer Cell* **13**, 507–518 (2008). doi: [10.1016/j.ccr.2008.04.001](#); pmid: [18538734](#)
- H. Eyob *et al.*, Inhibition of ron kinase blocks conversion of micrometastases to overt metastases by boosting antitumor immunity. *Cancer Discov.* **3**, 751–760 (2013). doi: [10.1158/2159-8290.CD-12-0480](#); pmid: [23612011](#)
- D. C. Strauss, J. M. Thomas, Transmission of donor melanoma by organ transplantation. *Lancet Oncol.* **11**, 790–796 (2010). doi: [10.1016/S1470-2045\(10\)70024-3](#); pmid: [20451456](#)
- D. C. Martin, M. Rubini, V. J. Rosen, Cadaveric Renal Homotransplantation with Inadvertent Transplantation of Carcinoma. *JAMA* **192**, 752–754 (1965). doi: [10.1001/jama.1965.03080220016003](#); pmid: [14285706](#)
- C. Feig *et al.*, Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti-PD-L1 immunotherapy in pancreatic cancer. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20212–20217 (2013). doi: [10.1073/pnas.1320318110](#); pmid: [24277834](#)
- I. Poschke *et al.*, Identification of a tumor-reactive T-cell repertoire in the immune infiltrate of patients with resectable pancreatic ductal adenocarcinoma. *Oncotarget* **5**, e1240859 (2016). doi: [10.1080/2162402X.2016.1240859](#); pmid: [28123878](#)
- A. A. Connor *et al.*, Association of Distinct Mutational Signatures With Correlates of Increased Immune Activity in Pancreatic Ductal Adenocarcinoma. *JAMA Oncol.* **3**, 774–783 (2017). doi: [10.1001/jamaoncol.2016.3916](#); pmid: [27768182](#)
- D. Balli, A. J. Rech, B. Z. Stanger, R. H. Vonderheide, Immune Cytolytic Activity Stratifies Molecular Subsets of Human Pancreatic Cancer. *Clin. Cancer Res.* **23**, 3129–3138 (2017). doi: [10.1158/1078-0432.CCR-16-2128](#); pmid: [28007776](#)
- A. Maitra *et al.*, Multicomponent analysis of the pancreatic adenocarcinoma progression model using a pancreatic intraepithelial neoplasia tissue microarray. *Mod. Pathol.* **16**, 902–912 (2003). doi: [10.1097/01.MP.0000086072.56290.FB](#); pmid: [13679454](#)
- S. R. Hingorani *et al.*, Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005). doi: [10.1016/j.ccr.2005.04.023](#); pmid: [15894267](#)
- A. D. Rhim *et al.*, Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer Cell* **25**, 735–747 (2014). doi: [10.1016/j.ccr.2014.04.021](#); pmid: [24856585](#)
- N. M. Aiello *et al.*, Metastatic progression is associated with dynamic changes in the local microenvironment. *Nat. Commun.* **7**, 12819 (2016). doi: [10.1038/ncomms12819](#); pmid: [27628423](#)
- I. M. Strommes, A. Hulbert, R. H. Pierce, P. D. Greenberg, S. R. Hingorani, T-cell Localization, Activation, and Clonal Expansion in Human Pancreatic Ductal Adenocarcinoma. *Cancer Immunol. Res.* **5**, 978–991 (2017). doi: [10.1158/2326-6066.CIR-16-0322](#); pmid: [29066497](#)
- S. F. Boj *et al.*, Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015). doi: [10.1016/j.cell.2014.12.021](#); pmid: [25557080](#)
- L. Fang *et al.*, MICA/B expression is inhibited by unfolded protein response and associated with poor prognosis in human hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* **33**, 76 (2014). doi: [10.1186/s13046-014-0076-7](#); pmid: [25228093](#)
- D. P. Granados *et al.*, ER stress affects processing of MHC class I-associated peptides. *BMC Immunol.* **10**, 10 (2009). doi: [10.1186/1471-2172-10-10](#); pmid: [19220912](#)
- S. F. de Almeida, J. V. Fleming, J. E. Azevedo, M. Carmo-Fonseca, M. de Sousa, Stimulation of an unfolded protein response impairs MHC class I expression. *J. Immunol.* **178**, 3612–3619 (2007). doi: [10.4049/jimmunol.178.6.3612](#); pmid: [17339458](#)
- H. J. Clarke, J. E. Chambers, E. Liniker, S. J. Marciniak, Endoplasmic reticulum stress in malignancy. *Cancer Cell* **25**, 563–573 (2014). doi: [10.1016/j.ccr.2014.03.015](#); pmid: [24823636](#)
- H. Vanacker *et al.*, Emerging Role of the Unfolded Protein Response in Tumor Immunosurveillance. *Trends Cancer* **3**, 491–505 (2017). doi: [10.1016/j.trecan.2017.05.005](#); pmid: [28718404](#)
- M. Cuadrado-Tejedor, A. García-Osta, A. Ricobaraza, J. Oyarzabal, R. Franco, Defining the mechanism of action of 4-phenylbutyrate to develop a small-molecule-based therapy for Alzheimer's disease. *Curr. Med. Chem.* **18**, 5545–5553 (2011). doi: [10.2174/092986711798347315](#); pmid: [22172064](#)
- R. Ghosh *et al.*, Allosteric inhibition of the IRE1 $\alpha$  RNase preserves cell viability and function during endoplasmic reticulum stress. *Cell* **158**, 534–548 (2014). doi: [10.1016/j.cell.2014.07.002](#); pmid: [25018104](#)
- N. M. Aiello *et al.*, Upholding a role for EMT in pancreatic cancer metastasis. *Nature* **547**, E7–E8 (2017). doi: [10.1038/nature22963](#); pmid: [28682339](#)
- F. Garrido, N. Aptsiauri, E. M. Doorduijn, A. M. Garcia Lora, T. van Hall, The urgent need to recover MHC class I in cancers for effective immunotherapy. *Curr. Opin. Immunol.* **39**, 44–51 (2016). doi: [10.1016/j.coi.2015.12.007](#); pmid: [26796069](#)
- K. Pantel *et al.*, Differential expression of proliferation-associated molecules in individual micrometastatic carcinoma cells. *J. Natl. Cancer Inst.* **85**, 1419–1424 (1993). doi: [10.1093/jnci/85.17.1419](#); pmid: [7688814](#)
- M. S. Sosa, P. Bragado, J. Debnath, J. A. Aguirre-Ghiso, Regulation of tumor cell dormancy by tissue microenvironments and autophagy. *Adv. Exp. Med. Biol.* **734**, 73–89 (2013). doi: [10.1007/978-1-4614-1445-2\\_5](#); pmid: [23143976](#)
- K. Bartkowiak *et al.*, Disseminated Tumor Cells Persist in the Bone Marrow of Breast Cancer Patients through Sustained Activation of the Unfolded Protein Response. *Cancer Res.* **75**, 5367–5377 (2015). doi: [10.1158/0008-5472.CAN-14-3728](#); pmid: [26573792](#)
- X. X. Li *et al.*, Knockdown of IRE1 $\alpha$  inhibits colonic tumorigenesis through decreasing  $\beta$ -catenin and IRE1 $\alpha$  targeting suppresses colon cancer cells. *Oncogene* **36**, 6738–6746 (2017). doi: [10.1038/onc.2017.284](#); pmid: [28825721](#)
- J. Hollien, J. S. Weissman, Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response. *Science* **313**, 104–107 (2006). doi: [10.1126/science.1129631](#); pmid: [16825573](#)
- C. W. Steele *et al.*, CXCR2 Inhibition Profoundly Suppresses Metastases and Augments Immunotherapy in Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **29**, 832–845 (2016). doi: [10.1016/j.ccell.2016.04.014](#); pmid: [27265504](#)
- Y. Naito *et al.*, Responses of plasma adrenocorticotrophic hormone, cortisol, and cytokines during and after upper abdominal surgery. *Anesthesiology* **77**, 426–431 (1992). doi: [10.1097/0000542-199209000-00004](#); pmid: [1355636](#)
- T. R. Flint *et al.*, Tumor-Induced IL-6 Reprograms Host Metabolism to Suppress Anti-tumor Immunity. *Cell Metab.* **24**, 672–684 (2016). doi: [10.1016/j.cmet.2016.10.010](#); pmid: [27829137](#)
- V. Keim, E. Klar, M. Poll, M. H. Schoenberg, Postoperative care following pancreatic surgery: Surveillance and treatment. *Dtsch. Arztebl. Int.* **106**, 789–794 (2009). pmid: [20038981](#)
- E. E. Embuscado *et al.*, Immortalizing the complexity of cancer metastasis: Genetic features of lethal metastatic pancreatic cancer obtained from rapid autopsy. *Cancer Biol. Ther.* **4**, 548–554 (2005). doi: [10.4161/cbt.4.5.1663](#); pmid: [15846069](#)

### ACKNOWLEDGMENTS

We thank D. Tuveson for providing mM1 and 1242 cells and the KPCy liver tissues. We thank R. Kappagantula for her help in processing patients' information. **Funding:** This work was supported by a Distinguished Scholar Award to D.T.F. from the Lustgarten Foundation, an award from the Cedar Hill Foundation, and 5P30CA45508-29 from NIH-NCI. A.P. was supported by the Philippe Foundation. **Author contributions:** Conceptualization: A.P. and D.T.F. Acquisition of data: A.P., N.A., N.M., Z.L.K., A.G., R.Y., C.A., J.A., M.E., C.A.I.-D., and S.K.L. Analysis of data: A.P., N.A., Z.L.K., A.G., R.Y., J.A., M.E., and D.T.F. Writing (original draft and review and editing): A.P. and D.T.F. **Competing interests:** D.T.F. is a cofounder of Myositis (a company developing cancer immunotherapies) and is on the Scientific Advisory Boards of ITeOS Therapeutics (a company developing immuno-oncology drugs), IFM Therapeutics (a company developing therapies targeting the innate immune system), and Kymab (a company developing therapeutic antibodies). **Data and materials availability:** scRNA-seq data were deposited in GEO with the accession number GSE108811.

### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/eaao4908/suppl/DC1](#)  
Figs. S1 to S15  
Tables S1 to S3

31 July 2017; accepted 13 April 2018  
Published online 17 May 2018  
[10.1126/science.aao4908](#)



## RESEARCH ARTICLE SUMMARY

## NANOTECHNOLOGY

# Improving mechanical sensor performance through larger damping

Swapan K. Roy, Vincent T. K. Sauer, Jocelyn N. Westwood-Bachman, Anandram Venkatasubramanian, Wayne K. Hiebert\*

**INTRODUCTION:** Nano-optomechanical systems (NOMS) are very small resonating mechanical devices that have extraordinary sensitivity. The coupling of mechanical motion to an optical cavity allows the motion to be tracked with femtometer precision. When using NOMS (or their electrical cousin, nanoelectromechanical systems) as a stable frequency reference, tiny force and mass changes can be distinguished by small frequency shifts. This is useful in atomic force microscopy and ultrasensitive mass measurement. For example, improvements in mass sensitivity have enabled the resolution of single molecules and have launched a prospective new paradigm of mechanical mass spectrometry. Any method to improve stability improves the performance of these sensors. If stability could remain the same or improve with more damping, NOMS

ultrasensitivity could be deployed in a damping medium, like air or liquid, greatly enhancing their utility for use as biosensors or gas sensors or in the environment. Better stability could also benefit oscillator clock electronics, which could ultimately improve technologies such as GPS.

**RATIONALE:** The quality factor ( $Q$ ) is the inverse of the damping and indicates how sharp the resonance is in frequency.  $Q$  has been used as a proxy metric for frequency stability. However,  $Q$  only provides half the contribution; the other half comes from how large the resonance signal is compared to noise [the signal-to-noise ratio ( $SNR$ )]. This relationship is known as Robins' model. Although traditionally  $Q$  and  $SNR$  have been assumed to be correlated, we noted that when the resonance conditions

are limited only by intrinsic factors, the  $SNR$  should be inversely proportional to  $Q$ . In this case, stability should be independent of  $Q$ , and stable performance should be maintained in a variety of damping conditions.

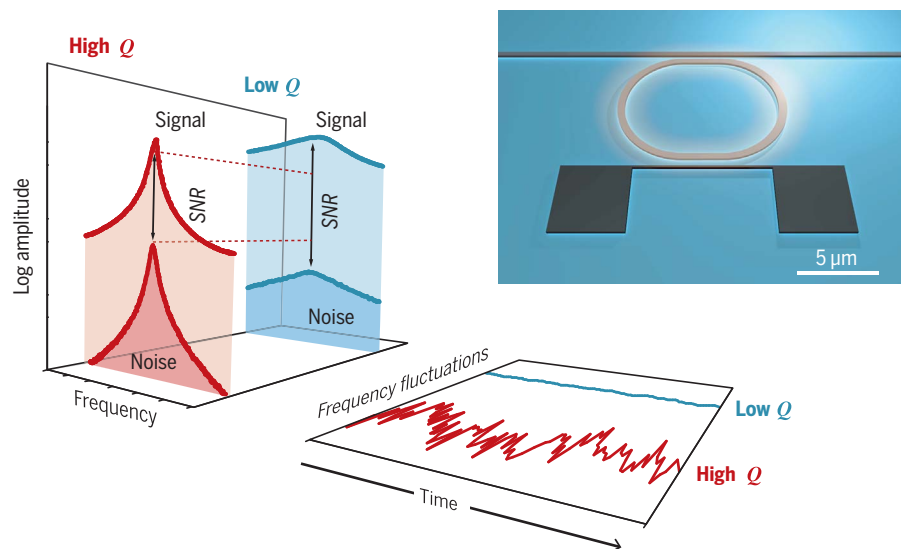
**RESULTS:** We measured intrinsic resonator stability in NOMS while decreasing  $Q$  by increasing the air pressure around the device. We found that  $SNR$  behaved inversely to  $Q$  as hypothesized; however, stability unexpectedly improved with decreasing  $Q$ . This improved performance with damping is diametrically opposed to conventional expectations that had been established for decades. We revisited Robins' model to find that it was based on a high- $Q$  approximation. Rederiving the model without approximation gave rise to a new flatband regime for large damping (low  $Q$ ). In this new regime, stability

## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aar5220>

tracked to  $SNR$  only and, correspondingly, improved with damping, explaining the measured data. We confirmed the improved performance at higher damping by monitoring temperature fluctuations at different pressures and found that the best stability occurred at highest  $SNR$ , consistent with the new model. Finally, there is a noise source called dephasing that is known to prevent mechanical resonators from reaching their stability limits. We confirmed that this extra noise source correlated with  $Q$  and therefore was mitigated at large damping and removed completely at atmospheric pressure.

**CONCLUSION:** Our measurements confirm that  $Q$  and  $SNR$  behave inversely for intrinsically limited resonators, refuting long-standing assumptions about  $Q$  as a stability proxy. More notably, we found stability improved with damping. A low- $Q$  approach was further shown to elegantly solve a vexing stability limitation caused by dephasing. We rederived Robins' model to find a new flatband regime in which stability is linked only to  $SNR$  and found the new model to be consistent with the measured data. The flatband model displayed intriguing properties (in addition to stability behaving inversely to  $Q$ ), including moderation of the trade-off between low noise and bandwidth and a route to frequency-scaling enhancement. The results offer a new paradigm for thinking about stability in mechanical resonators and suggest new pathways to improve stability in resonant sensors and crystal clock oscillators. ■



**Higher  $SNR$  leads to better stability.** A mechanical resonator operating near intrinsic limits (e.g., limited only by thermomechanical noise and driven to the end of its linear regime) has a higher  $SNR$  in a low- $Q$  state.  $SNR$  correlates to stability; frequency fluctuations are higher in the high- $Q$  case. The surprising result is that low- $Q$  devices can make better sensors and oscillators. (Inset) Cartoon of a NOMS chip with a waveguide bus (top), racetrack optical resonator (middle), and mechanical resonant doubly clamped beam (bottom). Beam motion is monitored with high precision through evanescent field interaction with and amplification by the optical cavity.

The list of author affiliations is available in the full article online.  
\*Corresponding author. Email: [wayne.hiebert@nrc-cnrc.gc.ca](mailto:wayne.hiebert@nrc-cnrc.gc.ca)  
Cite this article as S. K. Roy et al., *Science* 360, eaar5220 (2018). DOI: [10.1126/science.aar5220](https://doi.org/10.1126/science.aar5220)

## RESEARCH ARTICLE

## NANOTECHNOLOGY

# Improving mechanical sensor performance through larger damping

Swapan K. Roy<sup>1,2</sup>, Vincent T. K. Sauer<sup>1,3</sup>, Jocelyn N. Westwood-Bachman<sup>1,2</sup>, Anandram Venkatasubramanian<sup>1,3</sup>, Wayne K. Hiebert<sup>1,2,\*</sup>

Mechanical resonances are used in a wide variety of devices, from smartphone accelerometers to computer clocks and from wireless filters to atomic force microscopes. Frequency stability, a critical performance metric, is generally assumed to be tantamount to resonance quality factor (the inverse of the linewidth and of the damping). We show that the frequency stability of resonant nanomechanical sensors can be improved by lowering the quality factor. At high bandwidths, quality-factor reduction is completely mitigated by increases in signal-to-noise ratio. At low bandwidths, notably, increased damping leads to better stability and sensor resolution, with improvement proportional to damping. We confirm the findings by demonstrating temperature resolution of 60 microkelvin at 300-hertz bandwidth. These results open the door to high-performance ultrasensitive resonators in gaseous or liquid environments, single-cell nanocalorimetry, nanoscale gas chromatography, atmospheric-pressure nanoscale mass spectrometry, and new approaches in crystal oscillator stability.

Nanoelectromechanical systems (NEMS) are known for extraordinary sensitivity. Mass sensing has reached the single-proton level (1, 2), enabling NEMS gas chromatography (3–5) and mass spectrometry (6–9). Force sensing has produced single-spin magnetic resonance microscopy (10, 11) and single-molecule force spectroscopy (12). Torque resonance magnetometry has been reenvisioned (13), with applications in spintronics and magnetic skyrmions. The mechanical quantum ground state has even become accessible (14–16) and used for absolute thermometry (17). The best sensitivities, however, have generally been presumed to require the highest quality factors, limiting application to vacuum environments and low temperatures. A host of new applications could result from the availability of ultrasensitivity in air and liquid: biosensing, security screening, environmental monitoring, and chemical analysis. As an example, our group aims, over the long term, to combine mass spectrometry and gas chromatography functions into one via NEMS sensing in atmospheric pressure.

Exquisite NEMS sensitivity is enabled through ultrasmall mass or stiffness combined with precise resonant frequency  $f$  determination. This approach allows perturbations to that frequency  $\delta f$  (such as mass or force) to be probed (see Fig. 1A). Intuitively, the peak position (and ultimate sensitivity) is obscured by two factors: a finite

linewidth and noise on the signal. The former is quantified by resonant quality factor  $Q$  and the latter by signal-to-noise ratio  $SNR$ . Frequency stability should thus relate to these two quantities. Robins' formula (18), articulated in the atomic force microscope (AFM) community by Rugar and co-workers (19) and in NEMS by Roukes and co-workers (20, 21), forms the basis for force and mass sensitivity analyses and defines this relationship as follows:

$$\left\langle \frac{\delta f}{f} \right\rangle \sim \frac{1}{2Q} \frac{1}{SNR} = \frac{1}{2Q} 10^{-DR/20} \quad (1)$$

where  $SNR$  is the ratio of driven motional amplitude to equivalent noise amplitude on resonance

$$SNR = \frac{a_{\text{driven}}}{a_{\text{noise}}} \quad (2)$$

and the dynamic range  $DR$  is the power level associated with this  $SNR$ .

The  $Q$  factor in the denominator of Eq. 1 has led researchers to focus primarily on achieving higher  $Q$  for better resolution (22–26). Strategies for  $SNR$  enhancement have been scarcer (27), with almost no consideration given to relationships between  $Q$  and  $SNR$ . There is an implicit assumption that improving  $Q$  will also benefit the signal fidelity. Figure 1B presents a  $DR$  response to changing  $Q$  from this traditional view. An extrinsic noise floor (e.g., readout amplifier) sets a noise that does not change with  $Q$ . Drive power is also assumed to be unchanged, resulting in amplitude loss, and shrinking  $DR$ , for lower  $Q$ .

There is, however, a case when  $SNR \propto 1/Q$  that results in no sensitivity dependence on  $Q$ . This is not a special case. Indeed, it is the gen-

eral case if the  $DR$  is properly maximized. When instrument noise is negligible (Fig. 1B, right), lower- $Q$  resonances reveal fundamentally lower intrinsic noise floor peaks (e.g., thermomechanical noise). The intrinsic upper end of the dynamic range is associated with the end of linear response. The wider linewidth of a lower- $Q$  resonance tolerates more nonlinearity and extends this linear range to larger amplitude. Combined, the two effects give  $10^{-DR/20} \propto Q$ .

This peculiar observation implies that frequency-fluctuation noise should not depend on  $Q$  in the case when thermomechanical noise is well resolved and amplitude can be driven to nonlinearity. If true, the model provides a pathway to completely mitigate sensitivity loss due to low  $Q$ . Even more intriguing, a detailed inspection of the phase noise model used in NEMS (19–21) reveals that Eq. 1 is a high- $Q$  approximation. Removing this approximation, remarkably, implies frequency-fluctuation noise proportional to  $Q$  at low bandwidth: A highly damped system with full dynamic range should have better frequency stability (and sensitivity) than an equivalent lowly damped one. This is an exciting prospect with wide-ranging implications for (frequency modulated) scanning probe microscopy, mass sensing and biosensing, and inertial and timing microelectromechanical systems (gyroscopes, accelerometers, and crystal oscillators).

Using nano-optomechanical systems (NOMS), we demonstrated frequency stability improving with increased damping. We changed pressure from vacuum to atmosphere to vary the extrinsic  $Q$  within a single nanomechanical device. We observed that the  $SNR$  grew inverse proportionally to  $Q$  while the full dynamic range was maintained. Corresponding frequency stability (Allan deviation) dropped with increased damping and approached the theoretical limit, and was better in atmosphere than in vacuum. Notably, excess intrinsic frequency-fluctuation noise [also known as decoherence (23, 24, 28–32)] shrank with falling  $Q$  and did not limit stability at atmospheric pressure. We tested implied sensitivity improvement with temperature measurements, using the optical ring as calibration, and showed 60- $\mu$ K sensitivity at 300-Hz bandwidth. This performance is comparable to state-of-the-art sensitivity (33–36), even with the modest nanocalorimeter geometry of a doubly clamped beam. These results will allow proliferation of high-performance ultrasensitive resonators into gaseous and liquid environments.

## Maximizing dynamic range to minimize frequency fluctuations

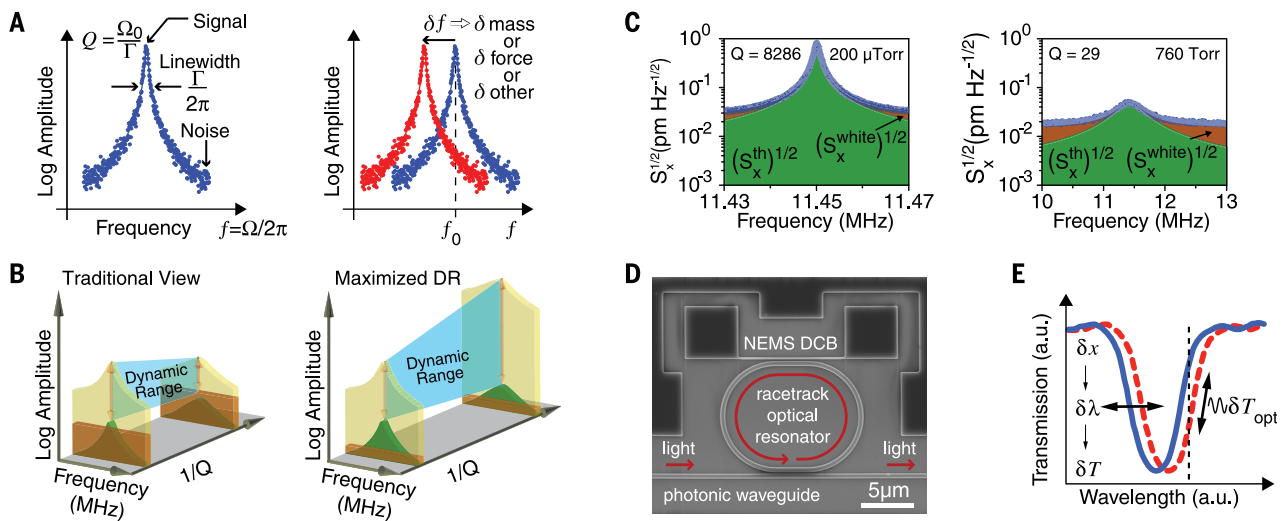
Thermomechanical noise was identified early on as the primary limit for AFM force detection (19). In contrast to macroscale mechanical resonators (such as quartz-crystal oscillators), the smaller stiffness ( $k$ ) and size of AFM beams result in nonnegligible fluctuations from the thermal bath. In essence,  $\frac{1}{2}k_B T$  of thermal energy (where  $k_B$  is the Boltzmann constant and  $T$  is temperature) populates  $\frac{1}{2}k\langle x^2 \rangle$  of modal energy, producing between picometer and nanometer

<sup>1</sup>Nanotechnology Research Centre, National Research Council of Canada, Edmonton, Alberta T6G 2M9, Canada.

<sup>2</sup>Department of Physics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada. <sup>3</sup>Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada.

\*Corresponding author. Email: wayne.hiebert@nrc-cnrc.gc.ca

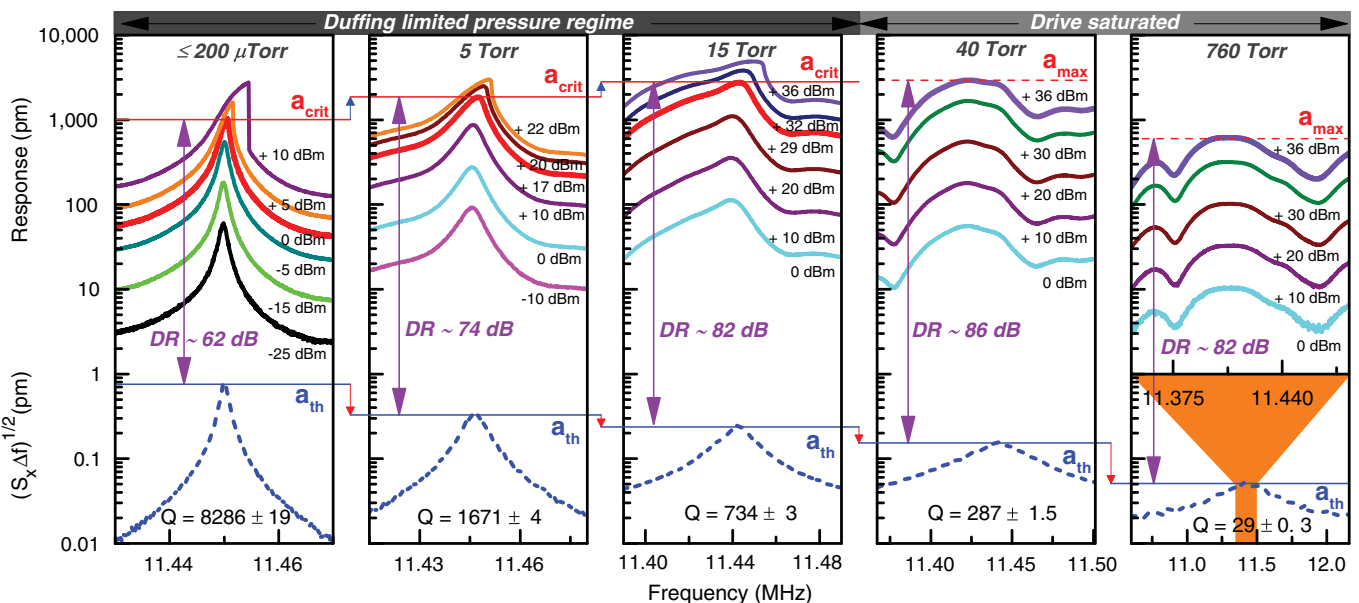




**Fig. 1. Frequency-shift sensing, improvement of dynamic range by smaller  $Q$ , and experimental setup for nano-optomechanical systems.**

(A) Concept of frequency-shift sensing: A mechanical resonance is perturbed by change in mass or force resulting in a frequency shift. At a first approximation, the minimum detectable shift improves with sharper linewidth and higher SNR, as per Robins' formula (Eq. 1). (B) Concept of maximizing linear dynamic range. In the traditional view (left), the DR (arrows) extends from the noise floor to the driven resonance peak (beige Lorentzian-peak shape). The instrumentation noise floor (brown rectangle) often obscures the thermomechanical noise floor (green Lorentzian-peak shape). Decreasing  $Q$  leads to a loss in system DR. In the maximized DR case (right), well-resolved thermomechanical noise leads to a drop in peak noise value during increased damping; simultaneously, the upper end of the linear range becomes higher as nonlinearity sets in at higher amplitude. System DR grows on both ends with falling  $Q$ . (C) Displacement noise  $S_x^{1/2}$  (blue) of the

doubly clamped silicon beam ( $9.75 \mu\text{m}$  by  $180 \text{ nm}$  by  $220 \text{ nm}$ ) shown in (D). Left graph is at high  $Q$  measured in vacuum; right graph is at low  $Q$  measured at atmospheric pressure. The green fit  $(S_x^{\text{th}})^{1/2}$  is resolved out of the orange (white noise) background  $(S_x^{\text{white}})^{1/2}$  near resonance. The white noise is the instrumentation noise floor (the displacement imprecision) set primarily by the photodetector dark current. The peak noise value is suppressed at lower  $Q$ . (D) Annotated scanning electron microscopy image of the NOMS device. A mechanically released doubly clamped beam (NEMS DCB) is adjacent to a racetrack optical cavity and bus photonic waveguide, all patterned in 220-nm-thick silicon-on-insulator. (E) Concept of the optical cavity resonance shift caused by mechanical beam motion. The solid blue line is cavity resonance, and the dashed red line is shifted cavity resonance. Oscillation in displacement  $\delta x$  of the mechanical beam modulates the optical resonance wavelength  $\delta \lambda$ , which, when probe light is situated on the side slope of the cavity, transduces to transmission modulation  $\delta T_{\text{optical}}$ . a.u., arbitrary units.



**Fig. 2. Dynamic range is pressure dependent.** Dashed lines (blue) are measured thermomechanical-noise frequency curves presented in a  $\Delta f = 1\text{-Hz}$  bandwidth;  $a_{\text{th}}$  is their peak value, which falls with increasing pressure. Solid lines are driven-response frequency curves for various drive

powers;  $a_{\text{crit}}$  (peak values of the thicker red lines) grows with increasing pressure in the Duffing-limited pressure regime. For the 760-torr case, the driven frequency axis is zoomed in (orange) with respect to the thermomechanical noise at the same pressure.

average displacements ( $\langle x \rangle$ ). These motion levels have been resolvable since the early 1990s. Mass-detection NEMS (20, 21) are stiffer and, although resolving picometer average displacements had been a challenge (37–40), new transduction techniques (38, 41, 42) have made it routine (14, 15, 17, 28, 29, 42–58).

The use of NOMS, in particular (14, 15, 17, 24, 29, 42–46, 48, 49, 51–62), has resolved thermomechanical noise by orders of magnitude above the instrumentation noise background. One example is our microring cavity optomechanical system (58–62), with displacement imprecision of approximately  $20 \text{ fm Hz}^{-1/2}$ . Figure 1C shows the measured displacement noise  $S_x^{1/2}$  in an example doubly clamped beam, measured in vacuum, where  $Q$  is high, and at atmospheric pressure, where  $Q$  is low. The device is shown in Fig. 1D, with the principle of detection illustrated in Fig. 1E. As per convention, values for  $S_x$  were calibrated from voltage signals ( $S_V$ ) by assuming the peak noise relation (derived via equipartition theorem):

$$S_x^{\text{th}}(\Omega_0) = \frac{4k_B T}{M\Omega^2 \Gamma} \quad (3)$$

where angular frequency  $\Omega = \Omega_0 = 2\pi f$  and  $f$  is the resonance frequency,  $M$  is the resonator effective mass, and  $\Gamma = \Omega/Q$  is the damping. We define the thermomechanical (th) noise amplitude on resonance  $a_{\text{th}}$  as

$$a_{\text{th}} = \sqrt{S_x^{\text{th}}(\Omega_0)\Delta f} = \sqrt{\frac{4k_B T Q}{M\Omega^3} \Delta f} \quad (4)$$

where  $\Delta f$  is the measurement bandwidth. Details are in (63), section 1.2. Note that  $a_{\text{th}}$  is proportional to  $Q^{+1/2}$ ; correspondingly, the high- $Q$  peak of Fig. 1C is higher. In both cases, the noise is dominated by the thermomechanical term near resonance, flattening to a white background far from resonance. In this way, we have access to the bottom end of the intrinsic  $DR$  out to 30-kHz measurement bandwidth.

Our devices were mechanically driven with a shear piezo (fig. S1), and a nonlinear response was reached owing to large drive power. As the doubly clamped beam was driven to larger amplitudes, the stiffness became amplitude dependent, resulting in a geometric nonlinearity (64–66). This Duffing nonlinearity created shark-fin-shaped resonance traces (e.g., Fig. 2, first panel, top trace) and amplitude-dependent resonance frequency. To avoid injecting amplitude noise into phase noise (which reduces stability), the driven amplitude must remain at or below a critical value  $a_{\text{crit}}$  that is generally used to define the end of the linear range and the top of the intrinsic  $DR$  (66):

$$a_{\text{crit}} = \frac{2(0.745)}{\pi} f_0 L^2 \sqrt{\frac{\rho \sqrt{3}}{QE}} \quad (5)$$

where  $L$  is the beam length,  $E$  is the Young's modulus, and  $\rho$  is the density [a version of the equation including tension (66) is in (63)]. It is known that  $a_{\text{crit}}$  has an inverse square root

dependence on  $Q$  that comes straight out of a Duffing derivation and applies to all nonlinear resonators (64–66). Intuitively, at a given amplitude, intrinsic nonlinearity causes a defined frequency shift large enough to tilt the resonance shape for a narrow linewidth while still being hidden by a wider one. When the full dynamic range was accessed, we could equate  $a_{\text{noise}}$  to  $a_{\text{th}}$  and  $a_{\text{driven}}$  to  $a_{\text{crit}}$ , and Eqs. 2, 4, and 5 combine to produce  $SNR$  proportional to  $1/Q$ .

To test  $SNR$  behavior, we measured properties of the same doubly clamped beam at different pressures (thus different  $Q$ s), from vacuum to atmospheric pressure. This approach kept all parameters except for  $Q$  identical. Results are presented in Fig. 2, with frequency sweeps for five representative pressures. At each pressure, the thermomechanical noise is plotted for a 1-Hz bandwidth along with the driven root mean square amplitude response for varying drive power. Indicated by thick red lines are traces for the drive power corresponding with Duffing critical amplitude (up to 15 torr) and, by thick purple lines, for the maximum driving power available (40 and 760 torr). For 15-torr pressures and up, the driven resonance line shape was distorted. This distortion was not caused by nonlinearity (note the conserved response shape, and see fig. S8); rather, the resonance broadened to the point where piezo drive efficiency was no longer a constant function of frequency (67, 68). The distorted features are related to bulk acoustic

resonances in the piezo-chip system. This distortion carries no information about the NEMS beam resonance [see (63), section 1.5, and fig. S7].

We note in Fig. 2 that the peak of the noise floor  $a_{\text{th}}$  diminished as the pressure increased and generally followed  $a_{\text{th}} \propto Q^{1/2}$  (compare Eq. 4). This result can be conceptually understood in the following way: The area under the thermomechanical resonance curve is conserved for a given temperature (in proportion to  $k_B T$ ); as the width of the curve increases ( $Q$  decreases), the peak value must fall to compensate. For the upper end of the  $DR$ , within the Duffing-limited pressure regime,  $a_{\text{crit}}$  increased in proportion to  $Q^{-1/2}$ , as predicted by Eq. 5. Accounting for both effects,  $SNR \propto 1/Q$ , up to 15-torr pressure. At 40 torr and up, we no longer had sufficient drive power to reach the Duffing critical amplitude and could no longer take advantage of the full intrinsic  $DR$  of the system. Nonetheless, the dynamic range was still higher at atmospheric pressure than in vacuum.

Figure 3 plots the peak amplitudes  $a_{\text{crit}}$  and  $a_{\text{max}}$ , the thermal amplitude  $a_{\text{th}}$ , quality factor  $Q$ ,  $SNR$ , and product of  $Q \times SNR$  as a function of pressure. We saw that  $Q \times SNR$  was conserved within the Duffing-limited regime. According to Robins' picture (Eq. 1), the frequency fluctuations in our system should be independent of  $Q$ , up to 15 torr.

### Frequency-fluctuation measurements (Allan deviation)

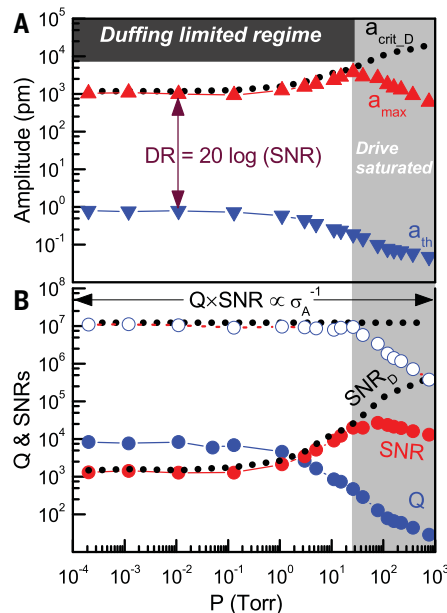
With  $Q \times SNR$  conserved, checking the fractional frequency stability  $\delta f/f$  in our device remained. We did this using the two-sample Allan variance, a standard method of characterizing frequency stability (69) related to Robins' formula [see (63), section 2.2]. The Allan deviation  $\sigma(\tau)$ , as the square root of the Allan variance, is an estimate of fractional-frequency stability for a given time  $\tau$  between frequency readings. The functional form for  $\sigma(\tau)$  is

$$\sigma_R(\tau) = \frac{1}{4Q} \frac{1}{SNR} \frac{1}{\sqrt{\Delta f}} \frac{1}{\sqrt{\tau}} \quad (6)$$

Figure 4 presents the measured Allan deviation data for our device at the five representative pressures and  $Q$ s. Surprisingly, rather than staying constant,  $\sigma$  improved as the pressure increased, and  $Q$  fell (up to 40-torr pressure). Furthermore, the measured data dipped well below the theoretical minimum set by Robins' formalism and Eq. 6 (blue lines). Frequency stability, and therefore performance, appeared to be correlated to  $SNR$  alone.

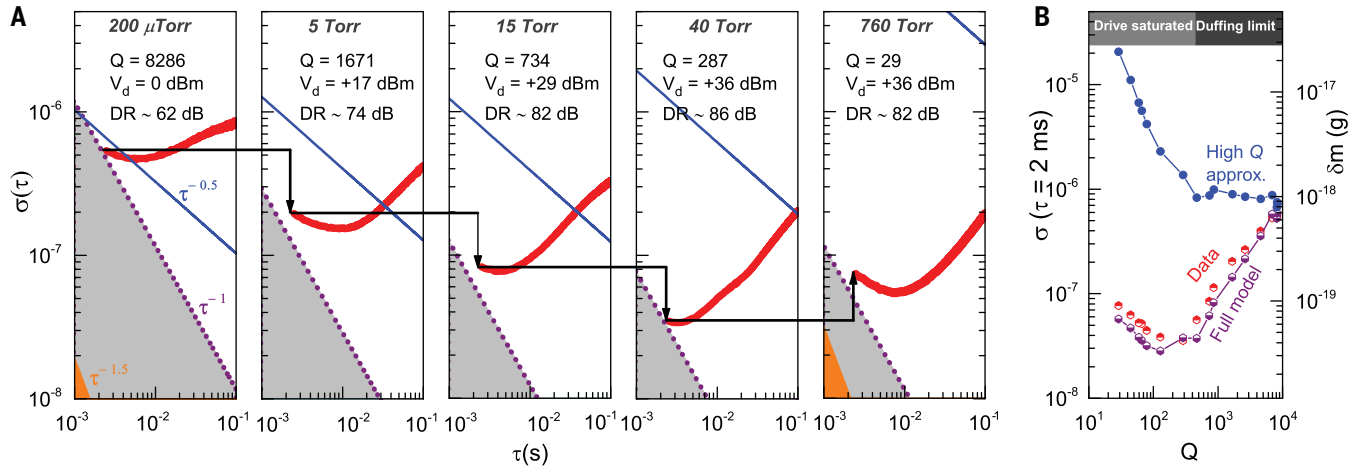
### Full analysis of Allan deviation from noise power

We could understand this response by revisiting the close connection between Allan deviation and phase noise (69). Allan variance  $\sigma^2$  is essentially an integration of close-in phase noise  $S_\phi(\omega)$  out to a measurement bandwidth  $\Delta f$ . Here sideband angular frequency  $\omega = 2\pi f_{\text{mod}}$ , where frequency-offset-from-carrier  $f_{\text{mod}} = f - f_0$ . The resulting



**Fig. 3. The product of  $Q \times SNR$  is constant in the Duffing-limited regime.** (A) Peak frequency-curve amplitude response versus pressure.  $a_{\text{crit,D}}$  is the theoretical Duffing amplitude defined by Eq. 5,  $a_{\text{max}}$  is the measured peak amplitude, and  $a_{\text{th}}$  is the thermomechanical peak amplitude. The  $DR$  grows with pressure. (B)  $Q$  and  $SNR$  versus pressure ( $P$ ).  $SNR_D$  is  $a_{\text{crit,D}}/a_{\text{th}}$ , and  $SNR$  is  $a_{\text{max}}/a_{\text{th}}$ .  $\sigma_A$ , theoretical Allan deviation.





**Fig. 4. Allan deviation  $\sigma_A$  falls (improves) with falling  $Q$ .** (A) Allan deviation (red line) versus sampling time at different pressures. Data are taken with a 4-kHz demodulation bandwidth and collected while tracking frequency in a 500-Hz phase-locked loop (PLL). The 4-kHz demodulation represents the integration bandwidth for the noise ( $\Delta f$ ), whereas 500 Hz sets the bound above which the PLL begins to attenuate fluctuations, effectively setting a minimum meaningful  $\tau$ . Details of the Zurich lock-in amplifier and PLL settings can be found in (63), section 1.6. The blue line is the theoretical

minimum from Eq. 6, whereas the dotted purple line with shading is the theoretical minimum from Eq. 10. The orange-shaded region (panels 1 and 5) is the theoretical minimum set by instrumentation noise floor (eq. S12).  $V_d$ , drive power. (B) Allan deviation at 2-ms sampling time versus  $Q$ . High- $Q$  approximation corresponds to Eq. 6, full model to Eq. 10, and data to the experimentally measured values. In the Duffing-limited regime, the data and full model are proportional to  $Q$ . In both regimes, the data reach close to the fundamental limit of Eq. 10.  $\delta m$ , theoretical mass sensitivity.

Allan deviation  $\sigma$  is proportional to  $\langle S_\phi \times \Delta f \rangle^{1/2}$ , where the  $\langle \rangle$  brackets here loosely represent the integration.

Understanding the frequency stability then reduces to understanding the behavior of  $S_\phi$ . In a Robins' formalism, displacement-derived phase noise  $S_\phi^x$  is related to  $S_x$  via normalization by driven energy:

$$S_\phi^x = \frac{1}{2} \frac{S_x}{a_{\text{driven}}^2} \quad (7)$$

In essence, displacement fluctuations obscure the zero-crossing (phase) during the oscillation cycle by the ratio of noise amplitude to driven amplitude (fig. S13). Full details are available in (63), section 2, and fig. S14. Displacement noise amplitude  $S_x$ , such as in Fig. 1C, turns into a low-pass filter with  $1/f^2$  roll-off when zoomed in near the resonance frequency (compare Fig. 5B):

$$S_x(\omega) = S_x(0) \frac{(\Gamma/2)^2}{\omega^2 + (\Gamma/2)^2} \quad (8)$$

Combining Eqs. 2, 4, 7, and 8 gives

$$S_\phi^x(\omega) = \frac{1}{SNR^2} \left( \frac{1}{2\Delta f} \right) \frac{(\Gamma/2)^2}{\omega^2 + (\Gamma/2)^2} \quad (9)$$

The analysis follows closely to previous Robins' analyses (19–21). At this point, the assumption is generally made that  $\omega^2 + (\Gamma/2)^2 \approx \omega^2$ , that is, that  $Q$  is high. This assumption turns Eqs. 8 and 9 from low-pass filters into pure roll-offs (see Fig. 5A and inset). Knowing that  $S_x(0) \propto 1/\Gamma$  (compare Eq. 3), it is concluded that  $S_\phi^x \sim S_x \sim \Gamma^{-1}$ ,

and, ultimately, that  $\sigma \propto \Gamma^{+1/2}$ . This result is generally well known in the AFM community.

When the high- $Q$  assumption is not made, the proportionality of  $S_x$  and  $S_\phi^x$  with  $\Gamma$  can vary considerably for different  $f_{\text{mod}}$ . Figure 5B shows our experimentally measured values of  $S_x(\omega)$  fit directly with Eq. 8. At high  $f_{\text{mod}}$ ,  $S_x \propto \Gamma^{-1}$ , as in Fig. 5A. For low  $f_{\text{mod}}$ , however,  $S_x \propto \Gamma^{-1}$ . When integrating to get  $\sigma^2$ , the high- $Q$  assumption overestimates the integration, needlessly adding the area between the flat pass and the  $f^{-2}$  line (blue shaded area in inset of Fig. 5A). This overestimate is negligible for large bandwidth ( $\Delta f$ ) but dominates for small bandwidth, leading to overestimated Allan variance.

The difference becomes more intriguing with increasing driven amplitude. Figure 5, C and D, shows Fig. 5B data in  $S_\phi^x(\omega)$  form (via Eq. 7). In Fig. 5C, for 15- and 40-torr pressures,  $a_{\text{driven}}$  had the same value,  $S_x$  and  $S_\phi^x$  maintained the same relation, and the noise dependence on damping was the same as in Fig. 5B. On the other hand, in Fig. 5D,  $a_{\text{driven}}$  was Duffing limited (that is,  $\langle a_{\text{driven}}^2 \rangle \propto \Gamma^{+1}$ ), causing  $S_\phi^x$  to shrink more quickly with damping than  $S_x$  does. This resulted in  $S_\phi^x$  being independent of  $\Gamma$  for large bandwidths and proportional to  $1/\Gamma^2$  at small bandwidths (compare Eq. 9). The right-hand portions of the data at different pressures and damping collapsed on top of each other. This correspondence is not a coincidence but rather the signature of  $SNR$  being inversely proportional to  $Q$ , resulting in no  $\Gamma$  dependence by Robins' equation (Eqs. 1 and 6).

However, for  $\Delta f \leq 1$  kHz, Fig. 5D implies  $\sigma^2 \propto \Gamma^{-2}$ , and therefore  $\sigma \propto \Gamma^{-1}$ . That is, better stability results from more damping. The full functional form of  $\sigma$  for this case, which we

refer to as the flatband (fb) regime, is [derived in (63), section 2]

$$\sigma_{\text{fb}}(\tau) = \left( \frac{3}{2} \right)^{1/2} \frac{1}{SNR} \frac{1}{\Omega \tau} \quad (10)$$

This regime is not usually considered, as it would normally result in prohibitively low bandwidths. However, as devices reach higher frequencies, and as  $Q$  is pushed purposefully down, the corner frequency of  $(\Gamma/2)/(2\pi)$  can become very large; in the present case, it is 200 kHz for atmospheric pressure.

What Eqs. 9 and 10 explicitly show is that flatband  $S_\phi$  and  $\sigma$  should correlate to  $SNR$  alone, not  $Q$ . We qualitatively confirmed this behavior with direct measurement of phase noise in fig. S19, where  $S_\phi$  decreased for increasing  $DR$ . Returning to Fig. 4, Eq. 10 sets the new theoretical lower limit. Here  $\sigma$  was dominated by drift at  $\tau = 0.1$  s but approached Eq. 10 at  $\tau = 2$  ms. Figure 4B plots  $\sigma(2 \text{ ms})$  versus  $Q$  with both theoretical minimum floors (Eqs. 6 and 10), showing close approach to the full model. Within the Duffing-limited regime, where  $SNR \propto 1/Q$ , the measured  $\sigma$  was proportional to  $Q$ , and stability became better in proportion to the amount of damping.

#### Application of damping improved stability: Temperature sensing

We demonstrate an application of damping-enhanced sensitivity, with NEMS temperature resolution improving with increasing pressure. NEMS can be used as a thermometer owing to subtle temperature changes to Young's modulus and device dimensions (33–36). Traditionally in the range of  $\sim 50$  parts per million (ppm)/K for silicon, intrinsic tension changes give our devices

enhanced temperature coefficients with resonant frequency (TCRF) as high as  $\sim 1200$  ppm/K [see (63), section 1.7, table S1, and (33)]. The optical microring cavity itself also had a resonance dependence on temperature, allowing its use

for calibration. The temperature responsivities for both microring and NEMS were simultaneously determined at each pressure tested by monitoring the change in resonant wavelength and mechanical resonant frequency for several

1-K temperature steps [see (63), section 1.7, and fig. S11].

Figure 6A shows the NEMS response at 3-torr pressure to a  $+0.3$ -K step change (followed later by a  $-0.3$ -K step change) in the temperature controller setting. The oscillations and long settling resulted from the PID (proportional integral differential) controller settings. The microring had a similar response (fig. S12). Zoom-ins of frequency (temperature) fluctuations for different pressures in Fig. 6B revealed a sensitivity sweet spot at 60 torr, where  $DR$  was maximum. Peak-to-peak fluctuations improved by one order of magnitude from vacuum to 60 torr, dropping to submillikelvin values. More formally, Fig. 6C presents the temperature resolution  $\sigma_{\Delta T}$  as a function of  $Q$ , where  $\sigma_{\Delta T}(\tau) = \sigma(\tau) \times f_0/S_{fT}$ , and  $S_{fT}$  is the NEMS temperature responsivity (table S1). Temperature resolution improved with falling  $Q$  to the 60-torr sweet spot, where it reaches  $60 \mu\text{K}$ . This performance is comparable to that in (33, 34).

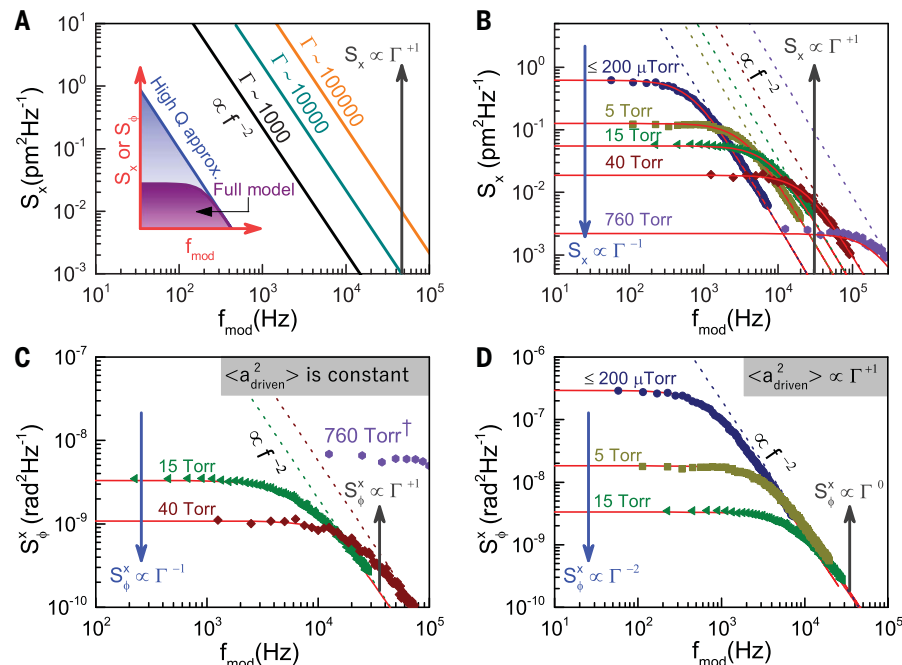
## Discussion

That resolution could be independent of  $Q$  in the Robins' picture has been hinted at (28, 29), but not tested, and is not widely appreciated in the NEMS community. The further revelation that low-bandwidth sensitivity and stability actually improve with damping is a momentous development with implications in NEMS, AFM, crystal clock oscillators, and other applications. As an example, the AFM community has long known of force noise proportional to square root damping and has tried to reduce the apparent thermal-force noise off resonance by increasing  $Q$ . This approach works for high bandwidth (above the corner) but increases noise on resonance, which is usually truncated and ignored. Instead, by purposefully suppressing  $Q$ , one simultaneously suppresses close-in noise while extending the corner frequency (and bandwidth). In essence, the usually inevitable trade-off between bandwidth and low noise is eliminated.

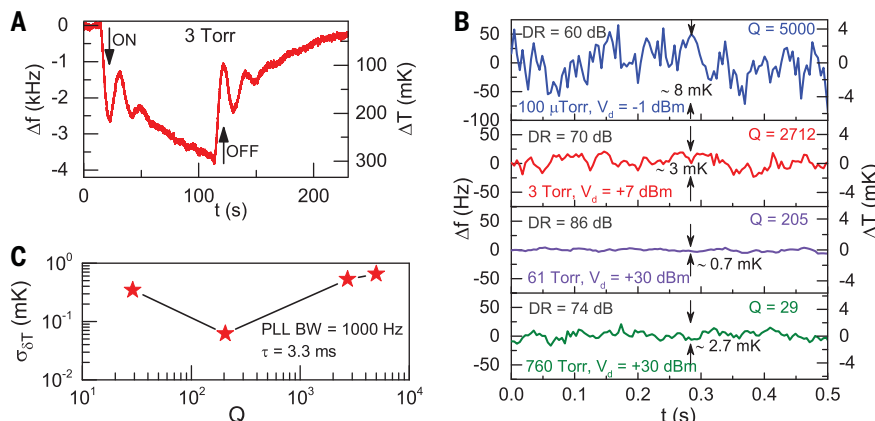
The flatband regime stability (Eq. 10) has no explicit dependence on  $Q$ ; our results derive from improving the  $DR$  with damping. The prevalent idea that stability improves with  $Q$  arises partly from the assumption that high  $Q$  and high  $DR$  are equivalent (compare Fig. 1B, left). Intrinsic noise-resolved and low-power resonators invert this assumption.

It is known that Eq. 6 has no explicit  $\Omega$  dependence. Equation 10, on the other hand, varies inversely with  $\Omega$ , opening additional paths to sensitivity improvement. Increasing the mechanical frequency should directly improve flatband sensitivity, while also extending the bandwidth available for a given  $Q$ . These enhancements are in addition to simultaneous sensitivity improvements coming from mass reduction.

The flatband suppression of the thermal-noise peak is reminiscent of cold damping and feedback cooling (14, 15, 70) but is distinct in that thermal noise is spread out rather than reduced. As such, feedback cooling could give cumulative



**Fig. 5. Noise-power behavior with respect to damping can be proportional, constant, inversely proportional, and inversely quadratic.** (A) Concept of thermomechanical displacement noise being proportional to damping for pure roll-off. Inset, pure roll-off (high- $Q$  approximation) overestimates integration in comparison to low-pass full model. (B) Measured thermomechanical noise fit to Eq. 8; noise is proportional to damping above the roll-off and inversely proportional below. (C) Measured displacement noise converted to phase noise (Eq. 9) with constant driven amplitude; noise is proportional to damping above the roll-off and inversely proportional below. †760 torr is at a lower driven amplitude than 40 and 15 torr and is included here for completeness. (D) Measured displacement noise converted to phase noise with squared driven amplitude proportional to damping; noise is independent of damping above the roll-off and inversely quadratic below.



**Fig. 6. NEMS thermometry.** (A) NEMS tracking of a  $0.3$ -K underdamped-controlled temperature change turned on at about 15 s and turned off at about 115 s. (B) Measured frequency (temperature) fluctuations in a  $0.5$ -s time window for different pressures and  $Q$ . Data have a  $1$ -kHz PLL bandwidth (BW) and are shown at a  $225$ -Hz sampling rate. Arrows indicate the maximum fluctuation over the full half-second. Higher  $DR$  correlates with improved stability. (C) Temperature resolution (red stars) as a function of  $Q$  for a  $1$ -kHz PLL bandwidth and  $3.3$ -ms sampling time. Black lines are guides for the eye.



benefit with the flatband technique. Similarly, techniques for using the nonlinear regime (71, 72) or parametric squeezing (73) can be piggybacked onto flatband.

Another side benefit of low  $Q$  is suppression of intrinsic resonator frequency-fluctuation noise (23, 24, 28–32). Sansa *et al.* (28) recently noted this noise as ubiquitous in preventing NEMS from reaching thermal limits, though Gavartin *et al.* (29) were able to mitigate it with sophisticated force feedback. The transfer function responsible for conveying this intrinsic noise is proportional to  $Q$  (24), which may help explain why we did not see it at atmospheric pressure and saw clear evidence of it only at long gate times in vacuum (figs. S17 to S19).

We note that the limitations of our drive power keep us from accessing the full dynamic range at atmospheric pressure. This problem can be solved by using optomechanical drive force, which can be turned up almost with impunity. Nonlinearities in the optomechanical transduction, in both readout and excitation, could eventually limit the present technique from extending dynamic range indefinitely.

## Methods summary

A full description of the materials and methods is available in the supplementary materials (63).

The optomechanical system was fabricated on a silicon photonics platform from silicon-on-insulator. A doubly clamped beam (9.75  $\mu\text{m}$  by 220 nm by 160 nm) was coupled to a race-track optical cavity resonator and probed with a tunable diode laser around 1550 nm. Mechanical excitations were made via shear piezo. The device chip resided in a vacuum chamber with controllable pressure. Other electronics supporting the measurement included lock-in amplification, high-power excitation amplification, temperature control, computerized data acquisition, and digitally implemented phase-lock techniques (section 1.1).

The thermomechanical noise calibration was implemented in a standard way, allowing voltage noise conversion to displacement noise (section 1.2), and the instrumentation noise floor was calibrated to be around 20 fm  $\text{Hz}^{-1/2}$  (arising primarily from photodetector dark current). The onset of nonlinearity and calculation for critical amplitudes is defined in section 1.3. At high drive power, heating needed to be accounted for in calibrating the critical amplitude (section 1.3.1). Notes on high drive powers (section 1.3.2) and different drive levels (section 1.3.3) discuss the piezo technique inefficiency and the inherent energy storage advantage of a lower- $Q$  device.

Optomechanical coupling was both simulated and calibrated from optomechanical spring effect (section 1.4), giving similar values of  $g_{\text{om}} = 2.8 \text{ rad GHz/nm}$ . Section 1.5 discusses bulk acoustic interference effects of the piezo-chip system and squeeze-film effects.

A Zurich instruments HF2LI lock-in amplifier was used for phase-lock looping, and Allan deviation measurements were taken from both open- and closed-loop conditions. In closed

loop, a typical setting was 4-kHz demodulation bandwidth (the integration bandwidth), allowing for a 500-Hz phase-locked loop bandwidth. Temperature calibration (section 1.7) was done using both the microring temperature tuning (found to be between 70 and 80 pm/K) and the NOMS temperature tuning (ranging from  $-12$  to  $-15 \text{ kHz/K}$  for TCRF values from  $-1050$  to  $-1270 \text{ ppm/K}$ ) with a PID temperature controller. The high value for TCRF was also verified from a thermal strain model that gives  $-1100 \text{ ppm/K}$ .

Section 2 provides detailed frequency stability analysis. The theorized conversion of thermal force noise to displacement noise and on to phase noise is shown in section 2.1. This derivation proceeds in typical fashion, except that no assumption is made that  $\omega \gg \Gamma$ . This results in a low-pass functional form of phase noise that reduces to the traditional form ( $S_\phi \propto 1/\omega^2$ ) for large  $\omega$ , but acts as frequency independent (flat) for small  $\omega$ . The Allan deviation is defined with the expected stability specified based on power law phase noise (section 2.2). This produces the traditional Robins' stability limit for large  $\omega$  (Eq. 6) and the new flatband stability limit for small  $\omega$  (Eq. 10). Section 2.3 confirms the power law integrations that produce Eqs. 6 and 10, and section 2.4 compares our derivation nomenclature with some existing Robins' derivations in the literature.

Signs of frequency-fluctuation noise (in distinction from additive white noise) in data at different pressures were sought via open-loop frequency tracking, Allan deviation measurements, and quadrature analysis (section 2.5). Clear signatures of frequency-fluctuation noise were seen at high  $Q$  in vacuum, and the absence of signatures of such noise was observed at low  $Q$  at atmospheric pressure. Finally, phase noise was measured directly in open loop at different pressures (section 2.6) and found to correlate with  $DR$ , consistent with behavior expected in flatband regime.

## REFERENCES AND NOTES

1. J. Chaste *et al.*, A nanomechanical mass sensor with yoctogram resolution. *Nat. Nanotechnol.* **7**, 301–304 (2012). doi: [10.1038/nnano.2012.42](https://doi.org/10.1038/nnano.2012.42); pmid: 22466856
2. W. Hiebert, Mass sensing: Devices reach single-proton limit. *Nat. Nanotechnol.* **7**, 278–280 (2012). doi: [10.1038/nnano.2012.66](https://doi.org/10.1038/nnano.2012.66); pmid: 22569423
3. A. Venkatasubramanian *et al.*, Nano-optomechanical systems for gas chromatography. *Nano Lett.* **16**, 6975–6981 (2016). doi: [10.1021/acs.nanolett.6b03066](https://doi.org/10.1021/acs.nanolett.6b03066); pmid: 27749074
4. I. Bargatin *et al.*, Large-scale integration of nanoelectromechanical systems for gas sensing applications. *Nano Lett.* **12**, 1269–1274 (2012). doi: [10.1021/nl2037479](https://doi.org/10.1021/nl2037479); pmid: 22280452
5. M. Li *et al.*, Nanoelectromechanical resonator arrays for ultrafast, gas-phase chromatographic chemical analysis. *Nano Lett.* **10**, 3899–3903 (2010). doi: [10.1021/nl101586s](https://doi.org/10.1021/nl101586s); pmid: 20795729
6. M. S. Hanay *et al.*, Inertial imaging with nanomechanical systems. *Nat. Nanotechnol.* **10**, 339–344 (2015). doi: [10.1038/nnano.2015.32](https://doi.org/10.1038/nnano.2015.32); pmid: 25822931
7. E. Sage *et al.*, Neutral particle mass spectrometry with nanomechanical systems. *Nat. Commun.* **6**, 6482 (2015). doi: [10.1038/ncomms7482](https://doi.org/10.1038/ncomms7482); pmid: 25753929
8. M. S. Hanay *et al.*, Single-protein nanomechanical mass spectrometry in real time. *Nat. Nanotechnol.* **7**, 602–608 (2012). doi: [10.1038/nnano.2012.119](https://doi.org/10.1038/nnano.2012.119); pmid: 22922541

9. A. K. Naik, M. S. Hanay, W. K. Hiebert, X. L. Feng, M. L. Roukes, Towards single-molecule nanomechanical mass spectrometry. *Nat. Nanotechnol.* **4**, 445–450 (2009). doi: [10.1038/nnano.2009.152](https://doi.org/10.1038/nnano.2009.152); pmid: 19581898
10. C. L. Degen, M. Poggio, H. J. Mamin, C. T. Rettner, D. Rugar, Nanoscale magnetic resonance imaging. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1313–1317 (2009). doi: [10.1073/pnas.0812068106](https://doi.org/10.1073/pnas.0812068106); pmid: 19139397
11. D. Rugar, R. Budakian, H. J. Mamin, B. W. Chui, Single spin detection by magnetic resonance force microscopy. *Nature* **430**, 329–332 (2004). doi: [10.1038/nature02658](https://doi.org/10.1038/nature02658); pmid: 15254532
12. M. Dong, O. Sahin, A nanomechanical interface to rapid single-molecule interactions. *Nat. Commun.* **2**, 247 (2011). doi: [10.1038/ncomms1246](https://doi.org/10.1038/ncomms1246); pmid: 21427718
13. J. E. Losby *et al.*, Torque-mixing magnetic resonance spectroscopy. *Science* **350**, 798–801 (2015). doi: [10.1126/science.124449](https://doi.org/10.1126/science.124449); pmid: 26564851
14. J. D. Teufel *et al.*, Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475**, 359–363 (2011). doi: [10.1038/nature10261](https://doi.org/10.1038/nature10261); pmid: 21734657
15. J. Chan *et al.*, Laser cooling of a nanomechanical oscillator into its quantum ground state. *Nature* **478**, 89–92 (2011). doi: [10.1038/nature10461](https://doi.org/10.1038/nature10461); pmid: 21979049
16. A. D. O'Connell *et al.*, Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010). doi: [10.1038/nature08967](https://doi.org/10.1038/nature08967); pmid: 20237473
17. T. P. Purdy, K. E. Grutter, K. Srinivasan, J. M. Taylor, Quantum correlations from a room-temperature optomechanical cavity. *Science* **356**, 1265–1268 (2017). doi: [10.1126/science.aag1407](https://doi.org/10.1126/science.aag1407); pmid: 28642433
18. W. P. Robins, *Phase Noise in Signal Sources: Theory and Applications* (Institution of Electrical Engineers Telecommunications Series, Institution of Engineering and Technology, London, 1982).
19. T. R. Albrecht, P. Grütter, D. Horne, D. Rugar, Frequency modulation detection using high- $Q$  cantilevers for enhanced force microscope sensitivity. *J. Appl. Phys.* **69**, 668–673 (1991). doi: [10.1063/1.347347](https://doi.org/10.1063/1.347347)
20. A. N. Cleland, M. L. Roukes, Noise processes in nanomechanical resonators. *J. Appl. Phys.* **92**, 2758–2769 (2002). doi: [10.1063/1.1499745](https://doi.org/10.1063/1.1499745)
21. K. L. Ekinci, Y. T. Yang, M. L. Roukes, Ultimate limits to inertial mass sensing based upon nanoelectromechanical systems. *J. Appl. Phys.* **95**, 2682–2689 (2004). doi: [10.1063/1.1642738](https://doi.org/10.1063/1.1642738)
22. Y. Tsaturyan, A. Barg, E. S. Polzik, A. Schliesser, Ultra-coherent nanomechanical resonators via soft clamping and dissipation dilution. *Nat. Nanotechnol.* **12**, 776–783 (2017). doi: [10.1038/nnano.2017.101](https://doi.org/10.1038/nnano.2017.101)
23. J. Moser, A. Eichler, J. Güttinger, M. I. Dykman, A. Bachtold, Nanotube mechanical resonators with quality factors of up to 5 million. *Nat. Nanotechnol.* **9**, 1007–1011 (2014). doi: [10.1038/nnano.2014.234](https://doi.org/10.1038/nnano.2014.234); pmid: 25344688
24. K. Y. Fong, W. H. P. Pernice, H. X. Tang, Frequency and phase noise of ultrahigh  $Q$  silicon nitride nanomechanical resonators. *Phys. Rev. B* **85**, 161410 (2012). doi: [10.1103/PhysRevB.85.161410](https://doi.org/10.1103/PhysRevB.85.161410)
25. Y. Tao, J. M. Boss, B. A. Moores, C. L. Degen, Single-crystal diamond nanomechanical resonators with quality factors exceeding one million. *Nat. Commun.* **5**, 3638 (2014). doi: [10.1038/ncomms4638](https://doi.org/10.1038/ncomms4638); pmid: 24710311
26. J. Gieseler, L. Novotny, R. Quidant, Thermal nonlinearities in a nanomechanical oscillator. *Nat. Phys.* **9**, 806–810 (2013). doi: [10.1038/nphys2798](https://doi.org/10.1038/nphys2798)
27. Z. Wang, P. X. L. Feng, Dynamic range of atomically thin vibrating nanomechanical resonators. *Appl. Phys. Lett.* **104**, 103109 (2014). doi: [10.1063/1.4868129](https://doi.org/10.1063/1.4868129)
28. M. Sansa *et al.*, Frequency fluctuations in silicon nanoresonators. *Nat. Nanotechnol.* **11**, 552–558 (2016). doi: [10.1038/nnano.2016.19](https://doi.org/10.1038/nnano.2016.19); pmid: 26925826
29. E. Gavartin, P. Verlot, T. J. Kippenberg, Stabilization of a linear nanomechanical oscillator to its thermodynamic limit. *Nat. Commun.* **4**, 2860 (2013). doi: [10.1038/ncomms3860](https://doi.org/10.1038/ncomms3860); pmid: 24326974
30. F. Sun, X. Dong, J. Zou, M. I. Dykman, H. B. Chan, Correlated anomalous phase diffusion of coupled phononic modes in a sideband-driven resonator. *Nat. Commun.* **7**, 12694 (2016). doi: [10.1038/ncomms12694](https://doi.org/10.1038/ncomms12694); pmid: 27576597
31. O. Maillet, F. Vavrek, D. Fefferman, O. Bourgeois, E. Collin, Classical decoherence in a nanomechanical resonator. *New J. Phys.* **18**, 073022 (2016). doi: [10.1088/1367-2630/18/7/073022](https://doi.org/10.1088/1367-2630/18/7/073022)

32. B. H. Schneider, V. Singh, W. J. Venstra, H. B. Meerwaldt, G. A. Steele, Observation of decoherence in a carbon nanotube mechanical resonator. *Nat. Commun.* **5**, 5819 (2014). doi: [10.1038/ncomms6819](#); pmid: [25524228](#)
33. N. Inomata, M. Toda, T. Ono, Highly sensitive thermometer using a vacuum-packed Si resonator in a microfluidic chip for the thermal measurement of single cells. *Lab Chip* **16**, 3597–3603 (2016). doi: [10.1039/C6LC00949B](#); pmid: [27526966](#)
34. X. C. Zhang, E. B. Myers, J. E. Sader, M. L. Roukes, Nanomechanical torsional resonators for frequency-shift infrared thermal sensing. *Nano Lett.* **13**, 1528–1534 (2013). doi: [10.1021/nl304687p](#); pmid: [23458733](#)
35. N. Rouger, L. Chrostowski, R. Vafaei, Temperature effects on silicon-on-insulator (SOI) racetrack resonators: A coupled analytic and 2-D finite difference approach. *J. Lightwave Technol.* **28**, 1380–1391 (2010). doi: [10.1109/JLT.2010.2041528](#)
36. R. Melamud *et al.*, Temperature-compensated high-stability silicon resonators. *Appl. Phys. Lett.* **90**, 244107 (2007). doi: [10.1063/1.2748092](#)
37. J. S. Bunch *et al.*, Electromechanical resonators from graphene sheets. *Science* **315**, 490–493 (2007). doi: [10.1126/science.1136836](#); pmid: [17255506](#)
38. M. Li, H. X. Tang, M. L. Roukes, Ultra-sensitive NEMS-based cantilevers for sensing, scanned probe and very high-frequency applications. *Nat. Nanotechnol.* **2**, 114–120 (2007). doi: [10.1038/nnano.2006.208](#); pmid: [18654230](#)
39. J. L. Ariett, J. R. Maloney, B. Gudlewski, M. Muluneh, M. L. Roukes, Self-sensing micro- and nanocantilevers with Attonewton-scale force resolution. *Nano Lett.* **6**, 1000–1006 (2006). doi: [10.1021/nl060275y](#)
40. W. K. Hiebert, D. Vick, V. Sauer, M. R. Freeman, Optical interferometric displacement calibration and thermomechanical noise detection in bulk focused ion beam-fabricated nanoelectromechanical systems. *J. Micromech. Microeng.* **20**, 115038 (2010). doi: [10.1088/0960-1317/20/11/115038](#)
41. Q. P. Unterreithmeier, E. M. Weig, J. P. Kotthaus, Universal transduction scheme for nanomechanical systems based on dielectric forces. *Nature* **458**, 1001–1004 (2009). doi: [10.1038/nature07932](#); pmid: [19396140](#)
42. M. Li *et al.*, Harnessing optical forces in integrated photonic circuits. *Nature* **456**, 480–484 (2008). doi: [10.1038/nature07545](#); pmid: [19037311](#)
43. M. Wu *et al.*, Nanocavity optomechanical torque magnetometry and radiofrequency susceptometry. *Nat. Nanotechnol.* **12**, 127–131 (2017). doi: [10.1038/nnano.2016.226](#); pmid: [27798605](#)
44. P. H. Kim, B. D. Hauer, C. Doolin, F. Souris, J. P. Davis, Approaching the standard quantum limit of mechanical torque sensing. *Nat. Commun.* **7**, 13165 (2016). doi: [10.1038/ncomms13165](#); pmid: [27762273](#)
45. P. Weber, J. Güttinger, A. Noury, J. Vergara-Cruz, A. Bachtold, Force sensitivity of multilayer graphene optomechanical devices. *Nat. Commun.* **7**, 12496 (2016). doi: [10.1038/ncomms12496](#); pmid: [27502017](#)
46. L. M. de Lépinay *et al.*, A universal and ultrasensitive vectorial nanomechanical sensor for imaging 2D force fields. *Nat. Nanotechnol.* **12**, 156–162 (2017). doi: [10.1038/nnano.2016.193](#); pmid: [27749835](#)
47. S. Olcum, N. Cermak, S. C. Wasserman, S. R. Manalis, High-speed multiple-mode mass-sensing resolves dynamic nanoscale mass distributions. *Nat. Commun.* **6**, 7070 (2015). doi: [10.1038/ncomms8070](#); pmid: [25963304](#)
48. E. Gil-Santos *et al.*, High-frequency nano-optomechanical disk resonators in liquids. *Nat. Nanotechnol.* **10**, 810–816 (2015). doi: [10.1038/nnano.2015.160](#); pmid: [26237347](#)
49. V. T. K. Sauer, Z. Diao, M. R. Freeman, W. K. Hiebert, Optical racetrack resonator transduction of nanomechanical cantilevers. *Nanotechnology* **25**, 055202 (2014). doi: [10.1088/0957-4484/25/5/055202](#); pmid: [24406727](#)
50. J. Moser *et al.*, Ultrasensitive force detection with a nanotube mechanical resonator. *Nat. Nanotechnol.* **8**, 493–496 (2013). doi: [10.1038/nnano.2013.97](#); pmid: [23748195](#)
51. E. Gavartin, P. Verlot, T. J. Kippenberg, A hybrid on-chip optomechanical transducer for ultrasensitive force measurements. *Nat. Nanotechnol.* **7**, 509–514 (2012). doi: [10.1038/nnano.2012.97](#); pmid: [22728341](#)
52. K. Srinivasan, H. Miao, M. T. Rakher, M. Davaño, V. Aksyuk, Optomechanical transduction of an integrated silicon cantilever probe using a microdisk resonator. *Nano Lett.* **11**, 791–797 (2011). doi: [10.1021/nl104018r](#); pmid: [21250747](#)
53. E. Gil-Santos *et al.*, Nanomechanical mass sensing and stiffness spectrometry based on two-dimensional vibrations of resonant nanowires. *Nat. Nanotechnol.* **5**, 641–645 (2010). doi: [10.1038/nnano.2010.151](#); pmid: [20693990](#)
54. M. Eichenfield, R. Camacho, J. Chan, K. J. Vahala, O. Painter, A picogram- and nanometre-scale photonic-crystal optomechanical cavity. *Nature* **459**, 550–555 (2009). doi: [10.1038/nature08061](#); pmid: [19489118](#)
55. J. D. Teufel, T. Donner, M. A. Castellanos-Beltrán, J. W. Harlow, K. W. Lehnert, Nanomechanical motion measured with an imprecision below that at the standard quantum limit. *Nat. Nanotechnol.* **4**, 820–823 (2009). doi: [10.1038/nnano.2009.343](#); pmid: [19893515](#)
56. G. Anetsberger *et al.*, Near-field cavity optomechanics with nanomechanical oscillators. *Nat. Phys.* **5**, 909–914 (2009). doi: [10.1038/nphys1425](#)
57. M. Belov *et al.*, Mechanical resonance of clamped silicon nanowires measured by optical interferometry. *J. Appl. Phys.* **103**, 074304 (2008). doi: [10.1063/1.2891002](#)
58. V. T. K. Sauer, Z. Diao, M. R. Freeman, W. K. Hiebert, Nanophotonic detection of side-coupled nanomechanical cantilevers. *Appl. Phys. Lett.* **100**, 261102 (2012). doi: [10.1063/1.4731210](#)
59. V. T. K. Sauer, Z. Diao, J. N. Westwood-Bachman, M. R. Freeman, W. K. Hiebert, Single laser modulated drive and detection of a nano-optomechanical cantilever. *AIP Adv.* **7**, 015115 (2017). doi: [10.1063/1.4975347](#)
60. V. T. Sauer, Z. Diao, M. R. Freeman, W. K. Hiebert, Wavelength-division multiplexing of nano-optomechanical doubly clamped beam systems. *Opt. Lett.* **40**, 1948–1951 (2015). doi: [10.1364/OL.40.001948](#); pmid: [25927755](#)
61. J. Westwood-Bachman, Z. Diao, V. Sauer, D. Bachman, W. Hiebert, Even nanomechanical modes transduced by integrated photonics. *Appl. Phys. Lett.* **108**, 061103 (2016). doi: [10.1063/1.4941542](#)
62. Z. Diao *et al.*, Confocal scanner for highly sensitive photonic transduction of nanomechanical resonators. *Appl. Phys. Express* **6**, 065202 (2013). doi: [10.7567/APEX.6.065202](#)
63. See supplementary materials.
64. S. Schmid, L. G. Villanueva, M. L. Roukes, *Fundamentals of Nanomechanical Resonators* (Springer, 2016).
65. N. Kacem, S. Hentz, D. Pinto, B. Reig, V. Nguyen, Nonlinear dynamics of nanomechanical beam resonators: Improving the performance of NEMS-based sensors. *Nanotechnology* **20**, 275501 (2009). doi: [10.1088/0957-4484/20/27/275501](#); pmid: [19528678](#)
66. H. W. C. Postma, I. Kozinsky, A. Husain, M. L. Roukes, Dynamic range of nanotube- and nanowire-based electromechanical systems. *Appl. Phys. Lett.* **86**, 223105 (2005). doi: [10.1063/1.1929098](#)
67. I. Bargatin, “High-frequency nanomechanical resonators for sensor applications,” thesis, California Institute of Technology (2008).
68. T. E. Schäffer, J. P. Cleveland, F. Ohnesorge, D. A. Walters, P. K. Hansma, Studies of vibrating atomic force microscope cantilevers in liquid. *J. Appl. Phys.* **80**, 3622–3627 (1996). doi: [10.1063/1.363308](#)
69. J. A. Barnes *et al.*, Characterization of frequency stability. *IEEE Trans. Instrum. Meas.* **IM-20**, 105–120 (1971). doi: [10.1109/TIM.1971.5570702](#)
70. M. Poggio, C. L. Degen, H. J. Mamin, D. Rugar, Feedback cooling of a cantilever’s fundamental mode below 5 mK. *Phys. Rev. Lett.* **99**, 017201 (2007). doi: [10.1103/PhysRevLett.99.017201](#); pmid: [17678185](#)
71. L. G. Villanueva *et al.*, Surpassing fundamental limits of oscillators using nonlinear resonators. *Phys. Rev. Lett.* **110**, 177208 (2013). doi: [10.1103/PhysRevLett.110.177208](#); pmid: [23679770](#)
72. D. Antonio, D. H. Zanette, D. López, Frequency stabilization in nonlinear micromechanical oscillators. *Nat. Commun.* **3**, 806 (2012). doi: [10.1038/ncomms1813](#); pmid: [22549835](#)
73. M. Poot, K. Y. Fong, H. X. Tang, Deep feedback-stabilized parametric squeezing in an opto-electromechanical system. *New J. Phys.* **17**, 043056 (2015). doi: [10.1088/1367-2630/17/4/043056](#)

## ACKNOWLEDGMENTS

The authors acknowledge the National Research Council’s Nanotechnology Research Centre and its fabrication, microscopy, and measurement facilities. The fabrication of the devices was facilitated through CMC Microsystems (silicon photonics services and CAD tools), and postprocessing was performed at the University of Alberta nanoFAB. We thank P. Barclay and M. Freeman for thoughtfully reviewing the manuscript. **Funding:** Funding was provided through author institutions as well as scholarship support from Alberta Innovates Technology Futures and the Vanier Canada Graduate Scholarship program. Grant support was provided from Alberta Innovates Health Solutions through a collaborative research innovation opportunity and from the Natural Sciences and Engineering Research Council, Canada, through the Discovery Grant program. **Author contributions:** W.K.H. conceived and designed the research. S.K.R. conducted the measurements with training and assistance from V.T.K.S., J.N.W.-B., and A.V. V.T.K.S. designed the devices for fabrication. W.K.H., S.K.R., and V.T.K.S. wrote the manuscript. All authors contributed to data interpretation and manuscript preparation. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/eaar5220/suppl/DC1](http://www.sciencemag.org/content/360/6394/eaar5220/suppl/DC1)  
Materials and Methods  
Supplementary Text  
Figs. S1 to S19  
Table S1  
References (74–90)

18 November 2017; accepted 23 April 2018  
10.1126/science.aar5220



## RESEARCH ARTICLE

## MACHINE LEARNING

## Neural scene representation and rendering

S. M. Ali Eslami<sup>\*†</sup>, Danilo Jimenez Rezende<sup>†</sup>, Frederic Besse, Fabio Viola, Ari S. Morcos, Marta Garnelo, Avraham Ruderman, Andrei A. Rusu, Ivo Danihelka, Karol Gregor, David P. Reichert, Lars Buesing, Theophane Weber, Oriol Vinyals, Dan Rosenbaum, Neil Rabinowitz, Helen King, Chloe Hillier, Matt Botvinick, Daan Wierstra, Koray Kavukcuoglu, Demis Hassabis

Scene representation—the process of converting visual sensory data into concise descriptions—is a requirement for intelligent behavior. Recent work has shown that neural networks excel at this task when provided with large, labeled datasets. However, removing the reliance on human labeling remains an important open problem. To this end, we introduce the Generative Query Network (GQN), a framework within which machines learn to represent scenes using only their own sensors. The GQN takes as input images of a scene taken from different viewpoints, constructs an internal representation, and uses this representation to predict the appearance of that scene from previously unobserved viewpoints. The GQN demonstrates representation learning without human labels or domain knowledge, paving the way toward machines that autonomously learn to understand the world around them.

Modern artificial vision systems are based on deep neural networks that consume large, labeled datasets to learn functions that map images to human-generated scene descriptions. They do so by, for example, categorizing the dominant object in the image (1), classifying the scene type (2), detecting object-bounding boxes (3), or labeling individual pixels into predetermined categories (4, 5). In contrast, intelligent agents in the natural world appear to require little to no explicit supervision for perception (6). Higher mammals, including human infants, learn to form representations that support motor control, memory, planning, imagination, and rapid skill acquisition without any social communication, and generative pro-

cesses have been hypothesized to be instrumental for this ability (7–10). It is thus desirable to create artificial systems that learn to represent scenes by modeling data [e.g., two-dimensional (2D) images and the agent's position in space] that agents can directly obtain while processing the scenes themselves, without recourse to semantic labels (e.g., object classes, object locations, scene types, or part labels) provided by a human (11).

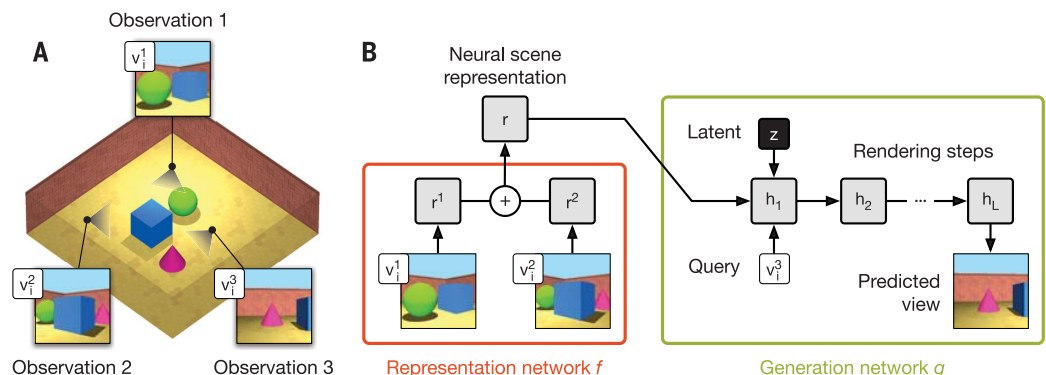
To that end, we present the Generative Query Network (GQN). In this framework, as an agent navigates a 3D scene  $i$ , it collects  $K$  images  $\mathbf{x}_i^k$  from 2D viewpoints  $\mathbf{v}_i^k$ , which we collectively refer to as its observations  $\mathbf{o}_i = \{(\mathbf{x}_i^k, \mathbf{v}_i^k)\}_{k=1, \dots, K}$ . The agent passes these observations to a GQN

composed of two main parts: a representation network  $f$  and a generation network  $g$  (Fig. 1). The representation network takes as input the agent's observations and produces a neural scene representation  $\mathbf{r}$ , which encodes information about the underlying scene (we omit scene subscript  $i$  where possible, for clarity). Each additional observation accumulates further evidence about the contents of the scene in the same representation. The generation network then predicts the scene from an arbitrary query viewpoint  $\mathbf{v}^q$ , using stochastic latent variables  $\mathbf{z}$  to create variability in its outputs where necessary. The two networks are trained jointly, in an end-to-end fashion, to maximize the likelihood of generating the ground-truth image that would be observed from the query viewpoint. More formally, (i)  $\mathbf{r} = f_{\theta}(\mathbf{o}_i)$ , (ii) the deep generation network  $g$  defines a probability density  $g_{\theta}(\mathbf{x}|\mathbf{v}^q, \mathbf{r}) = \int g_{\theta}(\mathbf{x}, \mathbf{z}|\mathbf{v}^q, \mathbf{r}) d\mathbf{z}$  of an image  $\mathbf{x}$  being observed at query viewpoint  $\mathbf{v}^q$  for a scene representation  $\mathbf{r}$  using latent variables  $\mathbf{z}$ , and (iii) the learnable parameters are denoted by  $\theta$ . Although the GQN training objective is intractable, owing to the presence of latent variables, we can employ variational approximations and optimize with stochastic gradient descent.

The representation network is unaware of the viewpoints that the generation network will be queried to predict. As a result, it will produce scene representations that contain all information (e.g., object identities, positions, colors, counts, and room layout) necessary for the generator to make accurate image predictions. In other words, the GQN will learn by itself what these factors are, as well as how to extract them from pixels. Moreover, the generator internalizes any statistical regularities (e.g., typical colors of the sky, as well as object shape regularities and symmetries, patterns, and textures) that are common across different scenes. This allows

**Fig. 1. Schematic illustration of the Generative Query Network.**

(A) The agent observes training scene  $i$  from different viewpoints (in this example, from  $\mathbf{v}_i^1$ ,  $\mathbf{v}_i^2$ , and  $\mathbf{v}_i^3$ ). (B) The inputs to the representation network  $f$  are observations made from viewpoints  $\mathbf{v}_i^1$  and  $\mathbf{v}_i^2$ , and the output is the scene representation  $\mathbf{r}$ , which is obtained by element-wise summing of the observations' representations. The generation network, a recurrent latent variable model, uses the representation to predict what the scene would look like from a different viewpoint  $\mathbf{v}_i^3$ . The generator can succeed only if  $\mathbf{r}$  contains accurate and complete information about the contents of the scene (e.g., the identities, positions, colors, and counts of the objects, as well as the room's colors). Training via back-propagation across many scenes, randomizing the number of observations, leads to learned scene representations that capture this information in a concise manner. Only a handful of observations need to be recorded from any single scene to train the GQN.  $h_1, h_2, \dots, h_L$  are the  $L$  layers of the generation network.



DeepMind, 5 New Street Square, London EC4A 3TW, UK.  
<sup>\*</sup>Corresponding author. Email: aesi@deepmind.com  
<sup>†</sup>These authors contributed equally to this work.

the GQN to reserve its representation capacity for a concise, abstract description of the scene, with the generator filling in the details where necessary. For instance, instead of specifying the precise shape of a robot arm, the representation network can succinctly communicate the configuration of its joints, and the generator knows how this high-level representation manifests itself as a fully rendered arm with its precise shapes and colors. In contrast, voxel (12–15) or point-cloud (16) methods (as typically obtained by classical structure-from-motion) employ literal representations and therefore typically scale poorly with scene complexity and size and are also difficult to apply to nonrigid objects (e.g., animals, vegetation, or cloth).

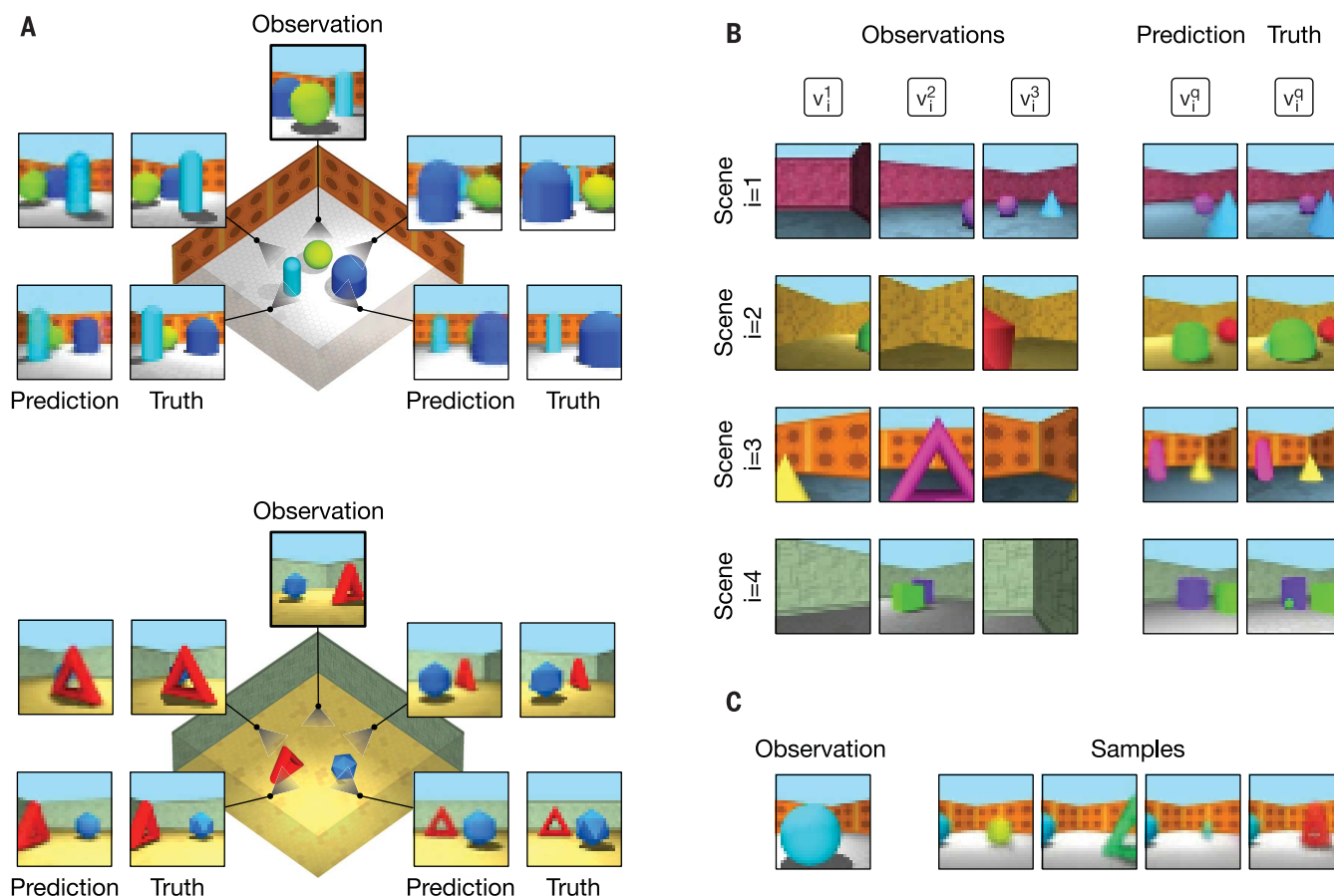
### Rooms with multiple objects

To evaluate the feasibility of the framework, we experimented with a collection of environments in a simulated 3D environment. In the first set

of experiments, we considered scenes in a square room containing a variety of objects. Wall textures—as well as the shapes, positions, and colors of the objects and lights—are randomized, allowing for an effectively infinite number of total scene configurations; however, we used finite datasets to train and test the model [see section 4 of (17) for details]. After training, the GQN computes its scene representation by observing one or more images of a previously unencountered, held-out test scene. With this representation, which can be as small as 256 dimensions, the generator's predictions at query viewpoints are highly accurate and mostly indistinguishable from ground truth (Fig. 2A). The only way in which the model can succeed at this task is by perceiving and compactly encoding in the scene representation vector  $\mathbf{r}$  the number of objects present in each scene, their positions in the room, the colors in which they appear, the colors of the walls, and the indirectly observed position of the light source.

Unlike in traditional supervised learning, GQNs learn to make these inferences from images without any explicit human labeling of the contents of scenes. Moreover, the GQN's generator learns an approximate 3D renderer (in other words, a program that can generate an image when given a scene representation and camera viewpoint) without any prior specification of the laws of perspective, occlusion, or lighting (Fig. 2B). When the contents of the scene are not explicitly specified by the observation (e.g., because of heavy occlusion), the model's uncertainty is reflected in the variability of the generator's samples (Fig. 2C). These properties are best observed in real-time, interactive querying of the generator (movie S1).

Notably, the model observes only a small number of images (in this experiment, fewer than five) from each scene during training, yet it is capable of rendering unseen training or test scenes from arbitrary viewpoints. We also monitored



**Fig. 2. Neural scene representation and rendering.** (A) After having made a single observation of a previously unencountered test scene, the representation network produces a neural description of that scene. Given this neural description, the generator is capable of predicting accurate images from arbitrary query viewpoints. This implies that the scene description captures the identities, positions, colors, and counts of the objects, as well as the position of the light and the colors of the room. (B) The generator's predictions are consistent with laws of perspective, occlusion,

and lighting (e.g., casting object shadows consistently). When observations provide views of different parts of the scene, the GQN correctly aggregates this information (scenes two and three). (C) Sample variability indicates uncertainty over scene contents (in this instance, owing to heavy occlusion). Samples depict plausible scenes, with complete objects rendered in varying positions and colors (see fig. S7 for further examples). The model's behavior is best visualized in movie format; see movie S1 for real-time, interactive querying of GQN's representation of test scenes.



the likelihood of predicted observations of training and test scenes (fig. S3) and found no noticeable difference between values of the two. Taken together, these points rule out the possibility of model overfitting.

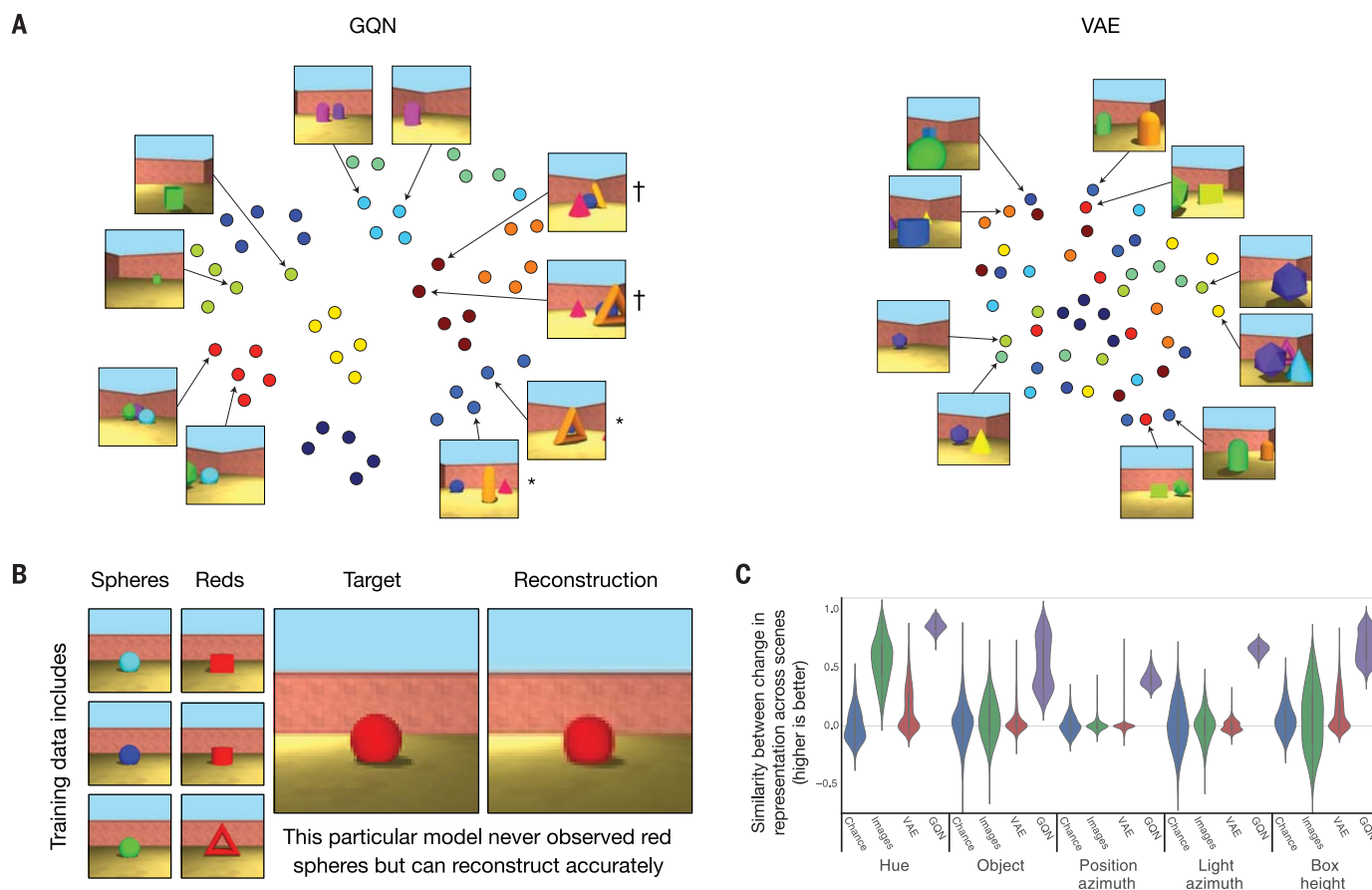
Analysis of the trained GQN highlights several desirable properties of its scene representation network. Two-dimensional t-distributed stochastic neighbor embedding (t-SNE) (18) visualization of GQN scene representation vectors shows clear clustering of images of the same scene, despite marked changes in viewpoint (Fig. 3A). In contrast, representations produced by autoencoding density models such as variational autoencoders (VAEs) (19) apparently fail to capture the contents of the underlying scenes [section 5 of (17)]; they appear to be representations of the observed images instead. Furthermore, when prompted to reconstruct a target image, GQN exhibits compositional behavior, as it is capable of both representing and rendering combinations of scene elements it has

never encountered during training (Fig. 3B), despite learning that these compositions are unlikely. To test whether the GQN learns a factorized representation, we investigated whether changing a single scene property (e.g., object color) while keeping others (e.g., object shape and position) fixed leads to similar changes in the scene representation (as defined by mean cosine similarity across scenes). We found that object color, shape, and size; light position; and, to a lesser extent, object positions are indeed factorized [Fig. 3C and sections 5.3 and 5.4 of (17)]. We also found that the GQN is able to carry out "scene algebra" [akin to word embedding algebra (20)]. By adding and subtracting representations of related scenes, we found that object and scene properties can be controlled, even across object positions [Fig. 4A and section 5.5 of (17)]. Finally, because it is a probabilistic model, GQN also learns to integrate information from different viewpoints in an efficient and consistent manner, as demonstrated by a reduction in its Bayesian

"surprise" at observing a held-out image of a scene as the number of views increases [Fig. 4B and section 3 of (17)]. We include analysis on the GQN's ability to generalize to out-of-distribution scenes, as well as further results on modeling of Shepard-Metzler objects, in sections 5.6 and 4.2 of (17).

### Control of a robotic arm

Representations that succinctly reflect the true state of the environment should also allow agents to learn to act in those environments more robustly and with fewer interactions. Therefore, we considered the canonical task of moving a robotic arm to reach a colored object, to test the GQN representation's suitability for control. The end-goal of deep reinforcement learning is to learn the control policy directly from pixels; however, such methods require a large amount of experience to learn from sparse rewards. Instead, we first trained a GQN and used it to succinctly represent the observations. A policy was then trained



**Fig. 3. Viewpoint invariance, compositionality, and factorization of the learned scene representations.** (A) t-SNE embeddings. t-SNE is a method for nonlinear dimensionality reduction that approximately preserves the metric properties of the original high-dimensional data. Each dot represents a different view of a different scene, with color indicating scene identity. Whereas the VAE clusters images mostly on the basis of wall angles, GQN clusters images of the same scene, independent of view (scene representations computed from each image individually). Two scenes with

the same objects (represented by asterisk and dagger symbols) but in different positions are clearly separated. (B) Compositionality demonstrated by reconstruction of holdout shape-color combinations. (C) GQN factorizes object and scene properties because the effect of changing a specific property is similar across diverse scenes (as defined by mean cosine similarity of the changes in the representation across scenes). For comparison, we plot chance factorization, as well as the factorization of the image-space and VAE representations. See section 5.3 of (17) for details.

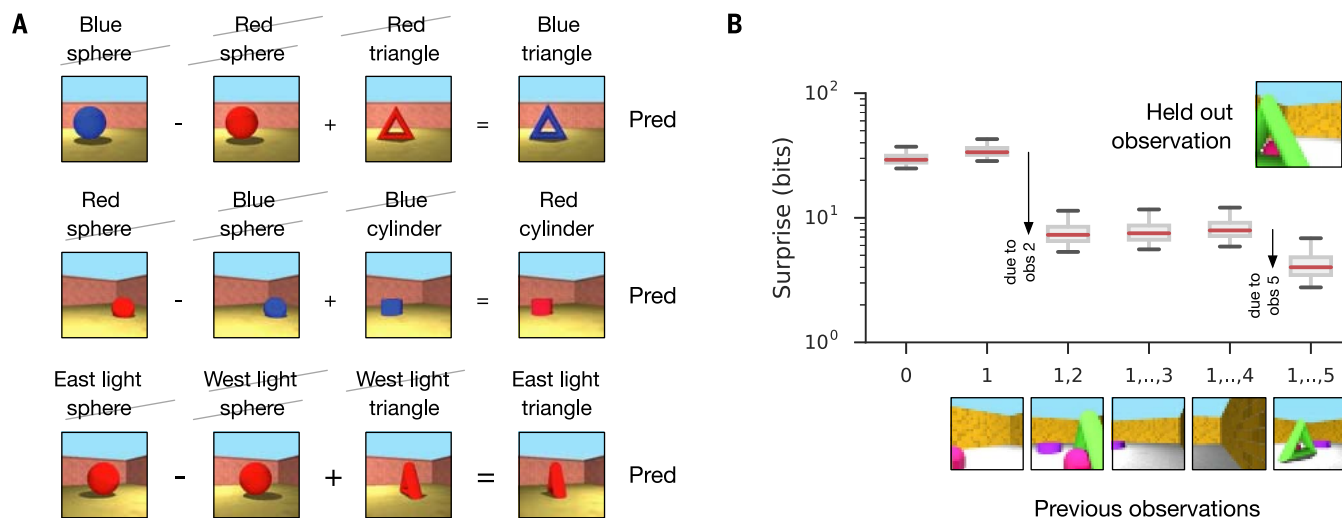
to control the arm directly from these representations. In this setting, the representation network must learn to communicate only the arm's joint angles, the position and color of the object, and the colors of the walls for the generator to be able to predict new views. Because this vector has much lower dimensionality than the raw input images, we observed substantially more robust and data-efficient policy learning, obtaining convergence-level control performance with approximately one-fourth as many interactions with the environment as a standard method using raw

pixels [Fig. 5 and section 4.4 of (17)]. The 3D nature of the GQN representation allows us to train a policy from any viewpoint around the arm and is sufficiently stable to allow for arm-joint velocity control from a freely moving camera.

### Partially observed maze environments

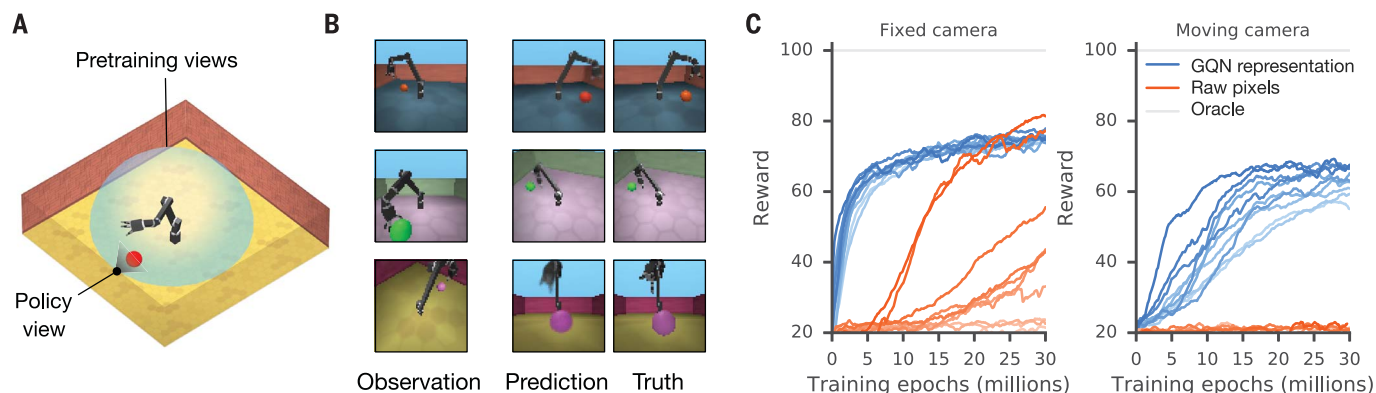
Finally, we considered more complex, procedural maze-like environments to test GQN's scaling properties. The mazes consist of multiple rooms connected via corridors, and the layout of each maze and the colors of the walls are randomized

in each scene. In this setting, any single observation provides a small amount of information about the current maze. As before, the training objective for GQN is to predict mazes from new viewpoints, which is possible only if GQN successfully aggregates multiple observations to determine the maze layout (i.e., the wall and floor colors, the number of rooms, their positions in space, and how they connect to one another via corridors). We observed that GQN is able to make correct predictions from new first-person viewpoints (Fig. 6A). We queried the GQN's



**Fig. 4. Scene algebra and Bayesian surprise.** (A) Adding and subtracting representations of related scenes enables control of object and scene properties via “scene algebra” and indicates factorization of shapes, colors, and positions. Pred, prediction. (B) Bayesian surprise at a new observation

after having made observations 1 to  $k$  for  $k = 1$  to 5. When the model observes images that contain information about the layout of the scene, its surprise (defined as the Kullback-Leibler divergence between conditional prior and posterior) at observing the held-out image decreases.



**Fig. 5. GQN representation enables more robust and data-efficient control.** (A) The goal is to learn to control a robotic arm to reach a randomly positioned colored object. The controlling policy observes the scene from a fixed or moving camera (gray). We pretrain a GQN representation network by observing random configurations from random viewpoints inside a dome around the arm (light blue). (B) The GQN infers a scene representation that can accurately reconstruct the scene. (C) (Left) For a fixed camera, an asynchronous advantage actor-critic reinforcement learning (RL) agent (44) learns to control the arm using roughly one-fourth as many experiences when using the GQN representation, as opposed to a standard method using raw pixels (lines correspond

to different hyperparameters; same hyperparameters explored for both standard and GQN agents; both agents also receive viewpoint coordinates as inputs). The final performance achieved by learning from raw pixels can be slightly higher for some hyperparameters, because some task-specific information might be lost when learning a compressed representation independently from the RL task as GQN does. (Right) The benefit of GQN is most pronounced when the policy network's view on the scene moves from frame to frame, suggesting viewpoint invariance in its representation. We normalize scores such that a random agent achieves 0 and an agent trained on “oracle” ground-truth state information achieves 100.



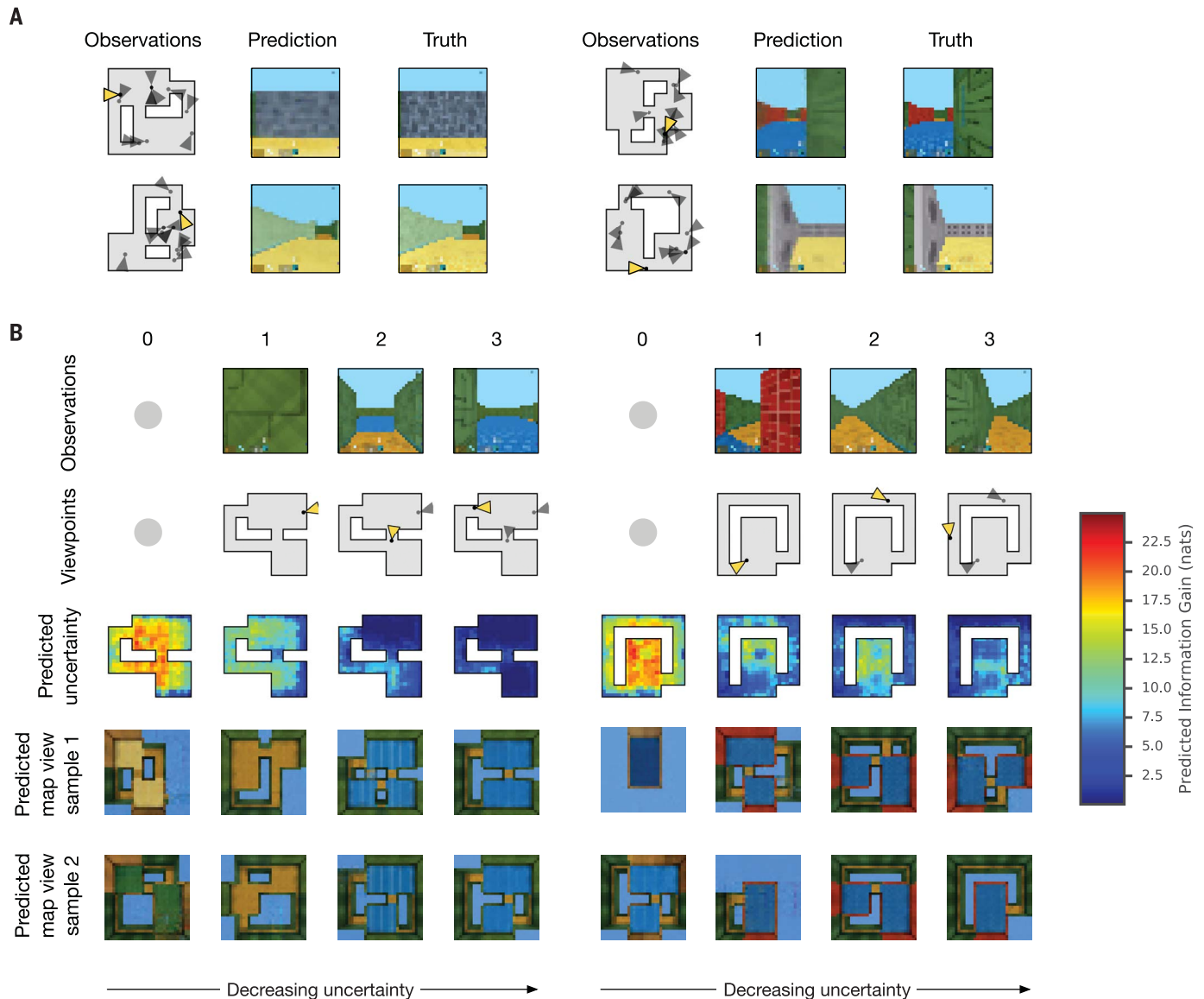
representation more directly by training a separate generator to predict a top-down view of the maze and found that it yields highly accurate predictions (Fig. 6B). The model's uncertainty, as measured by the entropy of its first-person and top-down samples, decreases as more observations are made [Fig. 6B and section 3 of (17)]. After about only five observations, the GQN's uncertainty disappears almost entirely.

### Related work

GQN offers key advantages over prior work. Traditional structure-from-motion, structure-

from-depth, and multiview geometry techniques (12–16, 21) prescribe the way in which the 3D structure of the environment is represented (for instance, as point clouds, mesh clouds, or a collection of predefined primitives). GQN, by contrast, learns this representational space, allowing it to express the presence of textures, parts, objects, lights, and scenes concisely and at a suitably high level of abstraction. Furthermore, its neural formulation enables task-specific fine-tuning of the representation via back-propagation (e.g., via further supervised or reinforced deep learning).

Classical neural approaches to this learning problem—e.g., autoencoding and density models (22–27)—are required to capture only the distribution of observed images, and there is no explicit mechanism to encourage learning of how different views of the same 3D scene relate to one another. The expectation is that statistical compression principles will be sufficient to enable networks to discover the 3D structure of the environment; however, in practice, they fall short of achieving this kind of meaningful representation and instead focus on regularities of colors and patches in the image space.



**Fig. 6. Partial observability and uncertainty.** (A) The agent (GQN) records several observations of a previously unencountered test maze (indicated by gray triangles). It is then capable of accurately predicting the image that would be observed at a query viewpoint (yellow triangle). It can accomplish this task only by aggregating information across multiple observations. (B) In the  $k$ th column, we condition GQN on observations 1 to  $k$  and show GQN's predicted uncertainty, as well as two of GQN's

sampled predictions of the top-down view of the maze. Predicted uncertainty is measured by computing the model's Bayesian surprise at each location, averaged over three different heading directions. The model's uncertainty decreases as more observations are made. As the number of observations increases, the model predicts the top-down view with increasing accuracy. See section 3 of (17), fig. S8, and movie S1 for further details and results. nats, natural units of information.

Viewpoint transformation networks do explicitly learn this relationship; however, they have thus far been nonprobabilistic and limited in scale—e.g., restricted to rotation around individual objects for which a single view is sufficient for prediction (15, 28–33) or to small camera displacements between stereo cameras (34–36).

By employing state-of-the-art deep, iterative, latent variable density models (25), GQN is capable of handling free agent movement around scenes containing multiple objects. In addition, owing to its probabilistic formulation, GQN can account for uncertainty in its understanding about a scene's contents in the face of severe occlusion and partial observability. Notably, the GQN framework is not specific to the particular choice of architecture of the generation network, and alternatives such as generative adversarial networks (37) or autoregressive models (38) could be employed.

A closely related body of work is that of discriminative pose estimation (39–41), in which networks are trained to predict camera motion between consecutive frames. The GQN formulation is advantageous, as it allows for aggregation of information from multiple images of a scene (see maze experiments); it is explicitly probabilistic, allowing for applications such as exploration through Bayesian information gain; and, unlike the aforementioned methods where scene representation and pose prediction are intertwined, the GQN architecture admits a clear architectural separation between the representation and generation networks. The idea of pose estimation is complementary, however—the GQN can be augmented with a second “generator” that, given an image of a scene, predicts the viewpoint from which it was taken, providing a new source of gradients with which to train the representation network.

## Outlook

In this work, we have shown that a single neural architecture can learn to perceive, interpret, and represent synthetic scenes without any human labeling of the contents of these scenes. It can also learn a powerful neural renderer that is capable of producing accurate and consistent images of scenes from new query viewpoints. The GQN learns representations that adapt to and compactly capture the important details of its environment (e.g., the positions, identities, and colors of multiple objects; the configuration of the joint angles of a robot arm; and the layout of a maze), without any of these semantics being built into the architecture of the networks. GQN uses analysis-by-synthesis to perform “inverse graphics,” but unlike existing methods (42), which require problem-specific engineering in the design of their generators, GQN learns this behavior by itself and in a generally applicable manner. However, the resulting representations are no longer directly interpretable by humans.

Our experiments have thus far been restricted to synthetic environments for three reasons: (i) a need for controlled analysis, (ii) limited availability of suitable real datasets, and (iii) limitations

of generative modeling with current hardware. Although the environments are relatively constrained in terms of their visual fidelity, they capture many of the fundamental difficulties of vision—namely, severe partial observability and occlusion—as well as the combinatorial, multi-object nature of scenes. As new sources of data become available (41) and advances are made in generative modeling capabilities (37, 43), we expect to be able to investigate application of the GQN framework to images of naturalistic scenes.

Total scene understanding involves more than just representation of the scene's 3D structure. In the future, it will be important to consider broader aspects of scene understanding—e.g., by querying across both space and time for modeling of dynamic and interactive scenes—as well as applications in virtual and augmented reality and exploration of simultaneous scene representation and localization of observations, which relates to the notion of simultaneous localization and mapping in computer vision.

Our work illustrates a powerful approach to machine learning of grounded representations of physical scenes, as well as of the associated perception systems that holistically extract these representations from images, paving the way toward fully unsupervised scene understanding, imagination, planning, and behavior.

## REFERENCES AND NOTES

1. A. Krizhevsky, I. Sutskever, G. E. Hinton, in *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, F. Pereira, C. J. C. Burges, L. Bottou, K. Q. Weinberger, Eds. (Curran Associates, 2012), pp. 1097–1105.
2. B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, A. Oliva, in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger, Eds. (Curran Associates, 2014), pp. 487–495.
3. S. Ren, K. He, R. Girshick, J. Sun, in *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, R. Garnett, Eds. (Curran Associates, 2015), pp. 91–99.
4. R. Girshick, J. Donahue, T. Darrell, J. Malik, in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2014), pp. 580–587.
5. M. C. Mozer, R. S. Zemel, M. Behrmann, in *Advances in Neural Information Processing Systems 4 (NIPS 1991)*, J. E. Moody, S. J. Hanson, R. P. Lippmann, Eds. (Morgan-Kaufmann, 1992), pp. 436–443.
6. J. Komorski, *Science* **160**, 652–653 (1968).
7. D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Henry Holt and Co., 1982).
8. D. Hassabis, E. A. Maguire, *Trends Cogn. Sci.* **11**, 299–306 (2007).
9. D. Kumaran, D. Hassabis, J. L. McClelland, *Trends Cogn. Sci.* **20**, 512–534 (2016).
10. B. M. Lake, R. Salakhutdinov, J. B. Tenenbaum, *Science* **350**, 1332–1338 (2015).
11. S. Becker, G. E. Hinton, *Nature* **355**, 161–163 (1992).
12. Z. Wu et al., in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015), pp. 1912–1920.
13. J. Wu, C. Zhang, T. Xue, W. Freeman, J. Tenenbaum, in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Curran Associates, 2016), pp. 82–90.
14. D. J. Rezende et al., in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Curran Associates, 2016), pp. 4996–5004.
15. X. Yan, J. Yang, E. Yumer, Y. Guo, H. Lee, in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Curran Associates, 2016), pp. 1696–1704.
16. M. Pollefeys et al., *Int. J. Comput. Vision* **59**, 207–232 (2004).
17. See supplementary materials.
18. L. van der Maaten, *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
19. I. Higgins et al., at International Conference on Learning Representations (ICLR) (2017).
20. T. Mikolov et al., in *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, K. Q. Weinberger, Eds. (Curran Associates, 2013), pp. 3111–3119.
21. Y. Zhang, W. Xu, Y. Tong, K. Zhou, *ACM Trans. Graph.* **34**, 159 (2015).
22. D. P. Kingma, M. Welling, arXiv:1312.6114 [stat.ML] (20 December 2013).
23. D. J. Rezende, S. Mohamed, D. Wierstra, in *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)* (JMLR, 2014), vol. 32, pp. 1278–1286.
24. I. Goodfellow et al., in *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger, Eds. (Curran Associates, 2014), pp. 2672–2680.
25. K. Gregor, F. Besse, D. J. Rezende, I. Danihelka, D. Wierstra, in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Curran Associates, 2016), pp. 3549–3557.
26. P. Vincent, H. Larochelle, Y. Bengio, P.-A. Manzagol, in *Proceedings of the 25th International Conference on Machine Learning (ICML 2008)* (ACM, 2008), pp. 1096–1103.
27. P. Dayan, G. E. Hinton, R. M. Neal, R. S. Zemel, *Neural Comput.* **7**, 889–904 (1995).
28. G. E. Hinton, A. Krizhevsky, S. D. Wang, in *Proceedings of the 21st International Conference on Artificial Neural Networks and Machine Learning (ICANN 2011)*, T. Honkela, W. Duch, M. Girolami, S. Kaski, Eds. (Lecture Notes in Computer Science Series, Springer, 2011), vol. 6791, pp. 44–51.
29. C. B. Choy, D. Xu, J. Gwak, K. Chen, S. Savarese, in *Proceedings of the 2016 European Conference on Computer Vision (ECCV)* (Lecture Notes in Computer Science Series, Springer, 2016), vol. 1, pp. 628–644.
30. M. Tatarchenko, A. Dosovitskiy, T. Brox, in *Proceedings of the 2016 European Conference on Computer Vision (ECCV)* (Lecture Notes in Computer Science Series, Springer, 2016), vol. 9911, pp. 322–337.
31. F. Anselmi et al., *Theor. Comput. Sci.* **633**, 112–121 (2016).
32. D. F. Fouhey, A. Gupta, A. Zisserman, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2016), pp. 1516–1524.
33. A. Dosovitskiy, J. T. Springenberg, M. Tatarchenko, T. Brox, *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 692–705 (2017).
34. C. Godard, O. Mac Aodha, G. J. Brostow, in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2017), pp. 6602–6611.
35. T. Zhou, S. Tulsiani, W. Sun, J. Malik, A. A. Efros, in *Proceedings of the 2016 European Conference on Computer Vision (ECCV)* (Lecture Notes in Computer Science Series, Springer, 2016), pp. 286–301.
36. J. Flynn, I. Neulander, J. Philbin, N. Snavely, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2016), pp. 5515–5524.
37. T. Karras, T. Aila, S. Laine, J. Lehtinen, arXiv:1710.10196 [cs.NE] (27 October 2017).
38. A. van den Oord et al., in *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett, Eds. (Curran Associates, 2016), pp. 4790–4798.
39. D. Jayaraman, K. Grauman, in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2015), pp. 1413–1421.
40. P. Agrawal, J. Carreira, J. Malik, arXiv:1505.01596 [cs.CV] (7 May 2015).
41. A. R. Zamir et al., in *Proceedings of the 2016 European Conference on Computer Vision (ECCV)* (Lecture Notes in Computer Science Series, Springer, 2016), pp. 535–553.
42. T. D. Kulkarni, P. Kohli, J. B. Tenenbaum, V. Mansinghka, in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2015), pp. 4390–4399.
43. Q. Chen, V. Koltun, in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE, 2017), pp. 1511–1520.
44. A. A. Rusu et al., arXiv:1610.04286 [cs.RO] (13 October 2016).



## ACKNOWLEDGMENTS

We thank M. Shanahan, A. Zisserman, P. Dayan, J. Leibo, P. Battaglia, and G. Wayne for helpful discussions and advice; G. Ostrovski, N. Heess, D. Zoran, V. Nair, and D. Silver for reviewing the paper; K. Anderson for help creating environments; and the rest of the DeepMind team for support and ideas. **Funding:** This research was funded by DeepMind. **Author contributions:** S.M.A.E. and D.J.R. conceived of the model. S.M.A.E., D.J.R., F.B., and F.V. designed and implemented the model, datasets, visualizations, figures, and videos. A.S.M. and A.R. designed and performed analysis

experiments. M.G. and A.A.R. performed robot arm experiments. I.D., D.P.R., O.V., and D.R. assisted with maze navigation experiments. L.B. and T.W. assisted with Shepard-Metzler experiments. H.K., C.H., K.G., M.B., D.W., N.R., K.K., and D.H. managed, advised, and contributed ideas to the project. S.M.A.E. and D.J.R. wrote the paper. **Competing interests:** The authors declare no competing financial interests. DeepMind has filed a U.K. patent application (GP-201495-00-PCT) related to this work. **Data and materials availability:** Datasets used in the experiments have been made available to download at <https://github.com/deepmind/gqn-datasets>.

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/1204/suppl/DC1](http://www.sciencemag.org/content/360/6394/1204/suppl/DC1)  
Supplementary Text  
Figs. S1 to S16  
Algorithms S1 to S3  
Table S1  
References (45–52)  
Movie S1

29 November 2017; accepted 10 April 2018  
10.1126/science.aar6170

## REPORT

## PHOTOSYNTHESIS

# Photochemistry beyond the red limit in chlorophyll f-containing photosystems

Dennis J. Nürnberg<sup>1\*</sup>, Jennifer Morton<sup>2</sup>, Stefano Santabarbara<sup>3</sup>, Alison Telfer<sup>1</sup>, Pierre Joliot<sup>4</sup>, Laura A. Antonaru<sup>1</sup>, Alexander V. Ruban<sup>5</sup>, Tanai Cardona<sup>1</sup>, Elmars Krausz<sup>3</sup>, Alain Boussac<sup>6</sup>, Andrea Fantuzzi<sup>1\*</sup>, A. William Rutherford<sup>1\*</sup>

Photosystems I and II convert solar energy into the chemical energy that powers life. Chlorophyll a photochemistry, using red light (680 to 700 nm), is near universal and is considered to define the energy "red limit" of oxygenic photosynthesis. We present biophysical studies on the photosystems from a cyanobacterium grown in far-red light (750 nm). The few long-wavelength chlorophylls present are well resolved from each other and from the majority pigment, chlorophyll a. Charge separation in photosystem I and II uses chlorophyll f at 745 nm and chlorophyll f (or d) at 727 nm, respectively. Each photosystem has a few even longer-wavelength chlorophylls f that collect light and pass excitation energy uphill to the photochemically active pigments. These photosystems function beyond the red limit using far-red pigments in only a few key positions.

Oxygenic photosynthesis uses chlorophyll a (chl a) to convert visible light into chemical energy. The photochemically active pigments of the two photosystems, photosystem I (PSI) and photosystem II (PSII), at 700 nm and 680 nm, respectively, represent the energy available for photochemistry (1–3) (see supplementary text S1). This was considered the "red limit" (4), the minimum energy required for oxygenic photosynthesis, until it was found that a cyanobacterium, *Acaryochloris*, extends this limit by using chl d, a pigment absorbing at a wavelength 40 nm longer than chl a (5, 6). *Acaryochloris* photosystems are dominated (~97%) by chl d with one or two chl a located in key positions (6, 7). More recently, chl f, the longest-wavelength chlorophyll known, was discovered (6, 8, 9). Chl f exists as a minority pigment in photosystems that contain ~90% chl a. Chl f is generally assumed to play a purely light-harvesting role, needing heat in the environment, the product of the Boltzmann constant and the temperature ( $k_B T$ ), for uphill excitation transfer to chl a for photochemistry to occur (5, 6, 10–13)

[see, however, (14)]. Thus, the presence of chl f in these photosystems was not considered an extension of the photochemical red limit. Here, we report data that change that view, showing that long-wavelength chlorophylls perform the photochemistry in both photosystems.

*Chroococcidiopsis thermalis*, a widely spread extremophile cyanobacterium, when grown under 750-nm far-red light (FRL), contains ~90% chl a, ~10% chl f and <1% chl d. The absorption and fluorescence spectra of the cells were shifted to longer wavelengths with a new absorption peak at 709 nm (Fig. 1A) and 80-K fluorescence at 740, 753, and 820 nm, suggesting changes in both photosystems (Fig. 1B) [see (10, 15)]. These data and decay-associated fluorescence (fig. S1) indicate that the FRL chlorophylls are the dominant terminal emitters and are present in essentially all the photosystems [see also (9, 10)]. The wavelength dependence of PSI and PSII activity (action spectra) in FRL cells were red-shifted (Fig. 1C), with new peaks at 745 nm for PSI and 715 nm for PSII. The action spectra correspond well to the absorption spectra of the isolated photosystems (Fig. 1C).  $P_{700}^{++}$  formation with far-red light was more efficient in FRL cells (Fig. 1D, right). Primary quinone ( $Q_A$ ) reduction under far-red excitation in white-light (WL) cells was slow and incomplete, but in FRL cells it was very rapid, faster than in WL cells with visible excitation (Fig. 1D, lower middle).  $O_2$  evolution under far-red light reached the same level as seen under visible light in WL cells (see supplementary text S2).

Thermoluminescence (TL) from  $S_2Q_B^-$  (Fig. 2A) and  $S_2Q_A^-$  (fig. S2A) (16) in FRL thylakoids is enhanced by a factor of >25 compared with WL

thylakoids. TL arises from repopulation of the excited state of the primary donor chlorophyll,  $^*Chl_{D1}$ , by charge recombination, which then emits at the same wavelength as the prompt fluorescence (17). The enhancement in TL intensity in FRL PSII indicates a decrease in the energy gap between the precursor charge pair and the luminescent excited-state chlorophyll (17), consistent with a lower-energy, longer-wavelength primary donor/emitter (supplementary text S3). Recombination takes place at temperatures close to those seen in the WL PSII, and the flash-number dependence is nearly identical to that seen in WL PSII, indicating no obvious change in the yields of charge separation and stabilization (fig. S2B).

PSII activity, measured as the rates of  $Q_A$  reduction at 293 K (Fig. 2B, inset) and 77 K (Fig. 2B) and as the rate of  $\beta$ -carotene cation radical formation at 15 K (18, 19), were comparable with red and far-red excitation (Fig. 2C and fig. S3). A purely antenna role for chl f (5, 6, 10–13) would require that heat ( $k_B T$ ) drives the excitation from the shortest-wavelength chl f (~720 nm) uphill to  $P_{680}$ , the chl a primary donor. The  $k_B T$  at room temperature (293 K) is 26 meV and too small to cover the ~100 meV energy gap between 720 and 680 nm. Although some overlap of the absorption bands could allow this process to occur inefficiently at 293 K, thermally induced excitation transfer will become vanishingly weak at 77 K ( $k_B T = 6.6$  meV) and 15 K ( $k_B T = 1.3$  meV) (supplementary text S4). The matching rates for PSII photochemistry with red and far-red excitation at cryogenic temperatures indicate that a FRL chlorophyll is involved in primary charge separation.

When the  $Pheo_{D1}^-$  is trapped by illumination in the presence of a reductant (20, 21) the light-minus-dark difference spectrum exhibits the expected bleaches at ~546 and ~680 nm (Fig. 2D) due to the disappearance of the  $Pheo_{D1} Q_y$  and  $Q_x$  absorptions. In WL PSII, the charge on  $Pheo_{D1}^-$  induces a blue shift of the  $Chl_{D1} Q_y$  absorption at ~680 nm. This electrochromic shift dominates the difference spectrum (20, 21) but is notably absent in the FRL PSII (Fig. 2D). Instead, a marked blue shift at ~720 nm is seen. This indicates that  $Chl_{D1}$  is a long-wavelength chlorophyll in FRL PSII. The smaller blue-shift feature at ~669 nm is attributed to  $P_{D1}/P_{D2}$ .

When  $Q_A^-$  is trapped by illumination at 77 K (Fig. 2E), the charge created on  $Q_A^-$  is known to induce a blue shift of the  $Q_y$  band of  $Chl_{D1}$  (21). In FRL PSII, the  $Chl_{D1}$  blue shift typically seen near ~680 nm (21) is replaced by one at ~727 nm. This indicates that  $Chl_{D1}$  is replaced by a FRL chlorophyll (Fig. 2F). The absence of the  $Chl_{D1}$  band shift at 680 nm allows a 685-nm red shift to become evident. This corresponds to the expected (but not previously visible)  $Q_A^-$ -induced shift of the  $Pheo_{D1} Q_y$  excitation (supplementary text S2).

High-performance liquid chromatography (HPLC) indicated that the isolated PSII contained 2 pheo a, 4 chl f, 1 chl d, and 30 chl a per center, based on the 35 chl per PSII (22) (fig. S5A). The 77-K absorption spectrum (Fig. 2E) shows four peaks above 700 nm. The 710-nm

<sup>1</sup>Department of Life Sciences, Imperial College, London SW7 2AZ, UK. <sup>2</sup>Research School of Chemistry, ANU, Canberra, Australia. <sup>3</sup>Istituto di Biofisica, Consiglio Nazionale delle Ricerche, via Celoria 26, 20133 Milano, Italy. <sup>4</sup>Institut de Biologie Physico-Chimique, Unité Mixte de Recherche 7141 Centre National de la Recherche Scientifique-Université Pierre et Marie Curie, 13 Rue Pierre et Marie Curie, 75005 Paris, France. <sup>5</sup>School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK. <sup>6</sup>Institut de Biologie Intégrative de la Cellule, UMR 9198, Bât 532, CEA Saclay, 91191 Gif-sur-Yvette, France. \*Corresponding author. Email: a.rutherford@imperial.ac.uk (A.W.R.); d.nuernberg@imperial.ac.uk (D.J.N.); a.fantuzzi@imperial.ac.uk (A.F.)



band is attributed to red-shifted allophycocyanins (9, 23), based on its wavelength, variable amplitude in different preparations, and absence of magnetic circular dichroism (MCD) (fig. S6). The 1.8-K absorption and MCD spectra were fitted with ~5 FRL chlorophylls at 721, 727, 734, 737, and 749 nm (fig. S7).

Intrinsically, chl d absorbs at shorter wavelengths than chl f; thus, it is the most obvious choice for the 721-nm pigment. This pigment could be a linker between the primary donor at 727 nm and the shorter-wavelength antenna (supplementary text S5). Given the small wavelength difference, we need to consider the possibility that chl d is the primary donor, Chl<sub>D1</sub>. The wavelength (727 nm) of the primary donor is similar to the chl d (725 nm) in this location in *Acaryochloris* (7, 24).

The conserved amino acid changes near Chl<sub>D1</sub> (figs. S8 and S9) suggest specific H bonding to a formyl group on either the C2 (chl f) or C3 (chl d)

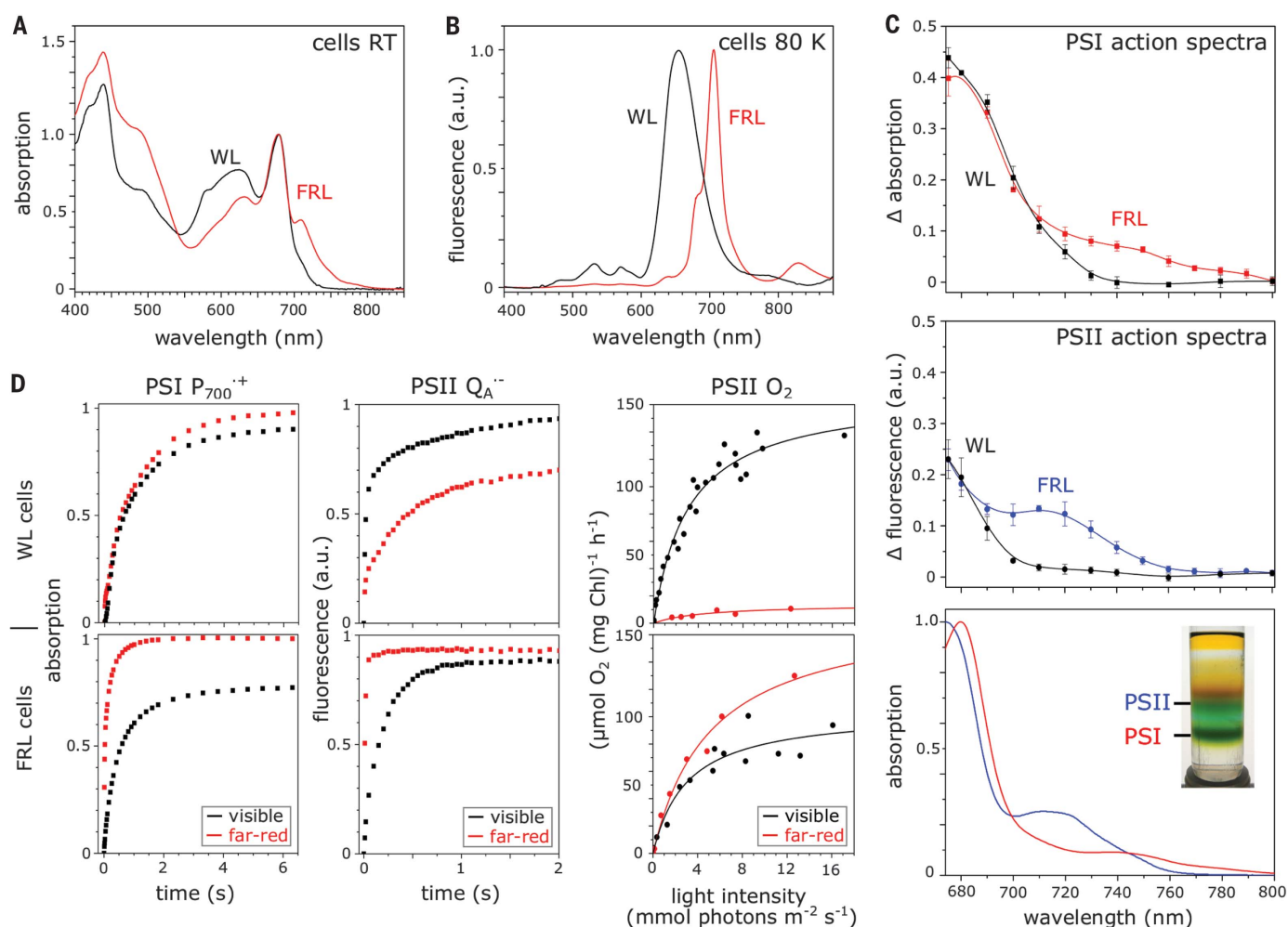
position of Chl<sub>D1</sub>. This provides further evidence that Chl<sub>D1</sub> in FRL PSII is a long-wavelength chlorophyll but does not discriminate between chl d and chl f (fig. S8). The three remaining long-wavelength chlorophylls appear to be tuned to span the energy gap from 749 nm up to the photochemically active chlorophyll at 727 nm, with gaps of 10 to 12 nm (23 to 26 meV), appropriate to the value of  $k_B T$  at ambient temperatures (~26 meV) (see supplementary text S5 and table S1 for other pigment assignments).

It seems likely that P<sub>D1</sub>, which bears the key long-lived chlorophyll cation radical, remains a chl a. This suggestion is based on (i) the presence of the blue shifts at 669 nm attributed to P<sub>D1</sub> (Fig. 2, D and E), (ii) structural considerations (figs. S8 and S12), (iii) the precedence of chl a being P<sub>D1</sub> in *Acaryochloris* (6, 7, 24), and (iv) conservation of specific chemical properties (e.g. oxidizing power, stability, and reactivity)

without the need for major redox and kinetic tuning. Nevertheless, sequence comparisons indicate changes in the environments of cofactors in the redox core, including P<sub>D1</sub>, likely reflecting the tuning needed to adjust to the presence of the long-wavelength pigment. (see table S2 and figs. S9 and S12 to S14).

When isolated FRL PSI was excited at 15 K using 610-, 730-, and 750-nm light, comparable photochemistry was seen (Fig. 3A and fig. S15), indicating that the chlorophyll absorbing at ~750 nm is involved in charge separation. In WL PSI, 750-nm excitation resulted in the expected low-quantum-yield photochemistry (Fig. 3A) due to weak optical charge-transfer absorption at 750 nm (supplementary text S4).

Light-minus-dark difference spectra at 293 K in FRL and WL thylakoids are similar to those obtained at 77 and 1.8 K from isolated FRL and WL PSI, except for thermal broadening at higher temperatures (Fig. 3B and fig. S16). The



**Fig. 1. Changes in the pigments and activities of the photosystems from *C. thermalis* cells grown in 750-nm light.** (A) Absorption spectra and (B) 80 K fluorescence spectra of WL and FRL cells. (C) Action spectra for PSI ( $P_{700}^{++}$  at 430 nm, 50  $\mu$ s after the flash) (top) and PSII ( $Q_A^{--}$  + DCMU, measured as fluorescence) (middle) in FRL cells (red and blue, respectively) compared to WL cells (black). Bottom panel: absorp-

tion spectra from isolated PSI (red) and PSII (blue); inset sucrose gradient separating the photosystems. (D) Photosystem activities measured in WL (top) and FRL cells using visible (black) and far-red (red points) excitation. Left: PSI activity ( $P_{700}^{++}$  at 705 nm 50  $\mu$ s after the flash). Middle: PSII activity ( $Q_A^{--}$  as in C). Right:  $O_2$  evolution as a function of light intensity. See materials and methods for details.

main bleach is at  $\sim 704$  nm with an absorption increase at 692 nm, indicating that the cation is located on the  $P_A P_B$  chlorophyll a/a' pair,  $P_{700}^{++}$ , as seen in WL PSI. The rest of the spectrum shows marked differences: the sharp trough at 684 nm, the peak at 674 nm (680 nm at 77 K), and the trough at 656 nm ( $\sim 660$  nm at 77 K), all of which are present in WL PSI (Fig. 3B and fig. S16) but are absent in FRL PSI. Instead, several changes appear at  $>700$  nm (Fig. 3B). These changes can be attributed to the replacement of chl a band

shifts in WL PSI with long-wavelength chlorophyll band shifts in FRL PSI. The band shifts at  $>700$  nm are better resolved at 77 and 1.8 K and show a broad blue-shift feature (or two overlapping blue shifts) at  $\sim 745$  nm, a sharp red shift at 756 nm, and a weak blue shift at 800 nm. In addition, there is a red shift at  $\sim 684$  nm, which is likely present in WL PSI but now resolved in the FRL PSI.

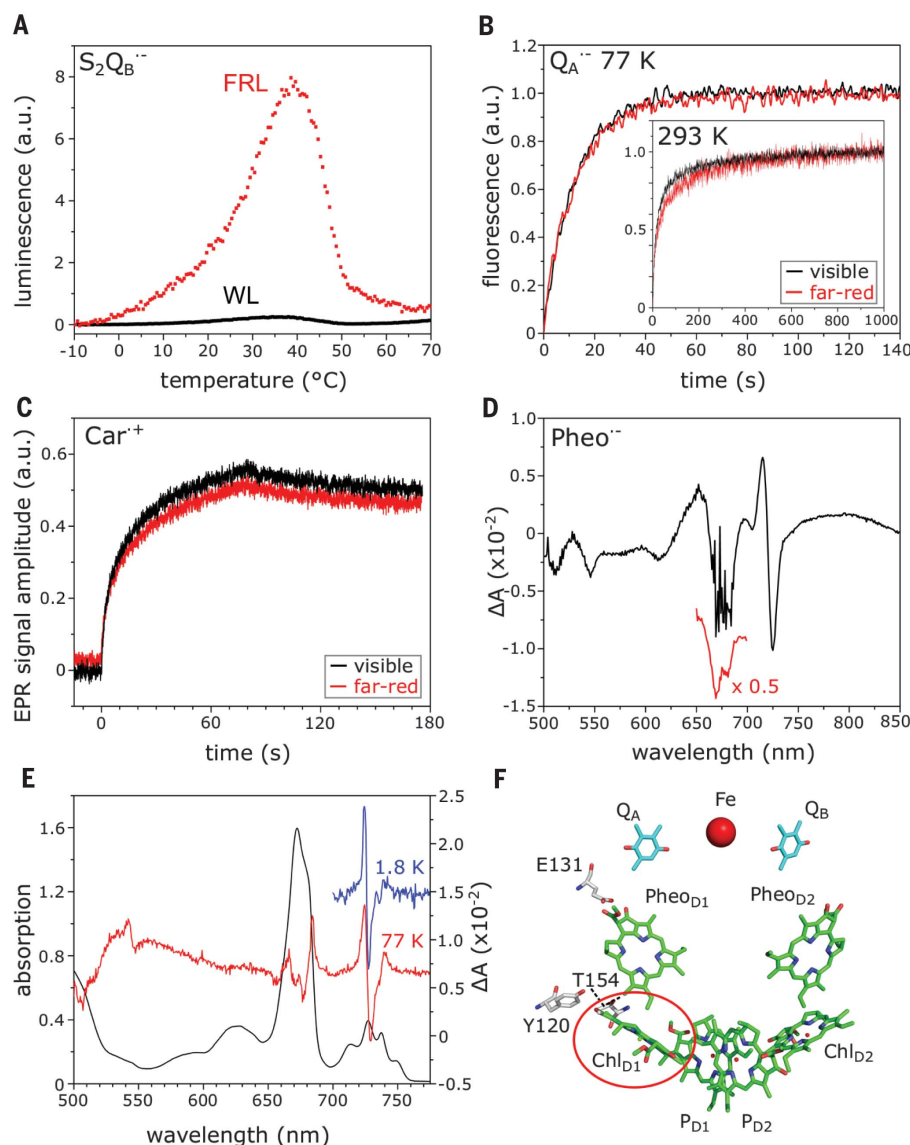
HPLC shows that FRL PSI contains 0 pheo, 7 to 8 chl f, 0 chl d, and  $\sim 88$  to 89 chl a (fig. S5B), assuming 96 chl per PSI (25). The 1.8 K absorp-

tion and MCD spectra at  $>700$  nm (fig. S17) can be fitted as follows: 1 chl f at 736 nm, 3 chl f at  $\sim 745$  nm, 1 chl f at 756 nm, 1 chl f at  $\sim 763$  nm, and  $\sim 2$  chl f at  $\sim 800$  nm (fig. S18). In addition, a chlorophyll feature at 709 nm is suggested to be a long-wavelength PSI chl a.

Based on the crystal structure (25), the difference spectra (Fig. 3B and fig. S16), the low-temperature photochemistry (Fig. 3A and fig. S15), and structural arguments (figs. S8 and S19), the following model emerges for the FRL PSI redox cofactors: (i)  $P_A$  and  $P_B$  remain chl a/chl a' and bear the  $P_{700}$  cation. (ii)  $A_{0A}$  and  $A_{0B}$ , the primary acceptors (1, 26), remain chl a, because the redox properties of chl f (13) make it unsuitable for a low potential role. This fits with the assignment of the  $\sim 684$ -nm red shift to  $A_0$  (Fig. 3B) [see (26)] and with  $A_0$  being chl a in *Acaryochloris* (24). (iii) One or both of the primary donors,  $A_{1A}$  and  $A_{1B}$  (27), are chl f in FRL PSI, absorbing at  $\sim 745$  nm. The difference spectra seem to arise from overlapping blue shifts at this wavelength, replacing the dominant  $\sim 684$ -nm blue shift in the WL PSI.  $A_{1A}$  and  $A_{1B}$  are the only redox-active chlorophylls oriented to allow a blue shift. Possible locations of the six chl f antenna are discussed in the supplementary materials (supplementary text S6).

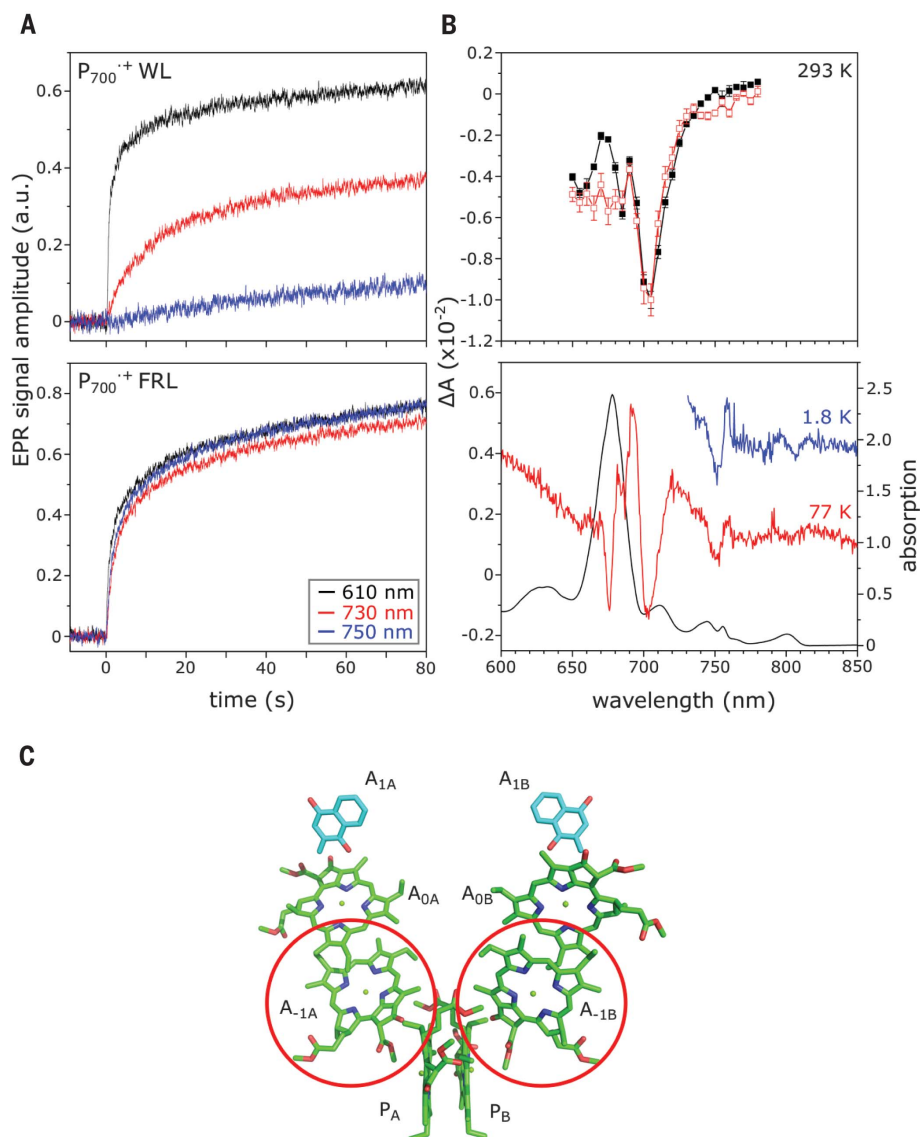
The far-red photosystems studied here represent a third paradigm for oxygenic photosynthesis (supplementary text S1). The first is the canonical type that use chl a for both charge separation and antenna absorption. The second is that of *Acaryochloris*, where chl d replaces all chl a for both photochemistry and light harvesting, with the exception of chl a being retained in two key redox roles:  $P_{D1}$ , where  $P_{D1}^{++}$  is the principle oxidant of PSII, and  $A_0$ , where  $A_0^{+}$  is the principle reductant in PSI. The new paradigm maintains chl a in nearly all positions except for a small number of chl f molecules (plus one chl d in PSII). These long-wavelength chlorophylls act as (i) the primary electron donors in both photosystems, (ii) linkers to shorter-wavelength antenna pigments, and (iii) a small, longer-wavelength antenna system. The far-red photosystems provide well-resolved absorption spectra at long wavelengths due to the decreased spectral overlap, which allows more definitive assignments and analyses.

Chl f can absorb at  $>760$  nm, yet the wavelength used to drive photochemistry in the FRL PSII is  $\sim 727$  nm (fig. S20). This wavelength is similar to that of the chl d primary donor ( $Chl_{D1}$ ) in *Acaryochloris* PSII (6, 7, 24). Thus, both of the known cases of PSII functioning "beyond the red limit" use primary donors with similar energies. This may represent a "second red limit" for PSII, one that applies in the stable, deep-shade environments where far-red photosynthesis occurs in nature. The  $\sim 110$  meV of energy sacrificed by the shift from  $\sim 680$  to 727 nm may correspond to the energy "headroom" needed by PSII to mitigate the photodamage caused by variable light intensities (3, 28, 29). This potential disadvantage may explain why oxygenic photosynthesis using far-red light is restricted to rare, deep-shade



**Fig. 2. FRL PSII photochemistry.** (A)  $S_2Q_B^{--}$  thermoluminescence in WL and FRL thylakoids from *C. thermalis*. (B) Kinetics of  $Q_A$  reduction in FRL thylakoids (measured as fluorescence) at 77 K (inset at 293 K). Samples were excited using visible (600-nm) and far-red (710-nm) light. (C) Kinetics of  $\beta$ -carotene cation radical formation in isolated FRL PSII at 15 K illuminated with visible (610-nm) and far-red (730-nm) light. (D) Light-minus-dark visible/near-IR spectrum at 293 K upon  $Pheo_{D1}^{--}$  formation in isolated FRL PSII. The 650- to 700-nm region is saturated in the main spectrum (black), the red inset is a 2x-diluted sample. (E) 77-K absorption spectrum of isolated FRL PSII (black) with the 77-K (red) and 1.8-K (blue) 725-nm illuminated-minus-dark difference spectra. (F) Cofactors in PSII showing  $Chl_{D1}$  as a FRL chlorophyll (circled) and potential H-bonding amino acids, based on the crystal structure PDB 3WU2 (22).





**Fig. 3. FRL PSI photochemistry.** (A) 15-K photo-accumulation of  $P_{700}^{++}$  monitored by EPR using visible and far-red light in (top) isolated FRL PSI and (bottom) isolated WL PSI. (B) (Top) Flash-induced  $P_{700}^{++}$  293-K absorption difference spectrum using FRL and WL thylakoids. (Bottom)  $P_{700}^{++}$  77-K (red) and 1.8-K (blue) difference spectra and 1.8-K absorption spectrum (black) of isolated FRL PSI. (C) FRL PSI model (PDB 1JB0) (25). Red circles show the proposed locations of chl f, with other positions as chl a (see the text).

environments, rich in far-red light (supplementary text S7).

#### REFERENCES AND NOTES

- J. H. Golbeck, Ed., *Photosystem I. The Light-Driven Plastocyanin: Ferredoxin Oxidoreductase* (Springer Netherlands, 2006).
- H. Dau, I. Zaharieva, *Acc. Chem. Res.* **42**, 1861–1870 (2009).
- A. W. Rutherford, A. Osyczka, F. Rappaport, *FEBS Lett.* **586**, 603–616 (2012).
- L. O. Björn, G. C. Papageorgiou, R. E. Blankenship, Govindjee, *Photosynth. Res.* **99**, 85–98 (2009).
- P. Loughlin, Y. Lin, M. Chen, *Photosynth. Res.* **116**, 277–293 (2013).
- H. Miyashita et al., *J. Phys. Chem. Biophys.* **4**, 149 (2014).
- T. Renger, E. Schlödder, *J. Phys. Chem. B* **112**, 7351–7354 (2008).
- M. Chen et al., *Science* **329**, 1318–1319 (2010).
- F. Gan et al., *Science* **345**, 1312–1317 (2014).
- S. Itoh et al., *Plant Cell Physiol.* **56**, 2024–2034 (2015).
- S. I. Allakhverdiev et al., *Biochem.* **81**, 201–212 (2016).
- D. M. Niedzwiedzki, H. Liu, M. Chen, R. E. Blankenship, *Photosynth. Res.* **121**, 25–34 (2014).
- M. Kobayashi et al., in *Photosynthesis*, Z. Dubinsky, Ed. (Intech, 2013).
- M. Kaucikas, D. Nürnberg, G. Dorhac, A. W. Rutherford, J. J. van Thor, *Biophys. J.* **112**, 234–249 (2017).
- C. Zhao, F. Gan, G. Shen, D. A. Bryant, *Front. Microbiol.* **6**, 1–13 (2015).
- A. W. Rutherford, A. R. Crofts, Y. Inoue, *Biochim. Biophys. Acta* **682**, 457–465 (1982).
- F. Rappaport, J. Lavergne, *Photosynth. Res.* **101**, 205–216 (2009).
- W. W. I. Adams 3rd, B. Demmig-Adams, K. Winter, U. Schreiber, *Planta* **180**, 166–174 (1990).
- J. Hanley, Y. Deligiannakis, A. Pascal, P. Faller, A. W. Rutherford, *Biochemistry* **38**, 8189–8195 (1999).
- V. V. Klimov, A. V. Klevanik, V. A. Shuvalov, A. A. Kransnovsky, *FEBS Lett.* **82**, 183–186 (1977).
- N. Cox et al., *J. Phys. Chem. B* **113**, 12364–12374 (2009).
- Y. Umena, K. Kawakami, J.-R. Shen, N. Kamiya, *Nature* **473**, 55–60 (2011).
- Y. Li et al., *Biochim. Biophys. Acta Bioenergetics* **1857**, 107–114 (2016).
- S. Itoh et al., *Biochemistry* **46**, 12473–12481 (2007).
- P. Jordan et al., *Nature* **411**, 909–917 (2001).
- A. Chauvet, N. Dashdorj, J. H. Golbeck, T. W. Johnson, S. Savikhin, *J. Phys. Chem. B* **116**, 3380–3386 (2012).
- M. G. Müller, C. Slavov, R. Luthra, K. E. Redding, A. R. Holzwarth, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4123–4128 (2010).
- C. A. R. Cotton et al., *Front. Bioeng. Biotechnol.* **3**, 36 (2015).
- G. A. Davis et al., *eLife* **5**, 1–27 (2016).

#### ACKNOWLEDGMENTS

We gratefully acknowledge our friend and colleague Fabrice Rappaport, who died in January 2016, with whom we worked in the early stages of this project. We thank J. Murray for discussions, B. Nwaobi for technical support, C. Mullineaux for access to a 77-K fluorometer, B. Baillieu and W. Remelli for technical help in obtaining the absorption spectra, J.-M. Ducruet for lending us the thermoluminescence setup, and L. Haigh for her assistance with the HPLC analysis. **Funding:** This work was supported by BBSRC grants BB/L011506/1 and BB/R001383/1, Leverhulme Trust grant RPG-2017-223 and a Wolfson Merit Award from the Royal Society to A.W.R., Australian Research Council grant DP150103137 to E.K., an Imperial College Junior Research Fellowship to T.C. and an ANR-10-INBS-05 (FRISBI) grant to A.B. **Author contributions.** D.J.N. initiated the study, built growth chambers, surveyed, selected and grew the strains. D.J.N., A.F., and A.W.R. conceived of the main experiments, collated results and interpretations, and wrote the article with input, edits, and approval from all authors. D.J.N. with A.T. and A.B. developed and did the isolation of the photosystems. J.M. with E.K. did the low-temperature optical spectroscopy (visible/near IR, CD, MCD, fluorescence). S.S. and D.J.N. did 293 K and 77 K visible/near IR spectroscopy on PSI. D.J.N. and A.T. did the HPLC. A.B. and A.W.R. did the EPR. L.A.A. with D.J.N. did the  $O_2$  evolution. A.F. and D.J.N. did the TL. A.F. did the Pheo<sup>•</sup> trapping and wavelength dependence of  $Q_A^{•-}$  at 293 and 80 K. P.J. and D.J.N. did the wavelength dependence of PSI and PSII activities in cells. D.J.N. and A.W.R. worked with F. Rappaport to obtain the action spectra with advice from P.J. A.V.R. and D.J.N. did the time-resolved fluorescence studies, and S.S. helped with the interpretation. A.F. and T.C. did the modeling studies. E.K. and D.J.N. did the curve fitting, with input from J.M., S.S., A.F., and A.W.R. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary material.

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6394/1210/suppl/DC1  
Materials and Methods  
Supplementary Text S1 to S7  
Figs. S1 to S20  
Tables S1 to S3  
References (30–92)  
Data S1

22 December 2017; accepted 18 April 2018  
10.1126/science.aar8313

## MAGNETISM

# Giant tunneling magnetoresistance in spin-filter van der Waals heterostructures

Tiancheng Song<sup>1\*</sup>, Xinghan Cai<sup>1\*</sup>, Matisse Wei-Yuan Tu<sup>2</sup>, Xiaou Zhang<sup>3</sup>, Bevin Huang<sup>1</sup>, Nathan P. Wilson<sup>1</sup>, Kyle L. Seyler<sup>1</sup>, Lin Zhu<sup>4</sup>, Takashi Taniguchi<sup>5</sup>, Kenji Watanabe<sup>5</sup>, Michael A. McGuire<sup>6</sup>, David H. Cobden<sup>1</sup>, Di Xiao<sup>3,†</sup>, Wang Yao<sup>2,†</sup>, Xiaodong Xu<sup>1,4,†</sup>

Magnetic multilayer devices that exploit magnetoresistance are the backbone of magnetic sensing and data storage technologies. Here, we report multiple-spin-filter magnetic tunnel junctions (sf-MTJs) based on van der Waals (vdW) heterostructures in which atomically thin chromium triiodide ( $\text{CrI}_3$ ) acts as a spin-filter tunnel barrier sandwiched between graphene contacts. We demonstrate tunneling magnetoresistance that is drastically enhanced with increasing  $\text{CrI}_3$  layer thickness, reaching a record 19,000% for magnetic multilayer structures using four-layer sf-MTJs at low temperatures. Using magnetic circular dichroism measurements, we attribute these effects to the intrinsic layer-by-layer antiferromagnetic ordering of the atomically thin  $\text{CrI}_3$ . Our work reveals the possibility to push magnetic information storage to the atomically thin limit and highlights  $\text{CrI}_3$  as a superlative magnetic tunnel barrier for vdW heterostructure spintronic devices.

Many two-dimensional (2D) materials can be incorporated into artificial heterostructures without the need for lattice matching. These materials thus provide a platform for exploring emerging phenomena and device function at the designed atomic interfaces (1, 2). However, magnetic memory and processing applications were out of reach in van der Waals (vdW) heterostructures before the

recent discovery of suitable 2D magnetic materials (3–10). One of these is the magnetic insulator chromium triiodide ( $\text{CrI}_3$ ), which in bilayer form has been found to possess a layered-antiferromagnetic ground state. Magneto-optical Kerr effect (MOKE) measurements suggest that the spins align ferromagnetically out of plane within each layer but antiferromagnetically between layers, resulting in vanishing net magnetization (Fig. 1A, left) (3).

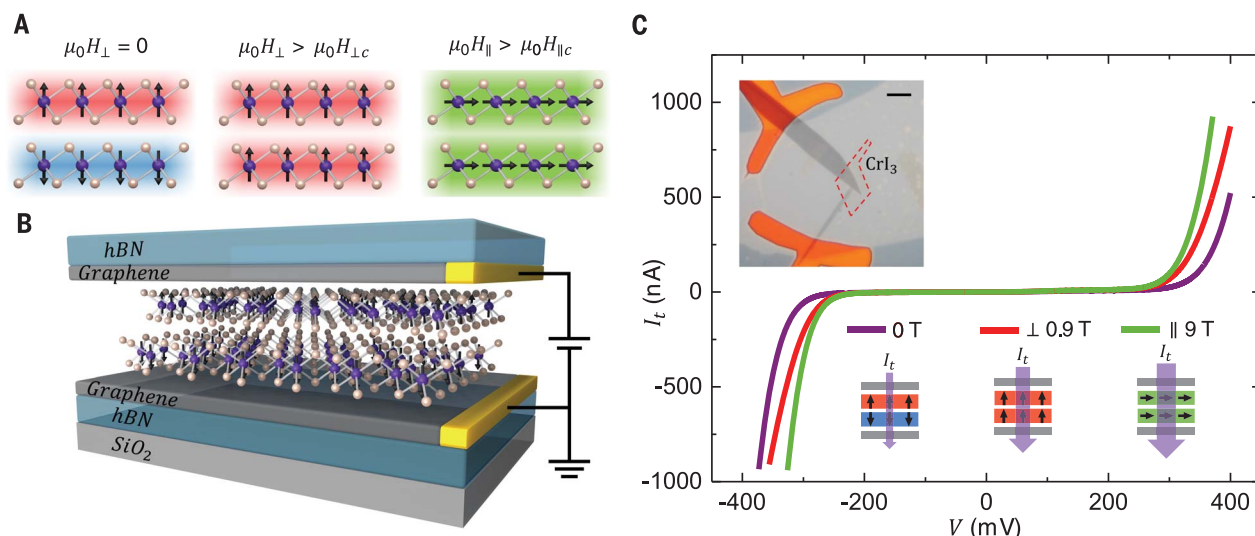
This layered-antiferromagnetic ordering makes  $\text{CrI}_3$  desirable for realizing atomically thin magnetic multilayer devices. When the magnetizations of the two layers in a bilayer are switched between antiparallel (Fig. 1A, left) and parallel states (Fig. 1A, middle and right), giant tunneling magnetoresistance (TMR) is produced by the double spin-filtering effect (11, 12). In general, spin filters, which create and control spin-polarized currents, are the fundamental element in magnetic multilayer devices, such as spin valves (13–15), magnetic tunnel junctions (MTJs) (16–21), and double spin-filter MTJs (sf-MTJs) (11, 12). Compared with the existing magnetic multilayer devices that require different choices of (metallic or insulating) magnets and spacers, the layered-antiferromagnetic structure in bilayer  $\text{CrI}_3$  avoids the need for fabricating separate spin filters with spacers. This guarantees sharp atomic interfaces between spin filters, crucial for achieving large sf-TMR.

An even more intriguing possibility arises if the intrinsic layered-antiferromagnetic structure of  $\text{CrI}_3$  extends beyond the bilayer. In this case,

<sup>1</sup>Department of Physics, University of Washington, Seattle, WA 98195, USA. <sup>2</sup>Department of Physics and Center of Theoretical and Computational Physics, University of Hong Kong, Hong Kong, China. <sup>3</sup>Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA. <sup>4</sup>Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, USA. <sup>5</sup>National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan. <sup>6</sup>Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.

\*These authors contributed equally to this work.

†Corresponding author. Email: xuxd@uw.edu (X.X.); wangyao@hku.hk (W.Y.); dixiao@cmu.edu (D.X.)



**Fig. 1. Spin-filter effects in layered-antiferromagnetic  $\text{CrI}_3$ .** (A) Schematic of magnetic states in bilayer  $\text{CrI}_3$ . (Left) Layered-antiferromagnetic state, which suppresses the tunneling current at zero magnetic field. (Middle and right) Fully spin-polarized states with out-of-plane and in-plane magnetizations, which do not suppress it. (B) Schematic of 2D spin-filter magnetic tunnel

junction (sf-MTJ), with bilayer  $\text{CrI}_3$  functioning as the spin-filter sandwiched between few-layer graphene contacts. (C) Tunneling current of a bilayer  $\text{CrI}_3$  sf-MTJ at selected magnetic fields. (Top inset) Optical microscope image of the device (scale bar, 5  $\mu\text{m}$ ). The red dashed line shows the position of the bilayer  $\text{CrI}_3$ . (Bottom) Schematic of the magnetic configuration for each  $I_t$ - $V$  curve.



every layer should act as another spin filter oppositely aligned in series, greatly enhancing the sf-TMR as the number of layers increases. The associated multiple magnetic states may also enable multiple magnetoresistance states for potentially encoding information in an individual sf-MTJ device. Moreover, being insulators, atomically thin  $\text{CrI}_3$  single crystals can be integrated into vdW heterostructures as tunnel barriers in place of nonmagnetic dielectrics, such as hexagonal boron nitride (hBN) (22, 23) or transition metal dichalcogenides (24), adding magnetic switching functionality. The realization of such vdW heterostructure sf-MTJs could produce novel 2D magnetic interface phenomena (25) and enable spintronics components such as spin current sources and magnetoresistive random-access memory (MRAM) (26).

Here, we demonstrate vdW-engineered sf-MTJs based on atomically thin  $\text{CrI}_3$  with extraordinarily large sf-TMR. Figure 1B shows the essential structure of the sf-MTJ, which consists of two few-layer graphene contacts separated by a thin  $\text{CrI}_3$  tunnel barrier. The sf-MTJ is sandwiched between two hexagonal boron nitride (hBN) flakes to avoid degradation. We have made and in-

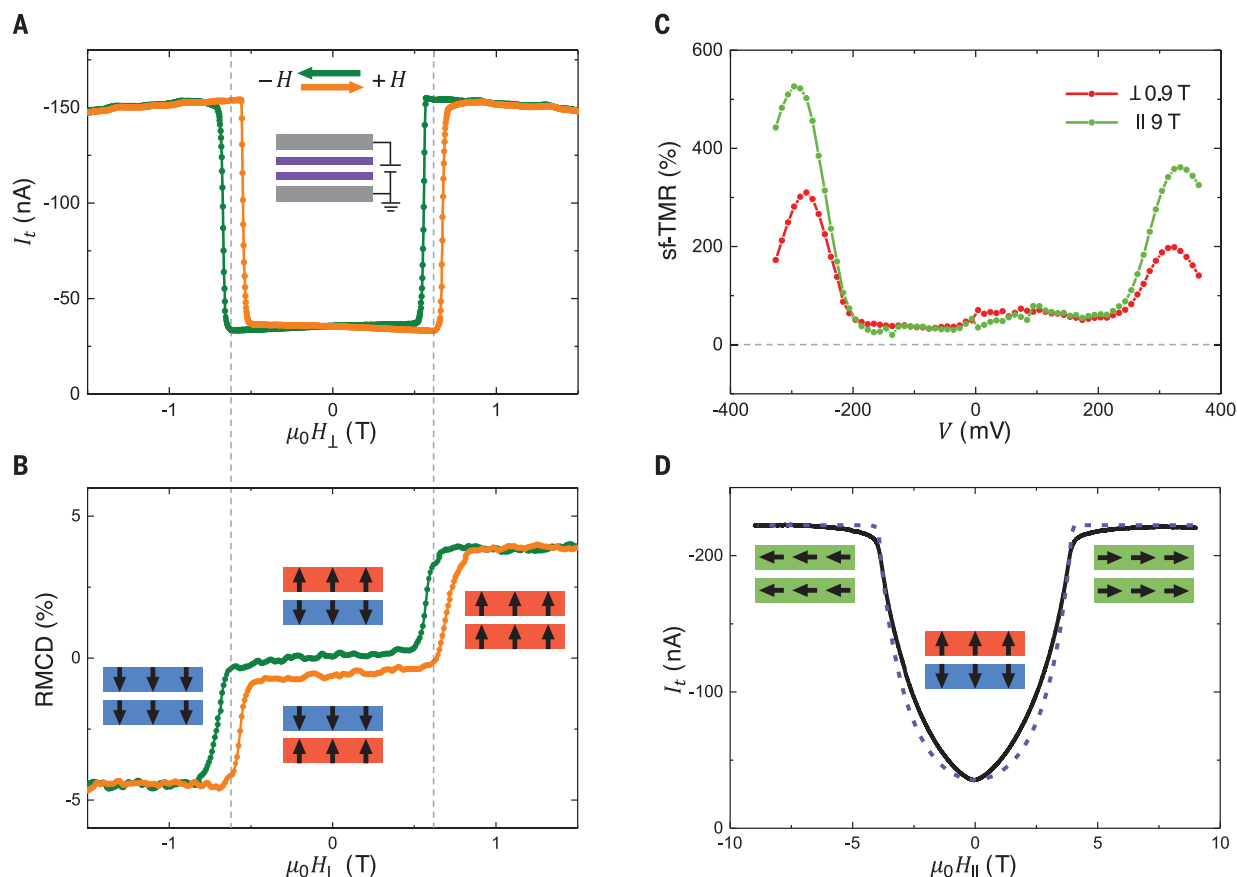
vestigated devices with bilayer, trilayer, and four-layer  $\text{CrI}_3$ . All measurements were carried out at a temperature of 2 K, unless otherwise specified.

We begin with the case of bilayer  $\text{CrI}_3$ . The inset of Fig. 1C is an optical micrograph of a device with the structure illustrated in Fig. 1B, obtained by stacking exfoliated 2D materials using a dry-transfer process in a glovebox (27). The tunneling junction area is less than  $\sim 1 \mu\text{m}^2$  to avoid effects caused by lateral magnetic domain structures (3, 4). Figure 1C shows the tunneling current ( $I_t$ ) as a function of DC bias voltage ( $V$ ) at selected magnetic fields ( $\mu_0 H$ ) (27). Unlike in tunneling devices using nonmagnetic hBN as the barrier (22, 23), it has a strong magnetic field dependence. As shown in Fig. 1C,  $I_t$  is much smaller at  $\mu_0 H = 0$  T (purple trace) than it is in the presence of an out-of-plane field ( $\mu_0 H_\perp$ , red trace) or an in-plane field ( $\mu_0 H_\parallel$ , green trace). This magnetic field-dependent tunneling current implies a spin-dependent tunneling probability related to the field-dependent magnetic structure of bilayer  $\text{CrI}_3$ .

To investigate the connection between the bilayer  $\text{CrI}_3$  magnetic states and the magnetoresistance, we measured  $I_t$  as a function of  $\mu_0 H_\perp$

at a particular bias voltage ( $\sim -290$  mV). The green and orange curves in Fig. 2A correspond to decreasing and increasing magnetic fields, respectively.  $I_t$  exhibits plateaus with two values, about  $-36$  nA and  $-155$  nA. The lower plateau is seen at low fields, and there is a sharp jump to the higher plateau when the magnetic field exceeds a critical value. We also employed reflective magnetic circular dichroism (RMCD) to probe the out-of-plane magnetization of the bilayer  $\text{CrI}_3$  near the tunneling area. Figure 2B shows the RMCD signal as a function of  $\mu_0 H_\perp$  under similar experimental conditions to the magnetoresistance measurements (27). The signal is small at low fields, corresponding to a layered-antiferromagnetic ground state ( $\uparrow\downarrow$  or  $\downarrow\uparrow$ ), where the arrows indicate the out-of-plane magnetizations in the top and bottom layers, respectively. As the magnitude of the field increases, there is a step up to a larger signal corresponding to the fully spin-polarized states ( $\uparrow\uparrow$  and  $\downarrow\downarrow$ ), consistent with earlier MOKE measurements on bilayer  $\text{CrI}_3$  (3). Additional bilayer device measurements can be found in (27).

A direct comparison of  $I_t$  and RMCD measurements provides the following explanation of the



**Fig. 2. Double spin-filter MTJ from bilayer  $\text{CrI}_3$ .** (A) Tunneling current as a function of out-of-plane magnetic field ( $\mu_0 H_\perp$ ) at a selected bias voltage ( $\sim -290$  mV). Green (orange) curve corresponds to decreasing (increasing) magnetic field. The junction area is about  $0.75 \mu\text{m}^2$ . (B) Reflective magnetic circular dichroism (RMCD) of the same device

at zero bias. Insets show the corresponding magnetic states. (C) Extracted sf-TMR ratio as a function of bias based on the  $I_t$ - $V$  curves in Fig. 1C. (D) Tunneling current as a function of in-plane magnetic field ( $\mu_0 H_\parallel$ ) (black) at a selected bias voltage ( $\sim -290$  mV) with simulations (dashed purple). Insets show the corresponding magnetic states.

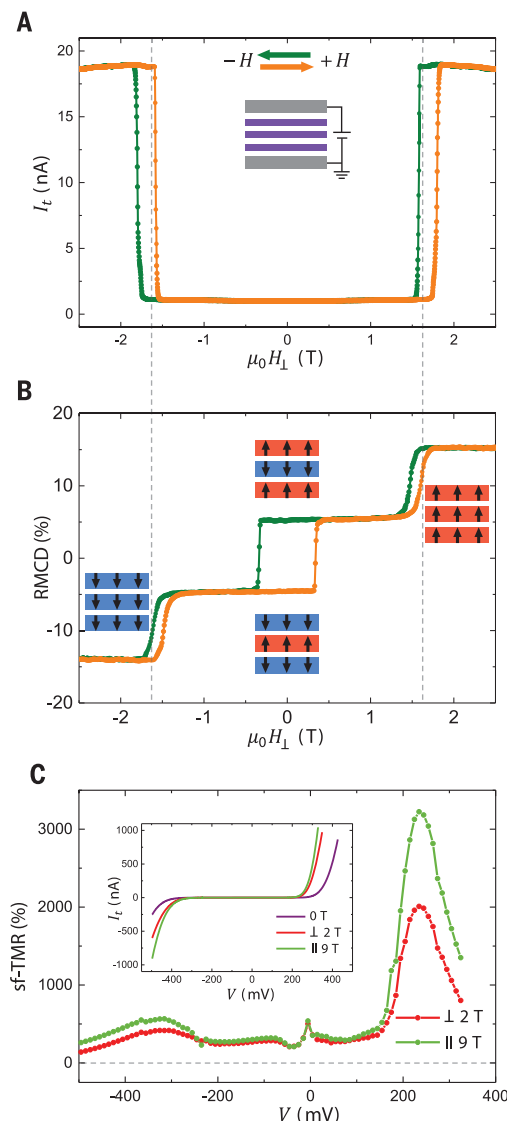
giant sf-TMR: In the  $\uparrow\downarrow$  or  $\downarrow\uparrow$  states at low field, the current is small because spin-conserving tunneling of an electron through the two layers in sequence is suppressed. The step in  $I_t$  occurs when the magnetic field drives the bilayer into the  $\uparrow\uparrow$  and  $\downarrow\downarrow$  states and this suppression is removed. This is known as the double spin-filtering effect (11, 12), and it can be modeled by treating the two monolayers as tunnel-coupled spin-dependent quantum wells (27).

We quantify the sf-TMR by  $(R_{\text{ap}} - R_{\text{p}})/R_{\text{p}}$ , where  $R_{\text{ap}}$  and  $R_{\text{p}}$  are the DC resistances with antiparallel and parallel spin alignment in bilayer  $\text{CrI}_3$ , respectively, measured at a given bias. Figure 2C shows the value of this quantity as a function of bias extracted from the  $I_t$ - $V$  curves in Fig. 1C. The highest sf-TMR achieved is 310% for magnetization fully aligned perpendicular to the plane and 530% for parallel alignment. The sf-TMR decreases as temperature increases and vanishes above the critical temperature at about 45 K (27).

The fact that the sf-TMR for in-plane magnetization is larger than for out-of-plane implies anisotropic magnetoresistance, which is a common feature in ferromagnets (28) and is a sign of anisotropic spin-orbit coupling stemming from the layered structure of  $\text{CrI}_3$ . The sf-TMR is also peaked at a certain bias and asymmetric between positive and negative bias. These observations are similar to the reported double sf-MTJs based on EuS thin films, where the asymmetry is caused by the different thickness and coercive fields of the two EuS spin filters (12). Likewise, our data imply that the device lacks up-down symmetry, possibly because the few-layer graphene contacts are not identical in thickness. This broken symmetry also manifests as tilting of the current plateaus (Fig. 2A) and the finite nonzero RMCD value (Fig. 2B) in the layered-antiferromagnetic states (27).

To further investigate magnetic anisotropy and the assignment of magnetic states in the bilayer, we measured  $I_t$  as a function of in-plane magnetic field. As shown in Fig. 2D (black curve),  $I_t$  is smallest at zero field, in the layered-antiferromagnetic state, and smoothly increases with the magnitude of the field. This behavior has a natural interpretation in terms of a spin-canting effect. Once the magnitude of  $\mu_0 H_{\parallel}$  exceeds about 4 T, the spins are completely aligned with the in-plane field and  $I_t$  saturates. Simulations of the canting effect to match the data (dashed purple curve) yield a magnetic anisotropy field of 3.8 T (27), much larger than the out-of-plane critical magnetic field of  $\pm 0.6$  T seen in Fig. 2A. These results therefore both demonstrate and quantify a large out-of-plane magnetic anisotropy in bilayer  $\text{CrI}_3$ .

We next consider the trilayer case. Figure 3, A and B, shows  $I_t$  and RMCD, respectively, for a trilayer  $\text{CrI}_3$  sf-MTJ as a function of out-of-plane field. There are four plateaus in the RMCD signal, at -14%, -5%, 5%, and 15%, the ratio between which is close to -3:-1:1:3. By analogy with the analysis of the  $\uparrow\downarrow$  and  $\downarrow\uparrow$  layered-



**Fig. 3. Giant sf-TMR of a trilayer  $\text{CrI}_3$  sf-MTJ.**

(A) Tunneling current as a function of out-of-plane magnetic field ( $\mu_0 H_{\perp}$ ) at a selected bias voltage (235 mV). The junction area is about  $0.06 \mu\text{m}^2$ . (B) RMCD of the same device at zero bias showing antiferromagnetic interlayer coupling. Insets show the corresponding magnetic states. (C) sf-TMR ratio calculated from the  $I_t$ - $V$  data shown in the inset.

antiferromagnetic states in the bilayer, we identify the trilayer ground state as  $\uparrow\uparrow\downarrow$  or  $\downarrow\downarrow\uparrow$  at zero field. We conclude that the interlayer coupling in trilayer  $\text{CrI}_3$  is also antiferromagnetic, and the net magnetization in the ground state—and thus the RMCD value—is 1/3 of the saturated magnetization when the applied field fully aligns the three layers (27). The jumps in  $I_t$  and RMCD in Fig. 3, A and B, are caused by the magnetization of an individual layer flipping, similar to what is seen in metallic layered antiferromagnets (29–32).

We deduce that the low current plateau at small fields in Fig. 3A occurs because the two layered-antiferromagnetic states ( $\uparrow\uparrow\downarrow$  and  $\downarrow\downarrow\uparrow$ ) of the trilayer function as three oppositely po-

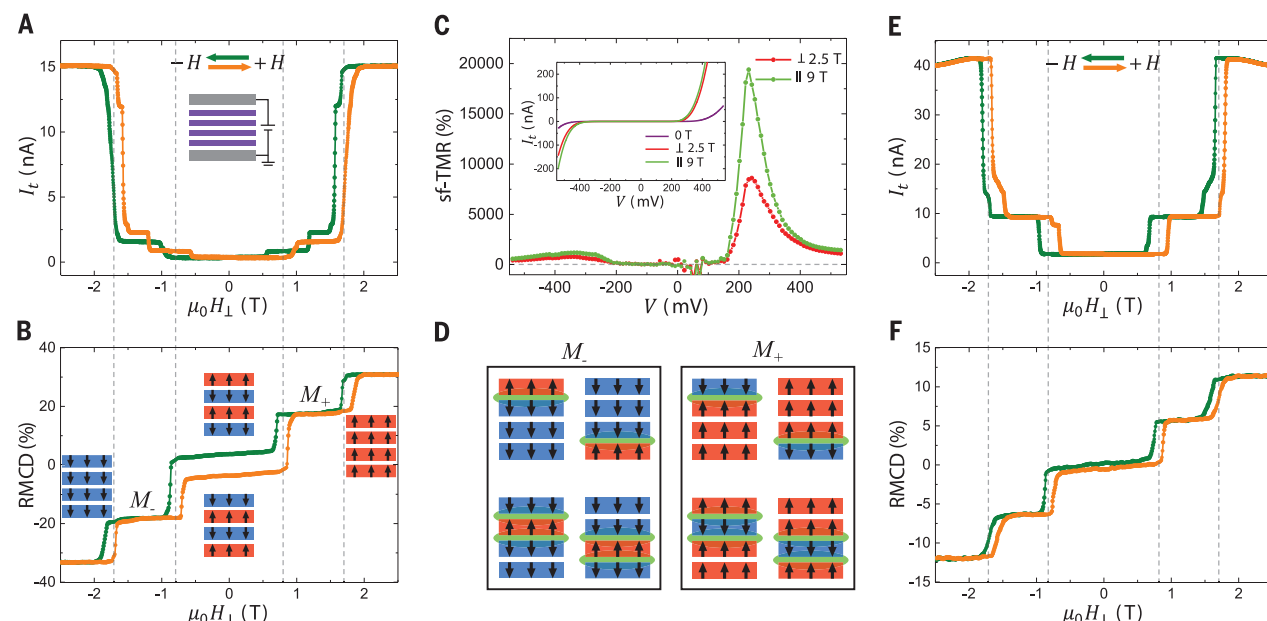
larized spin filters in series. Large enough fields drive the trilayer into fully spin-polarized states, which enhances tunneling and gives the high current plateaus. Figure 3C shows the sf-TMR as a function of bias derived from the  $I_t$ - $V$  curves shown in the inset. The peak values are about 2000% and 3200% for magnetization fully aligned perpendicular and parallel to the plane, respectively, revealing a drastically enhanced sf-TMR compared with bilayer devices.

Increasing the  $\text{CrI}_3$  thickness beyond three layers unlocks more complicated magnetic configurations. Figure 4, A and B, shows  $I_t$  and RMCD, respectively, for a four-layer device. There are multiple plateaus in each, signifying several magnetic configurations with different effects on the tunneling resistance. The small RMCD signal at low fields, below  $\sim 0.8$  T, corresponds to the fully antiferromagnetic ground state, either  $\uparrow\downarrow\downarrow\uparrow$  or  $\downarrow\uparrow\uparrow\downarrow$ . The fact that the RMCD is not zero (Fig. 4B) can be attributed to the asymmetry of the layers caused by the fabrication process, as in the bilayer case above (27). As expected, these fully antiferromagnetic states are very effective at suppressing the tunneling current because they act as four oppositely polarized spin filters in series, explaining the very low current plateau at small fields in Fig. 4A. Applying a large enough field fully aligns the magnetizations of all the layers ( $\uparrow\uparrow\uparrow\uparrow$  or  $\downarrow\downarrow\downarrow\downarrow$ ), producing the highest plateaus in both  $I_t$  and RMCD. Figure 4C shows the sf-TMR as a function of bias extracted from the  $I_t$ - $V$  curves in the inset. The peak values are now about 8600% and 19,000% for perpendicular and parallel field, respectively, representing a further enhancement of the sf-TMR compared to bilayer and trilayer cases.

The RMCD of four-layer  $\text{CrI}_3$  also shows intermediate plateaus at about half the values in the fully aligned states (Fig. 4, B and F, and fig. S10), corresponding to magnetic states with half the net magnetization of the fully aligned states. There are four possible magnetic states for the positive field plateau:  $M_{\perp}\{\uparrow\uparrow\downarrow, \uparrow\uparrow\downarrow, \downarrow\uparrow\uparrow, \uparrow\downarrow\uparrow\}$ , the four time-reversal copies ( $M_{\perp}$  being the negative field counterparts (Fig. 4D). The resulting spin-filter configuration should then correspond to one layer polarized opposite to the other three.

Remarkably, in the range of fields where these 1:3 configurations occur, the tunneling current displays multiple plateaus. The green curve in Fig. 4A shows three distinct intermediate  $I_t$  plateaus. Two are in the positive field corresponding to the same intermediate +18% RMCD plateau, and one is in the negative field range. The orange curve, sweeping in the opposite direction, is the time-reversal copy of the green one. The possibility of lateral domains with different net magnetizations being the cause of these extra plateaus is inconsistent with field-dependent RMCD maps of all three measured four-layer  $\text{CrI}_3$  sf-MTJs, none of which showed appreciable domains (27). In addition, the tunnel junction





**Fig. 4. Four-layer CrI<sub>3</sub> sf-MTJs with extraordinarily large sf-TMR and multiple resistance states.** (A) Tunneling current as a function of out-of-plane magnetic field ( $\mu_0 H_{\perp}$ ) at a selected bias voltage (300 mV) and (B) the corresponding RMCD of the same device at zero bias. Insets show the corresponding magnetic states. The junction area is about  $2.2 \mu\text{m}^2$ .

(C) Calculated sf-TMR ratio as a function of bias based on the  $I_t$ - $V$  curves in the inset. (D) Schematic of possible magnetic states corresponding to the intermediate plateaus in (A) and (B). Green lines show the current-blocking interfaces. (E) and (F) Tunneling current and RMCD from another four-layer CrI<sub>3</sub> sf-MTJ. The junction area is about  $1.3 \mu\text{m}^2$ .

area is quite small compared with the typical domain size of a few microns in CrI<sub>3</sub> (3, 4).

Instead, these current plateaus probably originate from distinct magnetic states. Whereas the four states in  $M_+$  are indistinguishable in RMCD because of the same net magnetization, the tunneling current is likely to be sensitive to the position of the one layer with minority magnetization. First, the  $\uparrow\uparrow\uparrow$  and  $\uparrow\uparrow\downarrow$  have only one current-blocking interface, whereas  $\uparrow\downarrow\uparrow$  and  $\uparrow\downarrow\downarrow$  have two (green lines between adjacent layers with opposite magnetizations shown in Fig. 4D). Second, the current flow direction as well as the possibly asymmetric few-layer graphene contacts may introduce distinct sf-TMR either between the  $\uparrow\downarrow\uparrow$  and  $\uparrow\downarrow\downarrow$  states or between the  $\downarrow\uparrow\uparrow$  and  $\uparrow\downarrow\downarrow$  states (27). This asymmetry may also help to stabilize  $\downarrow\uparrow\uparrow$  and  $\uparrow\downarrow\downarrow$ , which in general have higher energy than  $\uparrow\downarrow\uparrow$  and  $\uparrow\downarrow\downarrow$  in fully symmetric four-layer CrI<sub>3</sub>. However, to identify the specific magnetic states corresponding to the current plateaus will require a means to distinguish the magnetization of individual layers (27).

The four-layer CrI<sub>3</sub> sf-MTJ points to the potential for using layered antiferromagnets for engineering multiple magnetoresistance states in an individual sf-MTJ. Figure 4, E and F, and fig. S10 show  $I_t$  and RMCD for two other four-layer CrI<sub>3</sub> sf-MTJs. They exhibit one or two intermediate plateaus, rather than the three observed in Fig. 4A. The sample dependence suggests that these intermediate states are sensitive to the environment of the CrI<sub>3</sub>, such as the details of the contacts, implying potential tunability—for example, by electro-

statically doping the graphene contacts. One exciting future direction could be to seek electrically controlled switching between several different magnetoresistance states. Already the sf-TMR of up to 19,000% observed in four-layer devices is an order of magnitude larger than that of MgO-based conventional MTJs (19–21) and several orders of magnitude larger than achieved with existing sf-MTJs under similar experimental conditions (12). Although the demonstrated vdW sf-MTJs only work at low temperatures, these results highlight the potential of 2D magnets and their heterostructures for engineering novel spintronic devices with unrivaled performance (33, 34).

#### REFERENCES AND NOTES

1. A. K. Geim, I. V. Grigorieva, *Nature* **499**, 419–425 (2013).
2. K. S. Novoselov, A. Mishchenko, A. Carvalho, A. H. Castro Neto, *Science* **353**, aac9439 (2016).
3. B. Huang et al., *Nature* **546**, 270–273 (2017).
4. D. Zhong et al., *Sci. Adv.* **3**, e1603113 (2017).
5. C. Gong et al., *Nature* **546**, 265–269 (2017).
6. M. A. McGuire, H. Dixit, V. R. Cooper, B. C. Sales, *Chem. Mater.* **27**, 612–620 (2015).
7. M.-W. Lin et al., *J. Mater. Chem. C* **4**, 315–322 (2016).
8. Y. Tian, M. J. Gray, H. Ji, R. J. Cava, K. S. Burch, *2D Mater.* **3**, 025035 (2016).
9. X. Wang et al., *2D Mater.* **3**, 031009 (2016).
10. J.-U. Lee et al., *Nano Lett.* **16**, 7433–7438 (2016).
11. D. C. Worledge, T. H. Geballe, *J. Appl. Phys.* **88**, 5277–5279 (2000).
12. G.-X. Miao, M. Müller, J. S. Moodera, *Phys. Rev. Lett.* **102**, 076601 (2009).
13. M. N. Baibich et al., *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
14. G. Binashch, P. Grünberg, F. Saurenbach, W. Zinn, *Phys. Rev. B* **39**, 4828–4830 (1989).
15. B. Dieny et al., *Phys. Rev. B* **43**, 1297–1300 (1991).
16. M. Julliere, *Phys. Lett. A* **54**, 225–226 (1975).
17. J. S. Moodera, L. R. Kinder, T. M. Wong, R. Meservey, *Phys. Rev. Lett.* **74**, 3273–3276 (1995).

18. T. Miyazaki, N. Tezuka, *J. Magn. Magn. Mater.* **139**, L231–L234 (1995).
19. S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, K. Ando, *Nat. Mater.* **3**, 868–871 (2004).
20. S. S. P. Parkin et al., *Nat. Mater.* **3**, 862–867 (2004).
21. S. Ikeda et al., *Appl. Phys. Lett.* **93**, 082508 (2008).
22. L. Britnell et al., *Science* **335**, 947–950 (2012).
23. G.-H. Lee et al., *Appl. Phys. Lett.* **99**, 243114 (2011).
24. T. Georgiou et al., *Nat. Nanotechnol.* **8**, 100–103 (2013).
25. A. Soumyanarayanan, N. Reyren, A. Fert, C. Panagopoulos, *Nature* **539**, 509–517 (2016).
26. S. A. Wolf et al., *Science* **294**, 1488–1495 (2001).
27. See the supplementary materials.
28. T. R. McGuire, R. I. Potter, *IEEE Trans. Magn.* **11**, 1018–1038 (1975).
29. O. Hellwig, T. L. Kirk, J. B. Kortright, A. Berger, E. E. Fullerton, *Nat. Mater.* **2**, 112–116 (2003).
30. O. Hellwig, A. Berger, E. E. Fullerton, *Phys. Rev. Lett.* **91**, 197203 (2003).
31. B. Chen et al., *Science* **357**, 191–194 (2017).
32. M. Charilaou, C. Borden, F. Hellman, *Appl. Phys. Lett.* **104**, 212405 (2014).
33. D. C. Ralph, M. D. Stiles, *J. Magn. Magn. Mater.* **320**, 1190–1216 (2008).
34. D. MacNeill et al., *Nat. Phys.* **13**, 300–305 (2017).

#### ACKNOWLEDGMENTS

We thank Y. Cui, G. Miao, and S. Majetich for insightful discussion. **Funding:** Work at the University of Washington was mainly supported by the Department of Energy, Basic Energy Sciences, Materials Sciences and Engineering Division (DE-SC0018171). Device fabrication and part of transport measurements are supported by NSF-DMR-1708419, NSF MRSEC 1719797, and UW Innovation Award. D.H.C.'s contribution is supported by DE-SC0002197. Work at CMU is supported by DOE BES DE-SC0012509. Work at HKU is supported by the Croucher Foundation (Croucher Innovation Award), UGC of HKSAR (AoE/P-04/08), and the HKU ORA. Work at ORNL (M.A.M.) was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division. K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by the MEXT, Japan, and JSPS KAKENHI grant number JP15K21722. D.X. acknowledges the support of a Cottrell Scholar Award. X.X.

acknowledges the support from the State of Washington–funded Clean Energy Institute and from the Boeing Distinguished Professorship in Physics. **Author contributions:** W.Y. and X.X. conceived the project. T.S. and X.C. fabricated the devices and performed the experiments, assisted by B.H., N.P.W., K.L.S., and L.Z., supervised by X.X., W.Y., D.X., and D.H.C. T.S. and X.X. analyzed the data, with theory support from M.W.-Y.T., W.Y., X.Z., and D.X. M.A.M. provided and characterized bulk CrI<sub>3</sub> crystals. T.T. and K.W. provided and characterized bulk hBN crystals. T.S., X.X., W.Y., D.X., and D.H.C.

wrote the manuscript, with input from all authors. **Competing interests:** A provisional patent based on the content of this paper has been filed on behalf of the authors. **Data availability:** All data files are available from Harvard Dataverse at <https://doi.org/10.7910/DVN/TOTIWA>.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/1214/suppl/DC1](http://www.sciencemag.org/content/360/6394/1214/suppl/DC1)  
Materials and Methods

Supplementary Text  
Figs. S1 to S11  
Table S1  
References (35–39)

20 November 2017; accepted 24 April 2018  
Published online 3 May 2018  
[10.1126/science.aar4851](https://doi.org/10.1126/science.aar4851)



## MAGNETISM

# Probing magnetism in 2D van der Waals crystalline insulators via electron tunneling

D. R. Klein<sup>1\*</sup>, D. MacNeill<sup>1\*</sup>, J. L. Lado<sup>2,3</sup>, D. Soriano<sup>2</sup>, E. Navarro-Moratalla<sup>4</sup>, K. Watanabe<sup>5</sup>, T. Taniguchi<sup>5</sup>, S. Manni<sup>6,7,8</sup>, P. Canfield<sup>6,7</sup>, J. Fernández-Rossier<sup>2</sup>, P. Jarillo-Herrero<sup>1†</sup>

Magnetic insulators are a key resource for next-generation spintronic and topological devices. The family of layered metal halides promises varied magnetic states, including ultrathin insulating multiferroics, spin liquids, and ferromagnets, but device-oriented characterization methods are needed to unlock their potential. Here, we report tunneling through the layered magnetic insulator CrI<sub>3</sub> as a function of temperature and applied magnetic field. We electrically detect the magnetic ground state and interlayer coupling and observe a field-induced metamagnetic transition. The metamagnetic transition results in magnetoresistances of 95, 300, and 550% for bilayer, trilayer, and tetralayer CrI<sub>3</sub> barriers, respectively. We further measure inelastic tunneling spectra for our junctions, unveiling a rich spectrum consistent with collective magnetic excitations (magnons) in CrI<sub>3</sub>.

**V**an der Waals magnetic insulators are a materials system that may enable designer topological states (1) and spintronic technologies (2). The recent isolation (3, 4) of few-layer magnets with either ferromagnetic (CrI<sub>3</sub>, Cr<sub>2</sub>Ge<sub>2</sub>Te<sub>6</sub>) or antiferromagnetic order (5, 6) is just the tip of the iceberg. The vast family of layered metal halides (7) contains spin orders from multiferroics (8) to proximate spin liquids (9), of key interest to both fundamental and applied physics. Existing studies have focused on magneto-optical effects (3, 4, 10, 11) as a characterization tool, but a more general, device-oriented, approach is needed.

Here we demonstrate that tunneling through layered insulators is a versatile probe of magnetism on the nanoscale in these materials. We report the conductance of graphite/CrI<sub>3</sub>/graphite junctions (Fig. 1A) as a function of magnetic field and temperature and electrically detect an antiferromagnetic ground state and a field-induced metamagnetic transition. The metamagnetic transition is revealed by large magnetoresistances (up to 550%) arising from the antiparallel-to-parallel reorientation of chromium spins in adjacent crystal layers. A similar effect was previously proposed (12) for synthetic multilayer magnets, but experimental realizations (13) were limited to mag-

netoresistances below 70%. The performance of our devices is an order of magnitude higher, corresponding to estimated spin polarization above 95%. Furthermore, the two-dimensional magnetism of CrI<sub>3</sub> enables ultrathin tunnel barriers (<3 nm) and a concomitant 10,000-fold increase in conductance (per unit area) compared to previous results (13). The noninvasive van der Waals transfer of the magnetic layer ensures substrate-independent device integration, and together with high magnetoresistance, spin polarization, and conductance, may enable noninvasive spin injectors and detectors for next-generation spintronics experiments incorporating topological insulators (14), superconductors (15), antiferromagnets (16), and low-symmetry crystals (17–20).

Tunneling through magnetic insulators was first studied in the pioneering experiments of (21) and later in (22, 23). When electrons tunnel through a ferromagnetic insulator, spin-up and spin-down electrons see different barrier heights (Fig. 1B). As a result, the tunneling rate can vary by orders of magnitude for electrons of opposite spins (12, 22), called the spin filter effect. The smaller gap for spin-up electrons tends to decrease the junction resistance as the barrier is cooled below its Curie temperature, *T*<sub>C</sub>. The situation is more complicated for spatially textured magnetism. For example, the resistance of Ag/EuSe/Al tunnel junctions increases significantly when the EuSe becomes antiferromagnetic (23). However, the exponential dependence of the tunneling current on the barrier electronic structure generally provides a clear resistive signature of magnetism. We will use these effects to electrically detect the magnetic ground state and field-induced metamagnetic transition of few-layer CrI<sub>3</sub>.

The resistance of a graphite/tetralayer CrI<sub>3</sub>/graphite junction (device D1) as a function of

temperature is shown in Fig. 1C. We measure the resistance in a four-point geometry using a 30-mV root-mean-square AC excitation (24). The temperature dependence was measured by cooling the sample down with (purple line) and without (black line) the application of an external magnetic field. The magnetic field is applied perpendicular to the layers, along the magnetic easy axis of CrI<sub>3</sub>. Above 90 K, the resistance is independent of the applied field and shows Arrhenius behavior with a thermal activation gap of roughly 159 meV (fig. S1). The resistance becomes field-dependent as the temperature approaches the bulk *T*<sub>C</sub> of 61 K. When the sample is cooled in a 2.5-T magnetic field, the resistance plateaus below 80 K, signaling the onset of tunneling conductance (21, 23). By contrast, when the sample is cooled without an external field, the resistance exhibits a kink near *T*<sub>C</sub> and continues to increase below 60 K. The dependence of the tunneling resistance on magnetic field and temperature shows that the tunnel conductance is sensitive to the magnetization of the barrier.

To further investigate the magnetic phase diagram, we study the zero-bias conductance (500-μV AC excitation) of devices with two- to four-layer CrI<sub>3</sub> barriers as a function of applied magnetic field at low temperatures (300 mK to 4.2 K). We start with an analysis of a graphite/bilayer CrI<sub>3</sub>/graphite junction (D2, Fig. 2A). For this device, the junction conductance increases almost twofold in a sharp step as the external field is increased above 0.85 T. The corresponding magnetoresistance is 95%, defined as

$$MR = 100\% \times \frac{(G_{\text{HI}} - G_{\text{LO}})}{G_{\text{LO}}} \quad (1)$$

where *G*<sub>HI</sub> is the high-field conductance maximum and *G*<sub>LO</sub> is the low-field conductance minimum. No further steps are observed up to the largest fields studied (8 T; see fig. S2). As the field is reduced from 2.4 T, the conductance decreases to its original zero-field value in a sharp step at 0.35 T. The well-defined steps and hysteretic field dependence demonstrate that the conductance changes originate from switching events of the magnetization. The tunneling current is most sensitive to the interlayer magnetization alignment, so the large steps we observe likely arise from vertical domains, i.e., regions where the magnetization points in different directions in different layers of CrI<sub>3</sub>.

Recently, magneto-optical Kerr effect (MOKE) data have revealed an antiferromagnetic state in bilayer CrI<sub>3</sub> for fields below about 0.6 T (4). In this state, the Cr moments order ferromagnetically within each layer but point in opposite directions in adjacent layers (Fig. 2B). The layers are fully aligned when the external magnetic field is increased above a critical value (Fig. 2C); i.e., it undergoes a metamagnetic transition to a ferromagnetic state. When the field is reduced, the magnetization spontaneously reverts to the antiparallel configuration. The switching behavior we observe in magnetoconductance reflects these previous MOKE data, confirming that the conductance change arises from the metamagnetic

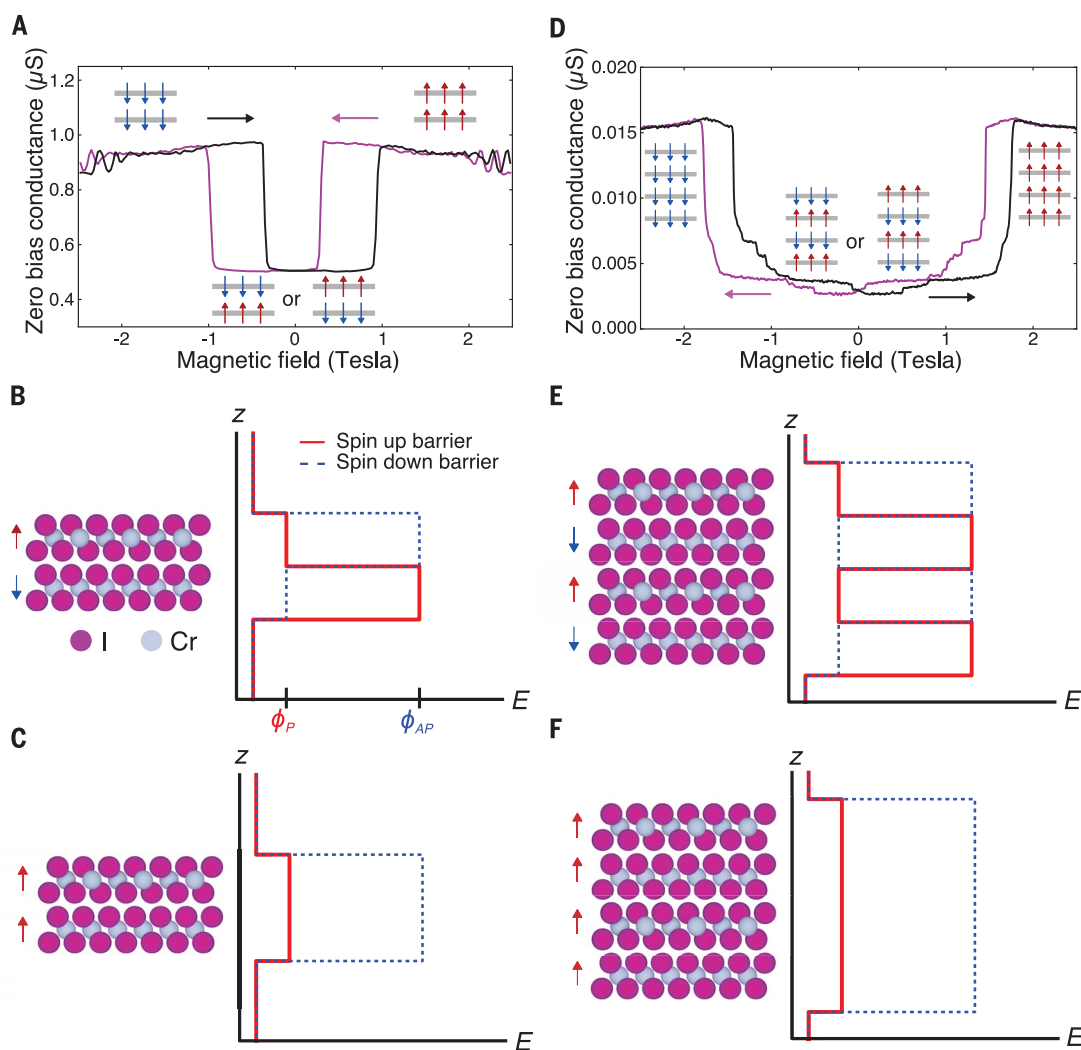
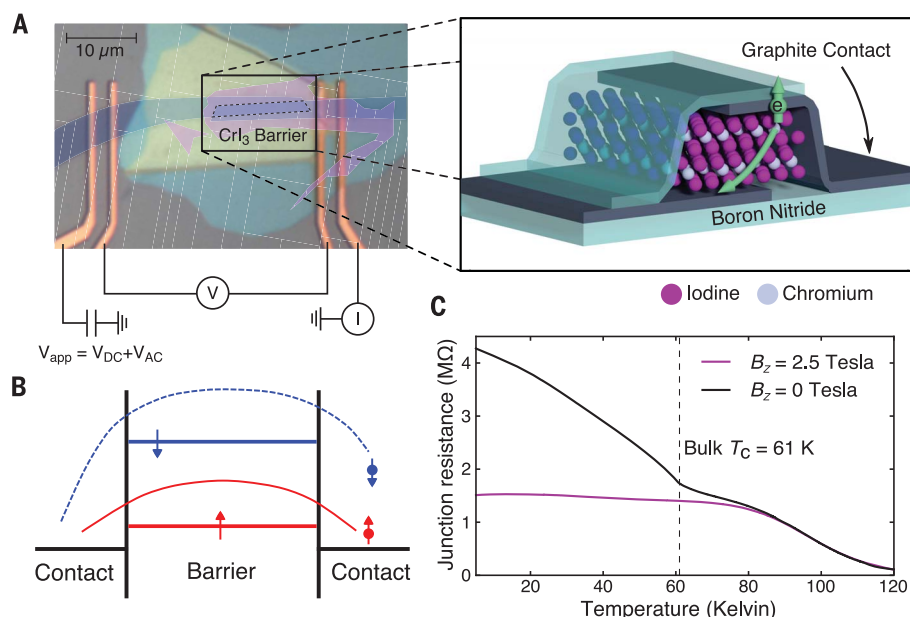
<sup>1</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>2</sup>Quantalab, International Iberian Nanotechnology Laboratory, 4715-310 Braga, Portugal.

<sup>3</sup>Institute for Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland. <sup>4</sup>Instituto de Ciencia Molecular, Universidad de Valencia, 46100 Burjassot, Spain. <sup>5</sup>National Institute for Materials Science, Tsukuba, Japan. <sup>6</sup>Ames Laboratory, U.S. Department of Energy, Iowa State University, Ames, IA 50011, USA.

<sup>7</sup>Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA. <sup>8</sup>Department of Condensed Matter Physics and Materials Science, Tata Institute of Fundamental Research, Mumbai 400005, India.

\*These authors contributed equally to this work.  
†Corresponding author. Email: pjarillo@mit.edu

**Fig. 1. Experimental setup.** (A) Optical micrograph of a tetralayer  $\text{CrI}_3$  tunnel junction device (device D1, false color). The dashed line encloses the tunnel junction area. The graphite contacts are themselves contacted by Au/Cr wires in a four-point geometry. Inset: Schematic of the van der Waals heterostructures studied in this work. Electrons tunnel between two graphite sheets separated by a magnetic  $\text{CrI}_3$  tunnel barrier. The entire stack is encapsulated in hexagonal boron nitride. (B) Schematic energy diagram of a metal/ferromagnetic insulator/metal junction. The red and blue lines in the barrier region represent the spin-up and spin-down energy barriers, respectively. The lower barrier for spin-up electrons leads to spin-polarized tunneling and reduced resistance for a ferromagnetic barrier. (C) Zero-bias junction resistance versus temperature for device D1 cooled with (purple) and without (black) an applied magnetic field. The curves begin to deviate around the bulk Curie temperature (61 K), giving evidence for magnetic order in the  $\text{CrI}_3$  barrier and for spin-polarized tunneling. The magnetic field was applied perpendicular to the  $\text{CrI}_3$  layers.



**Fig. 2. Magnetoconductance of few-layer  $\text{CrI}_3$ .**

(A) Conductance through a bilayer  $\text{CrI}_3$  tunnel barrier (device D2) as a function of an out-of-plane applied magnetic field with 500- $\mu\text{V}$  AC excitation. The data were taken both for decreasing (purple line, left arrow) and increasing (black line, right arrow) magnetic field. The magnetoconductance traces are consistent with previous magnetometry data (4) for bilayer  $\text{CrI}_3$  showing that the two layers are antiparallel at zero field but can be aligned with an external field below 1 T. (B and C) Schematic of barriers experienced by spin-up and spin-down electrons tunneling through bilayer  $\text{CrI}_3$  in the low-field (B) and high-field (C) states. In the low-field state, the two layers are antiparallel, and both spins see a high barrier. In the high-field state, the layers are aligned and up spins see a low-energy barrier, leading to increased conductance. (D to F) Analogous data and schematics for a tetralayer  $\text{CrI}_3$  barrier device (device D3). In both cases, the sample temperature was 300 mK.



transition. We note, however, that little to no hysteresis was previously observed in the MOKE results for bilayer  $\text{CrI}_3$  (4), whereas we observe a clear hysteresis in the tunneling measurements. The reasons for this may be related to the lower temperature for these tunneling experiments as well as the fact that the  $\text{CrI}_3$  remains closer to equilibrium (no photoexcitation).

We have also studied tunnel junctions with three- and four-layer  $\text{CrI}_3$  as the barrier. The zero-bias junction resistance of a graphite/4L  $\text{CrI}_3$ /graphite junction (D3) is shown as a function of external magnetic field in Fig. 2D. The overall phenomenology is similar to that of junctions with a bilayer barrier, with well-defined steps and a total magnetoresistance of 550%. In addition to the large jump around 1.8 T, we see multiple smaller steps that may correspond to lateral domains within the junction. This is consistent with a lateral domain size on the order of  $2\ \mu\text{m}$  observed in previous optical studies (2, 4). The behavior of our trilayer junctions is again similar, with magnetoresistances up to

300% (fig. S3). On the basis of these results, we hypothesize that few-layer  $\text{CrI}_3$  is antiferromagnetic without an external magnetic field (Fig. 2E). Such behavior is consistent with magneto-optical data for bilayer  $\text{CrI}_3$  (4), but those MOKE data suggested a ferromagnetic configuration for thicker crystals (e.g., 3L  $\text{CrI}_3$ ). Nevertheless, our data strongly support an antiparallel alignment between layers extending over most of the junction area. Once more, the different temperatures and absence of photoexcitation may be responsible for the different behavior observed.

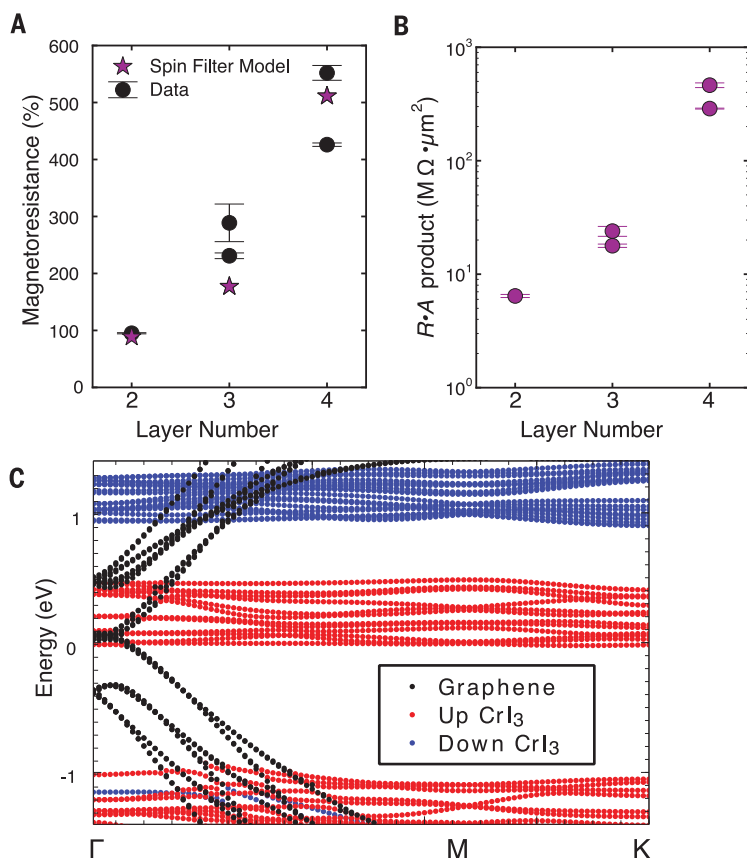
To understand the large magnetoresistance and its thickness dependence, we analyze a spin filter model (12) for transmission through a  $\text{CrI}_3$  barrier. The model treats each crystal layer of the  $\text{CrI}_3$  as an independent tunnel barrier, with a transmission coefficient of  $T_P$  and  $T_{AP}$  for spins parallel and antiparallel to the local spin direction, respectively. Ignoring multiple reflections and quantum interference effects, the transmission through the entire crystal is then a product

of the transmission coefficients for each layer. For example, for a  $\text{CrI}_3$  bilayer in the high-field magnetization configuration (Fig. 2C), spin-up electrons have a transmission probability  $T_P^2$  whereas spin-down electrons have transition probability  $T_{AP}^2$ . The high-field conductance is  $G_{\text{HI}} \propto T_P^2 + T_{AP}^2$ . Similarly, for the low-field configuration with antiparallel magnetizations (Fig. 2B), the conductance is  $G_{\text{LO}} \propto 2T_P T_{AP}$ . The ratio of high-field to low-field conductances is then  $G_{\text{HI}}/G_{\text{LO}} = (T_P^2 + T_{AP}^2)/2T_P T_{AP} \approx T_P/T_{AP}$ . We have carried out similar calculations for  $N = 3$ - and 4-layer  $\text{CrI}_3$  barriers, summarized in the supplementary text. In Fig. 3A, we plot the measured magnetoresistance (black circles) as a function of  $N$ , along with a one-parameter fit to the spin filter model (purple stars). The model reproduces the overall experimental trend with a best-fit value of  $T_P/T_{AP} = 3.5$ . For a summary of all measured devices, see table S1.

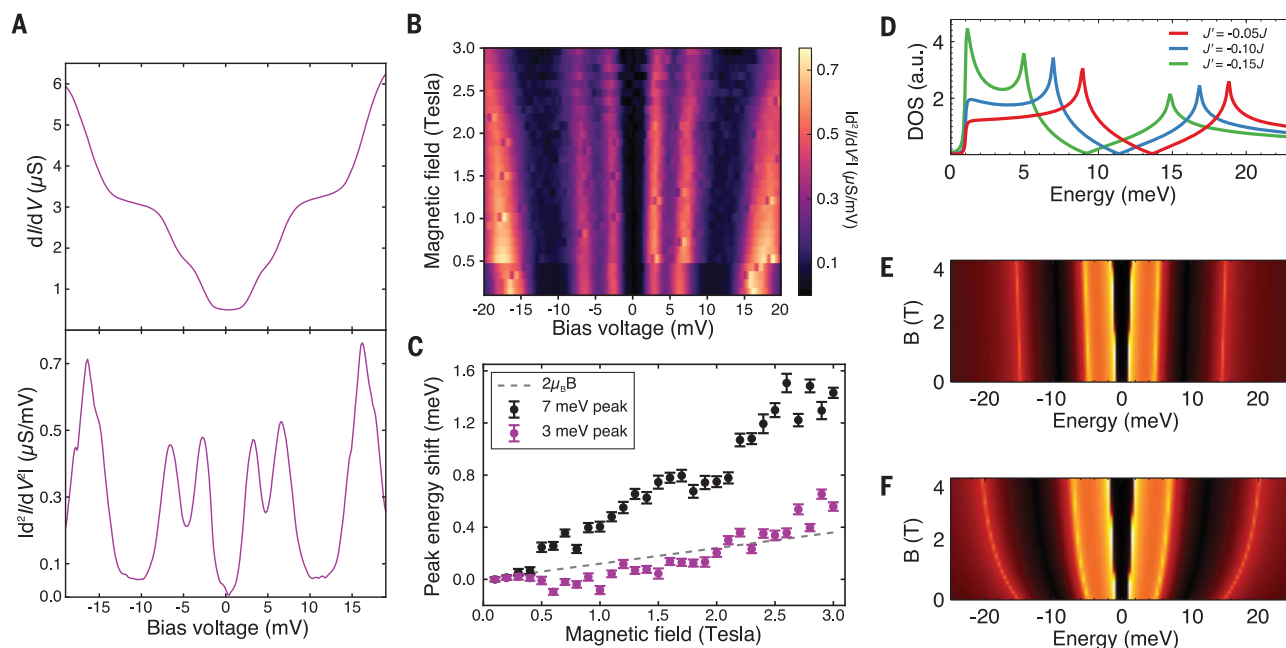
We can also estimate the spin polarization of the current within the spin filter model. When the  $\text{CrI}_3$  is fully polarized, the transmission probability of up and down spins through an  $N$ -layer  $\text{CrI}_3$  barrier is  $T_P^N$  and  $T_{AP}^N$ , respectively. Therefore, the ratio of spin-up to spin-down conductance is approximately  $G_{\uparrow}/G_{\downarrow} = (T_P/T_{AP})^N$ . From  $T_P/T_{AP} \approx 3.5$ , we estimate a spin polarization of  $(G_{\uparrow} - G_{\downarrow})/(G_{\uparrow} + G_{\downarrow}) \approx 85, 95$ , and 99% for  $N = 2, 3$ , and 4, respectively. These values are comparable to the largest values obtained with EuSe and EuS magnetic insulator barriers (13, 23), so that  $\text{CrI}_3$  tunnel barriers can enable future spin-sensitive transport devices.

In the spin filter approximation, the calculation of the magnetoresistance is reduced to a calculation of  $T_P/T_{AP}$ , related to the different barrier heights for spin-up and spin-down electrons. To investigate the barrier heights, we carried out density functional theory (DFT) calculations for three layers of  $\text{CrI}_3$  and three layers of graphite (see the supplementary text). Calculations portray  $\text{CrI}_3$  as a ferromagnetic insulator with magnetic moments localized on the chromium atoms and spin-split energy bands (Fig. 3C). Notably, when the magnetization of the three layers is aligned, we find that spin-up bands of  $\text{CrI}_3$  lie very close to the graphite Fermi energy, whereas the nearest spin-down bands are much higher in energy ( $>1\ \text{eV}$ ). Therefore, the transparency of the barrier has to be much smaller for spin-down electrons and provides a microscopic foundation for the large  $T_P/T_{AP}$ . Note that even though the DFT calculations show a  $\text{CrI}_3$  majority band very close to or crossing the graphite Fermi energy, the exponential thickness dependence of the junction resistance (Fig. 3B) shows that our junctions are in the tunneling-dominated regime with a finite barrier height. Further transport calculations should elucidate the precise tunneling pathways in  $\text{CrI}_3$ /graphite junctions leading to finite energy barriers with chromium 3d orbital bands very close to the Fermi level.

In addition to the zero-bias conductance, we measured the differential conductance  $dI/dV$  as a function of the applied DC offset  $V_{\text{DC}}$ . The



**Fig. 3. Origin of magnetoresistance in  $\text{CrI}_3$ .** (A) Magnetoresistance ratio (black circles) versus  $\text{CrI}_3$  layer number for multiple devices at 300 mK. We also plot a fit to the spin filter model (purple stars). The only fitting parameter,  $T_P/T_{AP} = 3.5$ , gives the ratio of spin-up to spin-down transmission through a  $\text{CrI}_3$  monolayer. (B) Resistance-area product versus  $\text{CrI}_3$  layer number for multiple devices. The resistances are measured in the fully aligned magnetic configuration and were taken at zero bias. (C) Electronic structure of a trilayer graphite/trilayer  $\text{CrI}_3$  heterostructure calculated with density functional theory. The  $\text{CrI}_3$  is in the fully ferromagnetic configuration, and its bands are projected on the spin-up and spin-down channels. Although the minority spins do not show states close to the Fermi energy, there are a large number of states in the majority channel. The difference establishes a microscopic basis for the large  $T_P/T_{AP}$  that we observe.



**Fig. 4. Inelastic tunneling spectroscopy.** (A) Top panel: Differential conductance versus a DC bias voltage for a bilayer  $\text{CrI}_3$  barrier device (D2) at zero applied magnetic field. The AC excitation was 200  $\mu\text{V}$  and the temperature was 300 mK. Bottom panel: Absolute value of  $d^2I/dV^2$  versus a DC bias voltage, obtained via numerical differentiation of the data in the top panel. According to the theory of inelastic tunneling spectroscopy, the peaks in  $d^2I/dV^2$  correspond to phonon or magnon excitations of the barrier or electrodes. (B)  $|d^2I/dV^2|$  (color scale at right) versus applied magnetic field and DC bias voltage. All three inelastic peaks increase in energy as the applied field is increased. (C) Energy of the two lowest-energy inelastic peaks versus applied magnetic field. The zero-field

energy is subtracted from both peaks for clarity. The peak locations were determined by Gaussian fits to the data. The error bars represent estimated standard deviations calculated from the least-squares fitting procedure. The dashed gray line shows the Zeeman energy shift of a  $2\mu_B$  magnetic moment (0.12 meV/T), which roughly matches the evolution of the 3-meV peak. (D) Calculated magnon density of states (DOS) for  $\text{CrI}_3$ . The details of the calculations are described in the supplementary text. (E) Calculated dispersion of magnons with applied magnetic field at zero temperature. (F) Calculated renormalized magnon dispersion with magnetic field at finite temperature ( $T = 0.033J$ , where  $J$  is the nearest-neighbor exchange).

$dI/dV$  versus  $V_{\text{DC}}$  traces reveal a rich spectrum, whose most prominent features are a series of steplike increases, symmetric in bias, below 25 meV (Fig. 4 and figs. S4 and S5). These steps are characteristic of inelastic electron tunneling where electrons lose energy to collective excitations of the barrier or electrodes. When the tunneling energy ( $eV_{\text{DC}}$ ) exceeds the collective excitation energy, the introduction of these additional tunneling pathways results in steps in the  $dI/dV$  versus  $V_{\text{DC}}$  trace. The energies of phonons (25–27) and magnons (28–31) can therefore be measured as peaks (dips) in  $d^2I/dV^2$  versus  $V_{\text{DC}}$  for positive (negative)  $V_{\text{DC}}$ . The bottom panel of Fig. 4A shows  $|d^2I/dV^2|$  obtained by numerical differentiation of the  $dI/dV$  data for a bilayer  $\text{CrI}_3$  tunneling device that we measured (figs. S4 to S6). Past IETS data on graphite/boron nitride/graphite heterostructures in a geometry similar to that of our junctions (26) do not contain any inelastic contributions from graphite phonons below 17 meV. Earlier scanning tunneling studies of graphite surfaces similarly find an onset of prominent graphite inelastic peaks at 16 meV (27). Thus, the inner two peaks must arise from  $\text{CrI}_3$  phonons or

magnons. The inelastic features start forming just below the onset of magnetism (fig. S7), suggesting a magnon excitation origin.

Another signature of magnon-assisted tunneling is the stiffening of the magnon modes as an external magnetic field is applied (29, 30). A single magnon corresponds to a delocalized spin-flip ( $|S_z| = 1$ ) within the  $\text{CrI}_3$  barrier, which carries a magnetic moment of approximately  $2\mu_B$  (where  $\mu_B$  is the Bohr magneton) antiparallel to the external magnetic field. Therefore, magnon IETS peaks should blueshift at 0.12 meV/T by the Zeeman effect. Figure 4B shows  $|d^2I/dV^2|$  as a function of both applied magnetic field and bias voltage. Even to the eye, a strong linear increase of all three IETS peaks is visible. In Fig. 4C, we plot the peak energies (determined by Gaussian fits) of the innermost peaks versus magnetic field. We also plot the expected energy shift  $2\mu_B B$  due to the Zeeman effect (dashed gray line). This line roughly fits the magnetic field dependence of the 3-meV peak, but the 7-meV peak clearly has much higher dispersion corresponding to  $8\mu_B$ . The latter effect might be caused by magnon renormalization effects, as discussed below.

To model the magnon spectrum, we write an effective spin Hamiltonian for  $\text{CrI}_3$  (32) that includes nearest- and next-nearest-neighbor ex-

change, together with an easy axis anisotropy term (see supplementary text for details). Using this model, we find that the calculated magnon density of states (Fig. 4D) can qualitatively reproduce the experimental inelastic spectrum. We used a nearest-neighbor exchange parameter consistent with previous first principles calculations and experiment (32–34) and chose the next-nearest-neighbor value to match our data (see supplementary text). At zero temperature, the magnon energies are still expected to blueshift at 0.12 meV/T in an applied magnetic field (Fig. 4E). However, at finite temperature and  $B = 0$ , thermally excited magnons deplete the magnetization, resulting in an effective reduction of the spin stiffness and a redshift of the magnon spectrum with respect to the case without thermal renormalization of the exchange constants. Application of a magnetic field increases the spin wave gap, decreasing the population of thermal spin waves and increasing the spin stiffness. This renormalizes the effective magnon hopping parameters, leading to a shift of the spin wave spectrum that adds to the Zeeman term and results in a nonlinear field dependence (Fig. 4F and supplementary text).

Our devices are an example of a “double spin filter” where a magnetic tunnel barrier with decoupled magnetic layers is used as a magnetic



memory bit (12). We overcome the limitations of previous double spin filters (13) owing to the unique decoupling of magnetic layers across the atomic-scale van der Waals gap. This decoupling provides electrical readout of the CrI<sub>3</sub> magnetization state without additional ferromagnetic sensor layers, enabling facile detection of spin-orbit torques on layered magnetic insulators. Further exploration is required to understand the electron-magnon coupling in these devices and to potentially study bosonic topological matter in honeycomb ferromagnets (35, 36).

## REFERENCES AND NOTES

- W.-K. Tse, Z. Qiao, Y. Yao, A. H. MacDonald, Q. Niu, *Phys. Rev. B* **83**, 155447 (2011).
- D. Zhong *et al.*, *Sci. Adv.* **3**, e1603113 (2017).
- M. A. McGuire, H. Dixit, V. R. Cooper, B. C. Sales, *Chem. Mater.* **27**, 612–620 (2015).
- B. Huang *et al.*, *Nature* **546**, 270–273 (2017).
- X. Wang *et al.*, *2D Mater.* **3**, 031009 (2016).
- J.-U. Lee *et al.*, *Nano Lett.* **16**, 7433–7438 (2016).
- M. A. McGuire, *Crystals (Basel)* **7**, 121 (2017).
- T. Kurumaji *et al.*, *Phys. Rev. Lett.* **106**, 167206 (2011).
- A. Banerjee *et al.*, *Science* **356**, 1055–1059 (2017).
- C. Gong *et al.*, *Nature* **546**, 265–269 (2017).
- K. L. Seyler *et al.*, *Nat. Phys.* **14**, 277–281 (2018).
- D. C. Worledge, T. H. Geballe, *J. Appl. Phys.* **88**, 5277–5279 (2000).
- G.-X. Miao, M. Müller, J. S. Moodera, *Phys. Rev. Lett.* **102**, 076601 (2009).
- A. R. Mellnik *et al.*, *Nature* **511**, 449–451 (2014).
- T. Wakamura *et al.*, *Nat. Mater.* **14**, 675–678 (2015).
- P. Wadley *et al.*, *Science* **351**, 587–590 (2016).
- D. MacNeill *et al.*, *Nat. Phys.* **13**, 300–305 (2017).
- D. MacNeill *et al.*, *Phys. Rev. B* **96**, 054450 (2017).
- Q. Shao *et al.*, *Nano Lett.* **16**, 7514–7520 (2016).
- T. D. Skinner *et al.*, *Nat. Commun.* **6**, 6730 (2015).
- L. Esaki, P. J. Stiles, S. von Molnar, *Phys. Rev. Lett.* **19**, 852–854 (1967).
- J. S. Moodera, X. Hao, G. A. Gibson, R. Meservey, *Phys. Rev. Lett.* **61**, 637–640 (1988).
- J. S. Moodera, R. Meservey, X. Hao, *Phys. Rev. Lett.* **70**, 853–856 (1993).
- Materials and methods are available as supplementary materials.
- R. C. Jaklevic, J. Lambe, *Phys. Rev. Lett.* **17**, 1139–1140 (1966).
- S. Jung *et al.*, *Sci. Rep.* **5**, 16642 (2015).
- L. Vitali, M. A. Schneider, K. Kern, L. Wirtz, A. Rubio, *Phys. Rev. B* **69**, 121414 (2004).
- D. C. Tsui, R. E. Dietz, L. R. Walker, *Phys. Rev. Lett.* **27**, 1729–1732 (1971).
- C. F. Hirjibehedin, C. P. Lutz, A. J. Heinrich, *Science* **312**, 1021–1024 (2006).
- A. Spinelli, B. Bryant, F. Delgado, J. Fernández-Rossier, A. F. Otte, *Nat. Mater.* **13**, 782–785 (2014).
- K. Yamaguchi, *Phys. Status Solidi, B Basic Res.* **236**, 634–639 (2003).
- J. L. Lado, J. Fernández-Rossier, *2D Mater.* **4**, 035002 (2017).
- W.-B. Zhang, Q. Qu, P. Zhu, C.-H. Lam, *J. Mater. Chem. C Mater. Opt. Electron. Devices* **3**, 12457–12468 (2015).
- A. Narath, *Phys. Rev.* **140**, A854–A862 (1965).
- S. A. Owerre, *Sci. Rep.* **7**, 6931 (2017).
- S. S. Pershoguba *et al.*, *Phys. Rev. X* **8**, 011010 (2018).
- D. Klein, Replication Data for: Probing magnetism in 2D van der Waals crystalline insulators via electron tunneling. Version 1, Harvard Dataverse (2018); <https://doi.org/10.7910/DVN/4RUPNW>.

## ACKNOWLEDGMENTS

We thank V. Fatemi and Y. Cao for helpful discussions and assistance with measurements. **Funding:** This work was supported by the Center for Integrated Quantum Materials under NSF grant DMR-1231319 as well as the Gordon and Betty Moore

Foundation's EPIQS Initiative through grant GBMF4541 to P.J.-H. Device fabrication was partly supported by the Center for Excitonics, an Energy Frontier Research Center funded by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences (BES), under award no. DESC0001088. D.R.K. acknowledges partial support by the NSF Graduate Research Fellowship Program under grant no. 1122374. J.L.L. acknowledges financial support from the ETH Zurich Postdoctoral Fellowship program. D.S. acknowledges the Marie Curie Cofund program at INL. J.F.-R. acknowledges support from PTDC/FIS-NAN/3668/2014. Growth of hexagonal boron nitride crystals at NIMS was supported by the Elemental Strategy Initiative conducted by the MEXT, Japan, and JSPS KAKENHI grant nos. JP15K21722 and JP25106006. Work done at Ames Laboratory (P.C. and S.M.) was supported by the DOE BES Division of Materials Sciences and Engineering. Ames Laboratory is operated for the DOE by Iowa State University under contract no. DE-AC02-07CH11358. S.M. was funded by the Gordon and Betty Moore Foundation's EPIQS Initiative through grant GBMF4411. **Author contributions:** D.R.K., D.M., and P.J.-H. conceived the project. D.R.K. and D.M. fabricated and measured devices and analyzed the data. K.W., T.T., S.M., and P.C. supplied the boron nitride crystals. E.N.-M. supplied the CrI<sub>3</sub> crystals. J.L.L., D.S., and J.F.-R. carried out theoretical calculations. All authors contributed to writing the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The data shown in the paper are available at Harvard Dataverse (37).

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/1218/suppl/DC1](http://www.sciencemag.org/content/360/6394/1218/suppl/DC1)  
Materials and Methods  
Supplementary Text  
Figs. S1 to S8  
Table S1  
References (38–41)

21 November 2017; accepted 24 April 2018  
Published online 3 May 2018  
10.1126/science.aar3617

## NEUROSCIENCE

# Normative brain size variation and brain shape diversity in humans

P. K. Reardon<sup>1,2,3\*</sup>, Jakob Seidlitz<sup>1,4\*</sup>, Simon Vandekar<sup>5</sup>, Siyuan Liu<sup>1</sup>, Raihaan Patel<sup>6,7</sup>, Min Tae M. Park<sup>6,8</sup>, Aaron Alexander-Bloch<sup>9</sup>, Liv S. Clasen<sup>1</sup>, Jonathan D. Blumenthal<sup>1</sup>, Francois M. Lalonde<sup>1</sup>, Jay N. Giedd<sup>10</sup>, Ruben C. Gur<sup>11</sup>, Raquel E. Gur<sup>11</sup>, Jason P. Lerch<sup>12</sup>, M. Mallar Chakravarty<sup>6,7</sup>, Theodore D. Satterthwaite<sup>11</sup>, Russell T. Shinohara<sup>5</sup>, Armin Raznahan<sup>1†</sup>

Brain size variation over primate evolution and human development is associated with shifts in the proportions of different brain regions. Individual brain size can vary almost twofold among typically developing humans, but the consequences of this for brain organization remain poorly understood. Using *in vivo* neuroimaging data from more than 3000 individuals, we find that larger human brains show greater areal expansion in distributed frontoparietal cortical networks and related subcortical regions than in limbic, sensory, and motor systems. This areal redistribution recapitulates cortical remodeling across evolution, manifests by early childhood in humans, and is linked to multiple markers of heightened metabolic cost and neuronal connectivity. Thus, human brain shape is systematically coupled to naturally occurring variations in brain size through a scaling map that integrates spatiotemporally diverse aspects of neurobiology.

**T**otal brain size can vary almost twofold among typically developing humans of the same age (1). Brain size variation has been linked to coordinated changes in the proportional size of different brain systems across primate evolution and development (2), but the relationship between interindividual variation in human brain size and brain shape remains less well understood (3). We mapped this relationship at high spatial resolution to identify organizational shifts accompanying human brain size variation and illuminate differential areal patterning in larger versus smaller brains.

Our study included 2904 structural magnetic resonance imaging brain scans from two independent primary cohorts: (i) a Philadelphia Neurodevelopmental Cohort (PNC) sample of 1373 cross-sectional scans from a 3 Tesla machine in youth aged 8 to 23 years (table S1A and fig. S1A) (4) and (ii) a National Institutes of Health (NIH) sample of 1531 longitudinally acquired brain scans

from a 1.5 Tesla machine in 792 youth aged 5 to 25 years (table S1B and fig. S1B) (1). To generate a reference map of areal scaling in the cortex, we measured the local surface area associated with each of ~80,000 cortical points per scan (henceforth “vertex area”) using an automated image-processing pipeline (5) and then used semiparametric generalized additive models (6) to estimate vertex-specific scaling as the log-log regression coefficient for total cortical area as a predictor of vertex area (Methods). Within this regression framework (7), a coefficient of 1 indicates linear scaling (e.g., doubling of vertex area with a doubling of cortical area), whereas deviation from 1 indicates nonlinear scaling: Coefficients >1 indicate that proportional vertex area increases with greater cortical size (“positive scaling”), and coefficients <1 indicate that proportional vertex area decreases (“negative scaling”). The models used to estimate scaling coefficients provided statistical control for age and sex effects on vertex area, after first ruling out statistically significant interactions between either of these variables and total cortical area (Methods). Thus, our results supported estimation of a single scaling map for each of the two developmental cohorts examined, which did not vary as a function of age and sex.

In both cohorts, scaling relationships between vertex area and total cortical area varied across the cortical sheet (Fig. 1A) in a manner that was broadly symmetric (but see fig. S2 for exceptions) and reproducible in the adult subset of each cohort (Methods, fig. S3A, and table S2). Intervertex differences in scaling were highly correlated between PNC and NIH cohorts (Pearson  $r = 0.7$ ), at levels above chance in a spatial permutation procedure that relies on random surface-based rotation of cortical maps (henceforth “spins”; i.e.,  $p_{\text{SPIN}} < 0.001$ ) (Methods, Fig. 1A, and table S2). Testing for statistically significant deviation from 1 of

vertex scaling coefficients and correcting for multiple comparisons across vertices (Methods) defined two distributed domains of statistically significant, nonlinear areal scaling in each cohort (Fig. 1B). Across cohorts, we observed regions of positive areal scaling in prefrontal, lateral temporoparietal, and medial parietal cortex, and negative areal scaling in limbic, primary visual, and primary sensorimotor regions (Fig. 1C). The reproducibility of areal scaling across cohorts was unlikely to reflect intrinsic methodological artifacts from a shared cortical analysis pipeline (5) because these scaling patterns were (i) lost after permuting scans across individuals in the PNC cohort (fig. S3B and table S2) and (ii) seen after processing a third independent magnetic resonance imaging (MRI) dataset from the Human Connectome Project (HCP) ( $n = 1113$ ) (Methods) with a different computational pipeline (FreeSurfer; Methods), at varying smoothing kernels (Fig. 1D, fig. S4, and table S2). Thus, log-log regression analysis revealed regionally specific patterns of nonlinear areal scaling in the human cortex that were broadly reproducible across three study cohorts, three image-acquisition platforms, two image processing pipelines, and a range of smoothing kernels.

To complement log-log regression analysis of scaling, we also generated person-level measures of cortical proportionality by expressing vertex area estimates as proportions of the total cortical surface area in each scan. Variability in proportional vertex area differed across the cortical sheet (fig. S5A), and maps for the relationship between proportional vertex area and total cortical area (fig. S5B) recapitulated the scaling gradients detected by log-log regression (Fig. 1A). Across individuals, raw surface area within regions of positive and negative scaling (Fig. 1B) increased with total cortical size, whereas the total proportional areas of these regions were positively and negatively (respectively) related to total cortical area (Fig. 1E). The ratio between mean proportional vertex area in regions of positive versus negative scaling—a summary “scaling index” that could be computed for each scan—showed a robust positive linear relationship with total cortical area in both cohorts (fig. S5C) and a positive relationship with measures of intelligence quotient (IQ) that were available for the NIH cohort ( $P = 0.004$ ) (Methods, Fig. 1F, and table S3). However, scaling index variation predicted a small fraction of IQ variance (~1%), and this association did not survive correction for total cortical size (Methods, Fig. 1F, and table S3), which was itself more robustly associated with IQ (see standardized coefficients and model  $R^2$ , table S3). Thus, mounting cortical size was simultaneously associated with a greater scaling index and IQ, but differences in scaling above and beyond those predicted by cortical size did not explain additional variance in IQ.

To assess whether regional differences in areal scaling were specific to the cortical sheet, we mapped areal scaling across five non-neocortical (henceforth “subcortical”) structures (thalamus, pallidum, striatum, hippocampus, and amygdala) using recently developed algorithms for automated subcortical shape analysis (MAGeT Brain, Methods). This

<sup>1</sup>Developmental Neurogenomics Unit, National Institute of Mental Health, NIH, Bethesda, MD, USA. <sup>2</sup>Department of Physiology, Anatomy and Genetics, Oxford University, UK. <sup>3</sup>Vagelos College of Physicians and Surgeons, New York, NY, USA. <sup>4</sup>Department of Psychiatry, Cambridge University, Cambridge, UK. <sup>5</sup>Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA. <sup>6</sup>Cerebral Imaging Center, Douglas Mental Health University Institute, Montreal, Quebec, Canada. <sup>7</sup>Department of Biological and Biomedical Engineering, McGill University, Montreal, Quebec, Canada. <sup>8</sup>Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada. <sup>9</sup>Department of Psychiatry, Yale University, New Haven, CT, USA. <sup>10</sup>Department of Psychiatry, University of California—San Diego, La Jolla, CA, USA. <sup>11</sup>Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104, USA. <sup>12</sup>Mouse Imaging Center, Hospital for Sick Children, Toronto, ON, Canada.

\*These authors contributed equally to this work.

†Corresponding author. Email: raznahan@mail.nih.gov

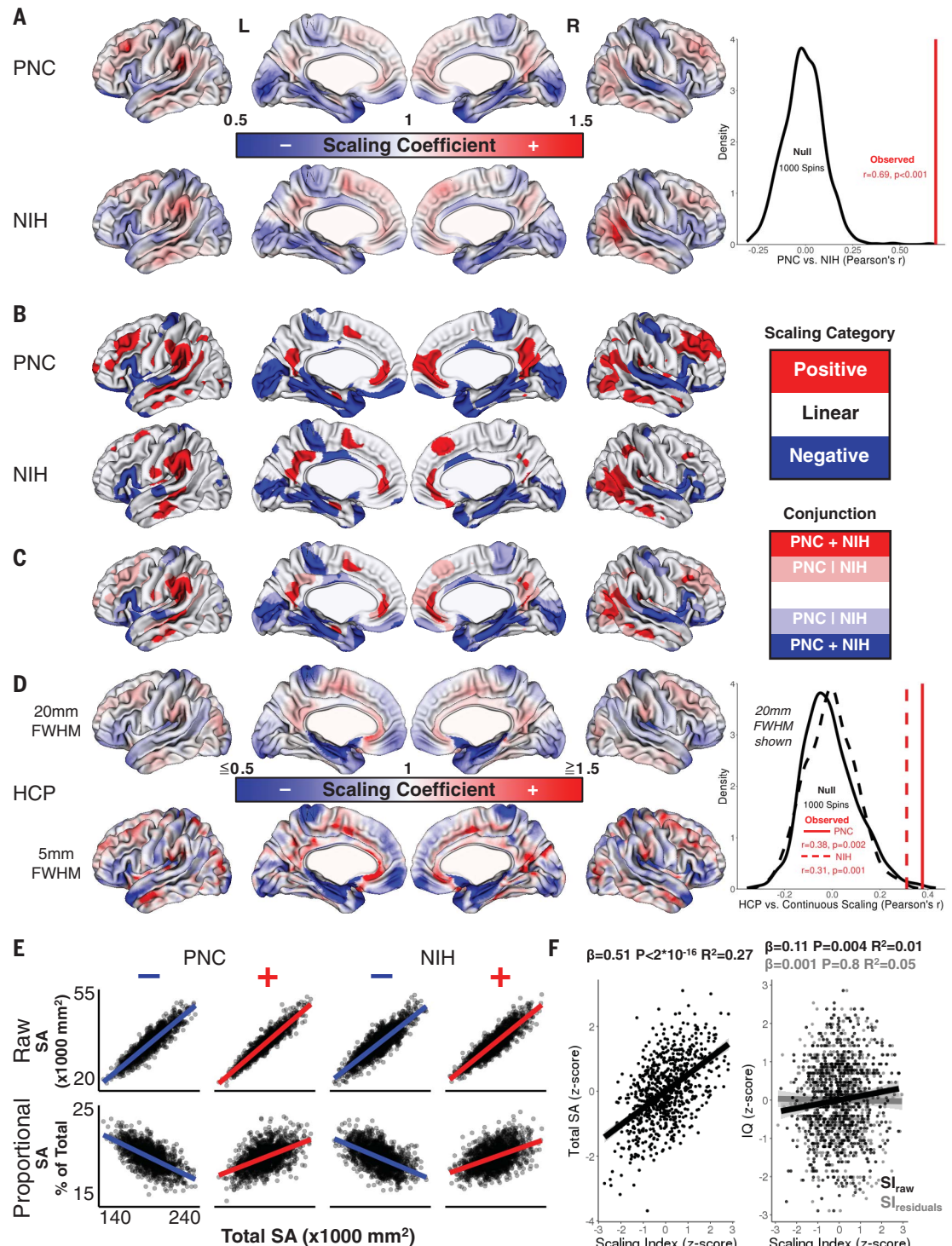


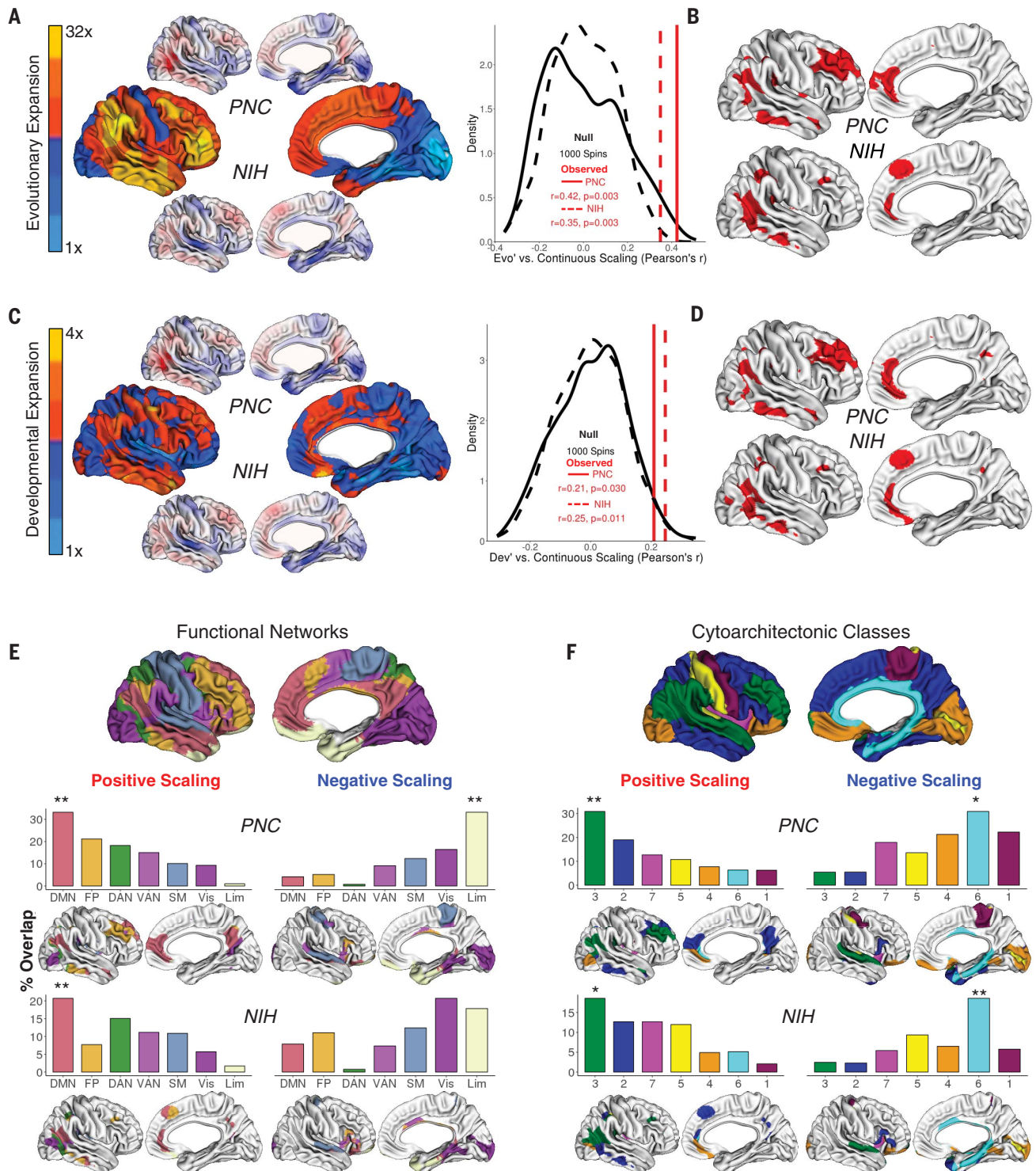
approach provided a homogeneous surface-based framework for quantification and visualization of areal scaling gradients across the cortex and subcortex (Methods, fig. S6, A and B). Vertex-level areal scaling coefficients varied within each subcortical structure examined (e.g., positive scaling in hippocampal head versus negative scaling in tail) (fig. S6C), revealing that size-related shifts in human brain shape are not restricted to the neocortex.

To assess the biological importance of spatially patterned areal scaling within the human brain, we compared regional differences in cortical scaling to several independent assays of cortical organization. We first investigated whether patterns of cortical area redistribution as a function of normative brain size variation in humans (Fig. 1A) aligned with those that accompany evolutionary and developmental changes in primate brain size (Fig. 2, A to D) (2). Intervertebral scaling

variations in PNC and NIH cohorts (Fig. 1A) were positively correlated with those seen in evolution and development (Fig. 2, A to C, and table S2). All three sources of primate brain size disparity (evolution, development, and naturally occurring size variation) involved disproportionate areal expansion within anterior cingulate, angular gyrus, superior parietal lobule, and lateral temporal cortex (Fig. 2, B and D). Within these regions, the magnitude of positive areal scaling between

**Fig. 1. Nonlinear areal scaling of the cortex with normative brain size variation in PNC, NIH, and HCP cohorts.** (A) Unthresholded vertex maps showing local surface area scaling with naturally occurring variations in total cortical area. Red, relative expansion in larger cortices ("positive scaling"); blue, relative contraction ("negative scaling"). The observed cross-vertex correlation in scaling between PNC/NIH cohorts is greater than that in 1000 surface-based rotations of the NIH scaling map (i.e.,  $p_{\text{SPIN}} < 0.001$ , density plot). (B) Categorical scaling maps showing regions of statistically significant positive and negative areal scaling (i.e.,  $H_0$ : scaling coefficient = 1 rejected) after correction for multiple comparisons across vertices. (C) Conjunction of PNC and NIH maps from Fig. 1B. (D) Regional scaling in a third independent dataset (Human Connectome Project,  $n = 1113$ ), across two different FWHM (full width at half maximum) smoothing kernels (for maps for all five kernels, see fig. S4). Density plot shows that observed alignment of HCP scaling with PNC(solid)/NIH (dashed) ( $r > 0.3$ ) exceeds chance ( $p_{\text{SPIN}} < 0.002$ ). (E) Scatter plots of raw (top row) and proportional (bottom row) surface area in regions of nonlinear scaling from Fig. 1B versus total cortical area (SA). (F) Interrelationships between age and sex residualized scaling index (SI, with and without residualization for SA), SA, and IQ in the NIH cohort.





**Fig. 2. Areal scaling aligns with patterns of cortical evolution, development, functional network topography, and cytoarchitecture.**

(A) Area expansion map in humans relative to macaques (2), with PNC/NIH scaling maps for comparison. Density plot shows that observed spatial correlation of scaling maps and evolutionary expansion is greater than chance for PNC (solid) and NIH (dashed) cohorts. (B) Conjunction between regions of positive areal scaling in PNC/NIH cohorts (Fig. 1B) and vertices with evolutionary expansion values above the 50th centile. (C and D) Identical analyses as Fig. 2, A and B, except for areal expansion map in human adults relative to human

infants (2). (E) Parcellation of cortex into seven canonical resting state functional connectivity networks (8), with bar plots and conjunction maps showing differential representation of positive versus negative scaling in each network (\*\* $p_{\text{SPIN}} < 0.001$ , \* $p_{\text{SPIN}} < 0.05$ ). Regions of positive scaling localize to the default mode network (DMN) and regions of negative scaling to the limbic network (Lim). (F) Cortical parcellation into seven different laminar types according to a classical cytoarchitectonic atlas (12), with bar plots and conjunction maps showing that regions of positive scaling localize to von Economo Type 3 cortex, and negative to Type 6 (\*\* $p_{\text{SPIN}} < 0.001$ , \* $p_{\text{SPIN}} < 0.05$ ).



macaques and humans tended to exceed that seen within humans (fig. S7A).

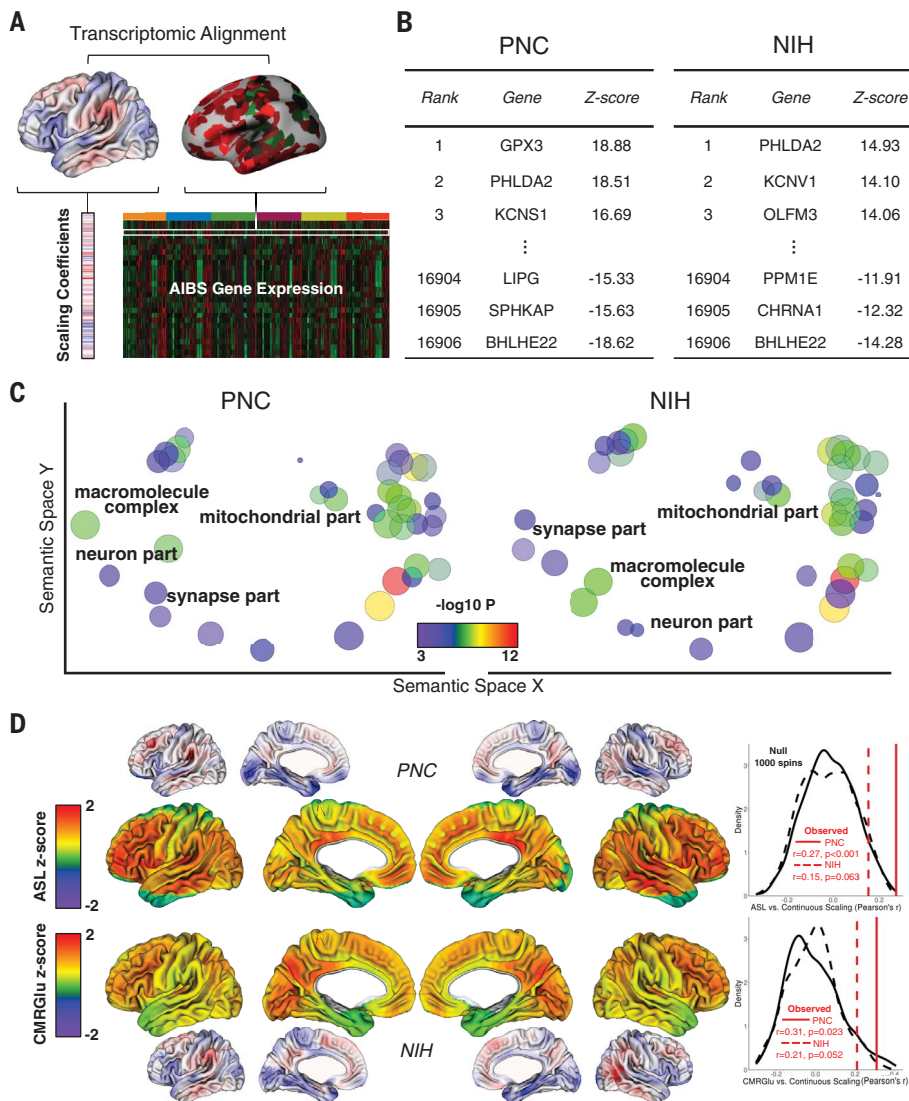
We next asked whether regional differences in human cortical scaling were related to functional and microstructural topography of the human cortex. Using a previously published parcellation of the cortical sheet into seven canonical resting-state functional connectivity networks (8), we found that (i) regions of positive scaling were concentrated within association cortices, including the default mode (DMN), dorsal attention, and frontoparietal networks ( $p_{\text{SPIN}} = 0.001$  for

DMN), whereas (ii) regions of negative scaling were overrepresented ( $p_{\text{SPIN}} = 0.007$ ) within the limbic network (Fig. 2E and table S4). These associations were replicated using a finer-grained functional parcellation of the human cortex (8) (fig. S7B and table S4), and indicated that larger human brains show preferential areal expansion within regions of association cortex that sit at the apex of a functional network hierarchy (9, 10) and provide high-level integration across lower systems (11). This theme was reinforced by comparison of scaling maps with a classical parcel-

lation of the cortex into seven cytoarchitectonic "types" with differing laminar organizations (12): Regions of positive areal scaling were overrepresented within "von Economo Type 3" cortices (Fig. 2F and table S4) that bear cytoarchitectonic specializations for long-range cortico-cortical connectivity (e.g., thickening of supragranular layers II/III). Thus, regions of preferential areal expansion in larger versus smaller human brains appeared to be both functionally and microstructurally suited to operate as hubs of information integration.

To next probe potential molecular correlates of regional differences in cortical scaling, we used a shared stereotaxic coordinate system to assign an areal scaling coefficient (Fig. 1A) to each of 1939 spatially distributed and transcriptomically characterized cortical samples from an independent cohort of six adult human donors provided by the Allen Institute for Brain Science (AIBS) (13) (Methods, Fig. 3A, and table S5). This data alignment allowed us to rank ~16,000 genes by the spatial correspondence between their expression and cortical scaling gradients in PNC and NIH cohorts (Methods, table S6), with expression of the high-ranking genes being most positively correlated with scaling (Fig. 3B). Rank-based gene ontology (GO) analysis (Methods) (14) revealed that high-ranking genes were significantly enriched for mitochondrial and synaptic GO annotations (Fig. 3C) and related processes of oxidative phosphorylation and transmembrane  $K^+$  transport (table S7). Observed gene ranks were robust to randomly excluding up to 70% of samples per donor, and the high rank of genes associated with mitochondrial and synaptic GO terms was lost after permutation of scaling values across AIBS samples (Methods). The positive association between areal scaling and postmortem expression of mitochondria-related genes (Fig. 3C and table S7) suggested that cortical regions that are preferentially expanded in larger versus smaller human brains may possess a distinct energy metabolism profile. This hypothesis was supported by convergent evidence for a statistically significant positive association between regional differences in cortical areal scaling and regional differences in two different neuroimaging proxies for cortical energy consumption at rest (15, 16) (Fig. 3D and table S2).

Our findings illuminate several aspects of cortical patterning. First, the spatial convergence of areal scaling maps across all three major axes of primate brain size variation (Fig. 2)—evolution, development, and standing interindividual variation—implies shared mechanisms for size-dependent patterning of the primate cortical sheet. These mechanisms presumably link variation in early progenitor cell proliferation to (i) the genesis of regional differences in cellular composition during prenatal corticogenesis (17, 18) or (ii) the subsequent emergence of regional differences in cellular morphology and neuropil composition (19). Second, the anabolic costs of tissue growth (20) and the overlap of positive areal scaling with diverse markers of biological investment (Fig. 3) suggest that preferential expansion of association cortex (e.g., DMN) may serve to maintain



**Fig. 3. Areal scaling aligns with cortical gene expression and metabolism.** (A) Vertex scaling coefficients from PNC and NIH maps (Fig. 1A) were uniquely assigned to each of ~2k cortical samples from the Allen Institute for Brain Sciences (AIBS) (13) by spatial proximity, allowing ranking of the ~16,000 genes measured across all AIBS samples by their spatial correlation with areal scaling. (B) Extreme-ranking genes for the PNC and NIH scaling maps. (C) Visualization of GO cellular component terms in semantic space showing statistically significant enrichment of mitochondria- and synapse-related terms among top-ranking genes. (D) Areal scaling is positively correlated with two neuroimaging-based proxies for cortical energy consumption at rest: arterial spin labeling (ASL) measures of arterial blood flow (15) and positron emission tomography measures of glucose uptake (CMRGlu) (16) (density plots, PNC solid, NIH dashed, black null, red observed, mean  $p_{\text{SPIN}} < 0.03$ ).

or enhance (21) brain function in larger versus smaller human brains. Testing this hypothesis will require new study designs that can probe diverse assays of human brain function beyond IQ while untangling the effects of brain size variation from linked changes in cortical patterning (Fig. 1F). Finally, the convergent evidence that cortical regions of positive scaling are specialized for integration of information across lower-order systems (Figs. 2 and 3) offers a potential mechanistic account for positive areal scaling in the primate brain. Just as the computational load of an integrative algorithm can increase supralinearly with the size of its inputs (22), larger cortices may need to disproportionately expand the anatomical substrates for integrative computation in association cortex. Based on our results (Figs. 2F and 3C) and available comparative histology (23), we speculate that these substrates include dendritic branching and synaptic spine density in supragranular neuropil.

In summary, naturally occurring variations in human brain size are accompanied by systematic changes in brain shape through scaling gradients that tie together macroscopic, microscopic, and evolutionary features of the human brain.

#### REFERENCES AND NOTES

1. J. N. Giedd *et al.*, *Neuropsychopharmacology* **40**, 43–49 (2015).
2. J. Hill *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 13135–13140 (2010).
3. C. Mankiw *et al.*, *J. Neurosci.* **37**, 5221–5231 (2017).
4. T. D. Satterthwaite *et al.*, *Neuroimage* **86**, 544–553 (2014).
5. Y. Ad-Dab'bagh *et al.*, The CIVET image-processing environment: A fully automated comprehensive pipeline for anatomical neuroimaging research. Proceedings of the 12th Annual Meeting of the Organization for Human Brain Mapping (2006).
6. M. Stasinopoulos, *Biometrics* **63**, 1298–1299 (2007).
7. J. Huxley, *Nature* **114**, 895–896 (1924).
8. B. T. T. Yeo *et al.*, *J. Neurophysiol.* **106**, 1125–1165 (2011).
9. O. Sporns, *Curr. Opin. Neurobiol.* **23**, 162–171 (2013).
10. D. S. Margulies *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12574–12579 (2016).
11. J. D. Medaglia, M.-E. Lynall, D. S. Bassett, *J. Cogn. Neurosci.* **27**, 1471–1491 (2015).
12. C. F. von Economo, G. N. Koskinas, *Die Cytoarchitektonik der Hirnrinde des Erwachsenen Menschen* (J. Springer, 1925).
13. M. J. Hawrylycz *et al.*, *Nature* **489**, 391–399 (2012).
14. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, *BMC Bioinformatics* **10**, 48 (2009).
15. T. D. Satterthwaite *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8643–8648 (2014).
16. M. F. Glasser, M. S. Goyal, T. M. Preuss, M. E. Raichle, D. C. Van Essen, *Neuroimage* **93**, 165–175 (2014).
17. I. Reillo, C. de Juan Romero, M. Á. García-Cabezas, V. Borrell, *Cereb. Cortex* **21**, 1674–1694 (2011).
18. D. J. Cahalane, C. J. Charvet, B. L. Finlay, *Front. Neuroanat.* **6**, 28 (2012).
19. G. N. Elston, I. Fujita, *Front. Neuroanat.* **8**, 78 (2014).
20. C. W. Kuzawa *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13010–13015 (2014).
21. M. Rubinov, *Nat. Commun.* **7**, 13812 (2016).
22. M. Sipser, *Introduction to the Theory of Computation* (Thompson Course Technology, ed. 2, 2006).
23. G. N. Elston, *Cereb. Cortex* **13**, 1124–1138 (2003).

#### ACKNOWLEDGMENTS

The authors thank all research volunteers within the cohorts studied for their participation and all the reviewers for their helpful comments. A.R. thanks SVR and AIR for their support. **Funding:** This research was partially funded by the NIMH Intramural (ZIA MH002794, PI A.R.) and Extramural (R01MH107235, R01NS085211, R01NS060910, R01MH112847, and R01MH107703, PIs S.V., R.C.G., R.E.G., T.D.S., and R.T.S.) Research Programs. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (PIs: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. **Author contributions:** Conceptualization, P.K.R. and A.R.; Methodology, all; Software, R.P., M.T.M.P., J.P.L., and M.M.C.; Investigation and Visualization, P.K.R., S.L., J.S., and A.R.; Resources, S.V., A.A.-B., L.S.C., J.D.B., F.M.L., R.C.G., R.E.G., T.D.S., and R.T.S.; Writing, all. **Competing interests:** None declared. **Data and materials availability:** PNC, Genotypes and Phenotypes (dbGaP accession phs000607.v2, p2). NIH, Data available on request through A.R. HCP, S1200 release from humanconnectome.org. Scaling maps presented in this manuscript are available for download at <https://neurovault.org>.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/1222/suppl/DC1](http://www.sciencemag.org/content/360/6394/1222/suppl/DC1)  
Materials and Methods  
Figs. S1 to S7  
Tables S1 to S7  
References (24–60)

19 October 2017; accepted 9 April 2018  
Published online 31 May 2018  
10.1126/science.aar2578



## NEUROSCIENCE

# Synapse-specific representation of the identity of overlapping memory engrams

Kareem Abdou<sup>1,2\*</sup>, Mohammad Shehata<sup>1,2\*†‡</sup>, Kiriko Choko<sup>1,2</sup>, Hirofumi Nishizono<sup>2,3</sup>, Mina Matsuo<sup>3</sup>, Shin-ichi Muramatsu<sup>4,5</sup>, Kaoru Inokuchi<sup>1,2§</sup>

Memories are integrated into interconnected networks; nevertheless, each memory has its own identity. How the brain defines specific memory identity out of intermingled memories stored in a shared cell ensemble has remained elusive. We found that after complete retrograde amnesia of auditory fear conditioning in mice, optogenetic stimulation of the auditory inputs to the lateral amygdala failed to induce memory recall, implying that the memory engram no longer existed in that circuit. Complete amnesia of a given fear memory did not affect another linked fear memory encoded in the shared ensemble. Optogenetic potentiation or depotentiation of the plasticity at synapses specific to one memory affected the recall of only that memory. Thus, the sharing of engram cells underlies the linkage between memories, whereas synapse-specific plasticity guarantees the identity and storage of individual memories.

**M**emories are formed through long-term changes in synaptic efficacy, a process known as synaptic plasticity (1–7), and are stored in the brain in specific neuronal ensembles called engram cells, which are reactivated during memory retrieval (8–13). When two memories are associated, cell ensembles corresponding to each memory overlap (14–19) and are responsible for the association (18). Although multiple associated memories can be encoded in the overlapping population of cells, each memory has its own identity (14, 18). Synaptic plasticity is essential for the retrieval, but not the storage, of associative fear memories (5, 20, 21). However, how the brain defines the identity of a particular memory amid the many memories stored in the same ensemble has been elusive.

We asked whether individual memories stored in a shared neuronal ensemble would maintain their identities and have a different fate if one memory was erased by complete retrograde amnesia. We subjected mice to auditory fear conditioning (AFC), in which a tone was associated with a foot shock. This association is mediated by synaptic plasticity between neuron terminals of the auditory cortex (AC) and the medial part of the medial geniculate nucleus (MGm)

and neurons of the lateral amygdala (LA) (22). Two different tones, at 2 and 7 kHz, were used. Mice discriminated between the two tones and showed a freezing response only to the 7-kHz tone that was paired with shock (figs. S1 and S2). To completely erase memories, we used autophagy, which is a major protein degradation pathway wherein the autophagosome sequesters a small portion of the cytoplasm and fuses with the endosome-lysosome system to degrade the entrapped contents. Autophagy contributes to synaptic plasticity (23, 24), and its induction by the peptide tat-beclin enhances destabilization of synaptic efficacy after reactivation of these synapses through the degradation of endocytosed  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (25, 26). When tat-beclin is combined with inhibition of protein synthesis after memory retrieval, complete retrograde amnesia is induced through enhanced memory destabilization and reconsolidation inhibition (25).

To optogenetically manipulate specific memories, we used c-Fos::tTA transgenic mice; we injected adeno-associated virus (AAV) expressing Cre recombinase under the control of tetracycline-responsive element (TRE) in combination with AAV encoding DIO-oChIEF-citrine, downstream of the human synapsin (hSyn) 1 promoter, into the AC and MGm (both of which relay auditory information to the LA) to label the activated ensemble with a channelrhodopsin variant, oChIEF (Fig. 1, A to D, and fig. S2). Mice were trained with AFC (7-kHz tone plus shock) 2 days after doxycycline withdrawal (OFF DOX). One day later, under the ON DOX condition, the LAs of these mice were infused with phosphate-buffered saline (PBS), anisomycin, or anisomycin combined with tat-beclin (Ani+tBC) immediately after the test session (day 5). The anisomycin infusion induced partial retrograde amnesia, whereas Ani+tBC accomplished complete amnesia, with the freezing level comparable to that of non-

shocked and unpaired control groups (Fig. 1E). Optogenetic activation of the axonal terminals of the AC and MGm engram cells in the LA induced fear memory recall in the PBS and anisomycin groups, which is consistent with a previous study (21), whereas it failed to do this in the Ani+tBC-treated mice (Fig. 1F).

To further confirm memory erasure, we tried to recover the erased memories by using optical long-term potentiation (LTP). High-frequency optical stimulation of the terminals of AC and MGm engram cells to the LA led to long-lasting potentiated field responses (fig. S3, A and B). Optical LTP allowed anisomycin-treated mice to completely recover from amnesia to the PBS group's freezing level, which was specific to the 7-kHz conditioned tone (i.e., it did not generalize to the 2-kHz tone) (Fig. 1G). In the Ani+tBC-treated mice, optical LTP failed to completely recover the fear memory; these mice showed only a slight increase in the freezing level, which was similar to that which occurred in the unpaired control group (Fig. 1G). Because the unpaired conditioning did not form an associative fear memory, this slight increase in the freezing response might be attributed to the formation of a new artificial associative memory, rather than restoration of a previously stored associative memory.

In a remote memory test, the Ani+tBC group displayed significantly lower freezing than the anisomycin or PBS groups in both natural cue and optogenetic tests, indicating that memory erasure was long-lasting and that the memory did not undergo spontaneous recovery over time (Fig. 1, H and I). The Ani+tBC-treated mice that received LTP showed light-induced freezing comparable to that of the PBS group (test 9), excluding the possibility of LA damage from Ani+tBC treatment. Furthermore, similar results were obtained when engram cells in the LA were labeled and manipulated similarly but optical (instead of tone) recall was used (fig. S4).

To examine the synaptic mechanism underlying the complete retrograde amnesia, we conducted a LTP occlusion experiment, in which artificial induction of LTP was occluded in circuits with potentiated synapses, whereas it was facilitated in circuits with unpotentiated synapses (27–29). High-frequency optical stimulation 1 day after test 1 induced LTP in the Ani+tBC group that was comparable to that in the nonshock group but significantly higher than that in the PBS and anisomycin groups (Fig. 2, A to D). Thus, synaptic plasticity was totally reset and returned to nonshock levels after complete amnesia.

This conclusion was further supported by analysis of functional connectivity. Using c-Fos::tTA/R26R::H2B-mCherry double transgenic mice (18), we measured the connectivity pattern between upstream and downstream engram cells after memory erasure. Engram cells in the LA were labeled with Cre-dependent mCherry, and the axonal terminals of the AC and MGm engram cells were labeled with oChIEF. The terminals were optogenetically stimulated at 10 Hz, and the number of cells that were double-positive for mCherry and c-Fos, which represented the degree of functional connectivity

<sup>1</sup>Department of Biochemistry, Graduate School of Medicine and Pharmaceutical Sciences, University of Toyama, Toyama 930-0194, Japan. <sup>2</sup>Japan Science and Technology Agency, CREST, University of Toyama, Toyama 930-0194, Japan.

<sup>3</sup>Division of Animal Experimental Laboratory, Life Science Research Center, University of Toyama, Toyama 930-0194, Japan. <sup>4</sup>Division of Neurology, Department of Medicine, Jichi Medical University, Tochigi 3290498, Japan. <sup>5</sup>Center for Gene and Cell Therapy, The Institute of Medical Science, The University of Tokyo, Tokyo 1088639, Japan.

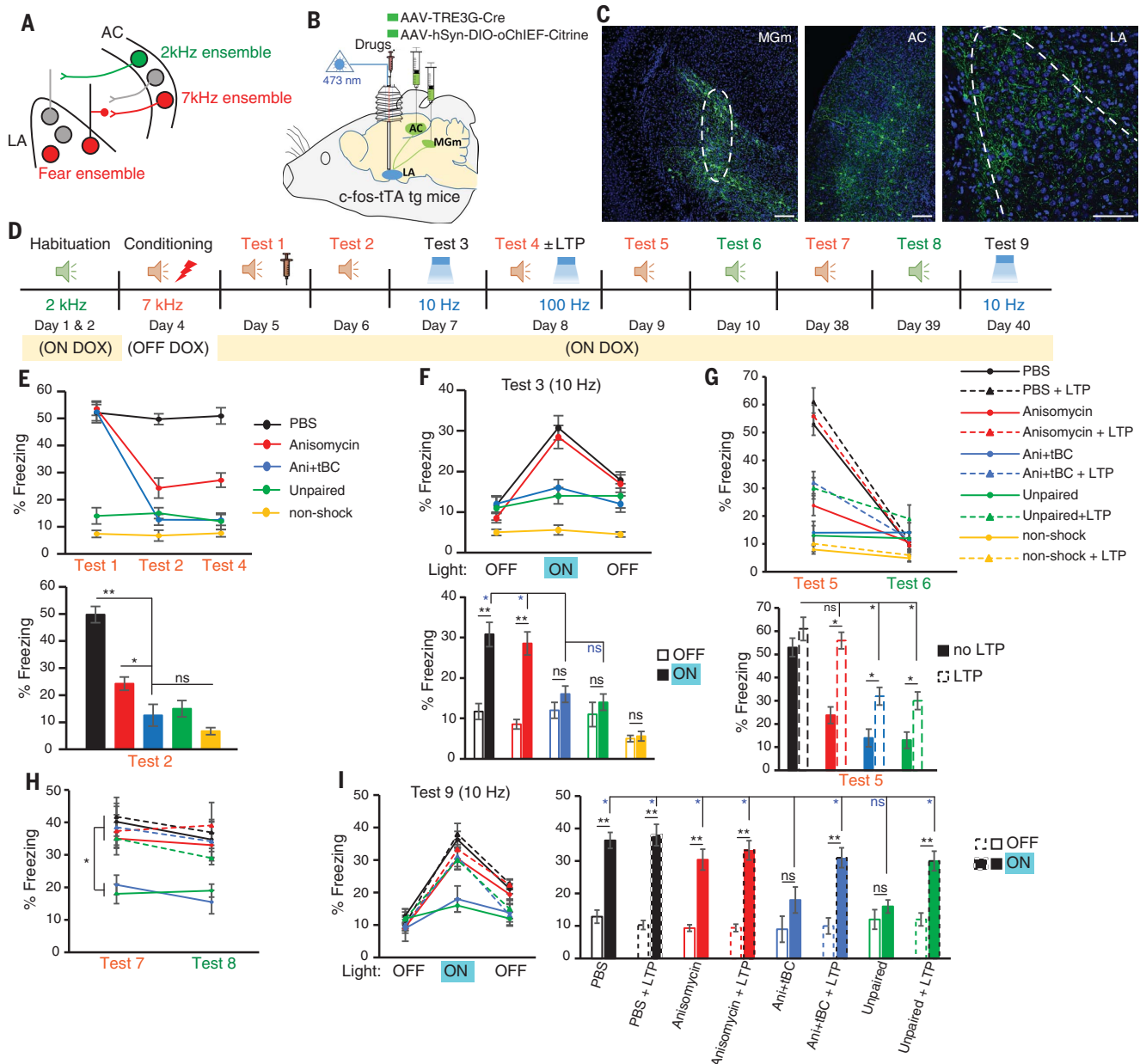
\*These authors contributed equally to this work. †Present address: Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA. ‡Present address: Mind and Brain Laboratory for Perceptual and Cognitive Processing, Toyohashi University of Technology, Toyohashi 441-8580, Japan. §Corresponding author. Email: inokuchi@med.u-toyama.ac.jp

between upstream and downstream engram cells, was counted (Fig. 2, E to I, and fig. S7, A and B). Complete amnesia resulted in a significant decrease in the c-Fos<sup>+</sup>-mCherry<sup>+</sup> overlap in the Ani+tBC group in comparison with the PBS and anisomycin groups, which is consistent with the

behavioral data and the total resetting of synaptic efficacy.

Considering that memories are stored in interconnected networks, and the brain can store two memories in a shared ensemble (14–18), we examined the effect of complete retrograde am-

nesia of one memory on another memory by using two different AFC events: a 7-kHz AFC (event 1) followed by a 2-kHz AFC (event 2) (Fig. 3). When these two events were separated by 5 hours, memory for event 2 was enhanced (fig. S5), indicating interaction between the memories



**Fig. 1. Complete and long-lasting erasure of fear memory trace from AC-LA and MGm-LA engram circuits.** (A) Model showing the ensemble responsive to the 2- and 7-kHz tones in the AC and the fear-responsive ensemble in the LA. (B) Labeling strategy for the AFC-responsive ensemble in the AC and MGm, using the c-Fos::tTA transgenic mice. (C) Expression of oChIEF in AC and MGm neurons and their axonal terminals in the LA. Dashed lines show the borders of the MGm and LA. Scale bars, 100  $\mu$ m. (D) Design of memory engram erasure experiment. Different chambers were used for each session. (E to I) Freezing levels (percent of time) before and after drug injection (E), during 10-Hz optical stimulation (F), in response to the conditioned and neutral tones after optical LTP (G), at a remote time point (H), and during 10-Hz

stimulation at a remote time point (I).  $n = 20$  mice per group in (E) and (F) and 10 mice per group in (G) to (I). Bottom panels of (E) to (G) show statistical significance between groups during test 2 (E), during light-off and light-on epochs (F), and during test 5 (G). The right panel of (I) shows statistical significance within and between groups during test 9. Statistical comparisons were performed using one-way analysis of variance (ANOVA) [(E), (G), and (H)] and two-way ANOVA [(F) and (I)]. \* $P < 0.05$ ; \*\* $P < 0.01$ ; ns, not significant. In the bottom panels of (F), (G), and (I), the colors of the upper asterisks indicate the comparison (e.g., blue asterisks indicate a comparison with the Ani+tBC group). Data are represented as mean  $\pm$  SEM. Ani, anisomycin; tBC, tat-beclin.



(17). The majority of the LA engram cells for event 1 ( $mCherry^+$ ) also encoded event 2 ( $c-Fos^+$ ), whereas the memories were encoded in two distinct populations in the AC (Fig. 3, A to E, and fig. S7, C to F). When the two memories were separated by 24 hours, they were allocated to distinct populations in both the LA and the AC.

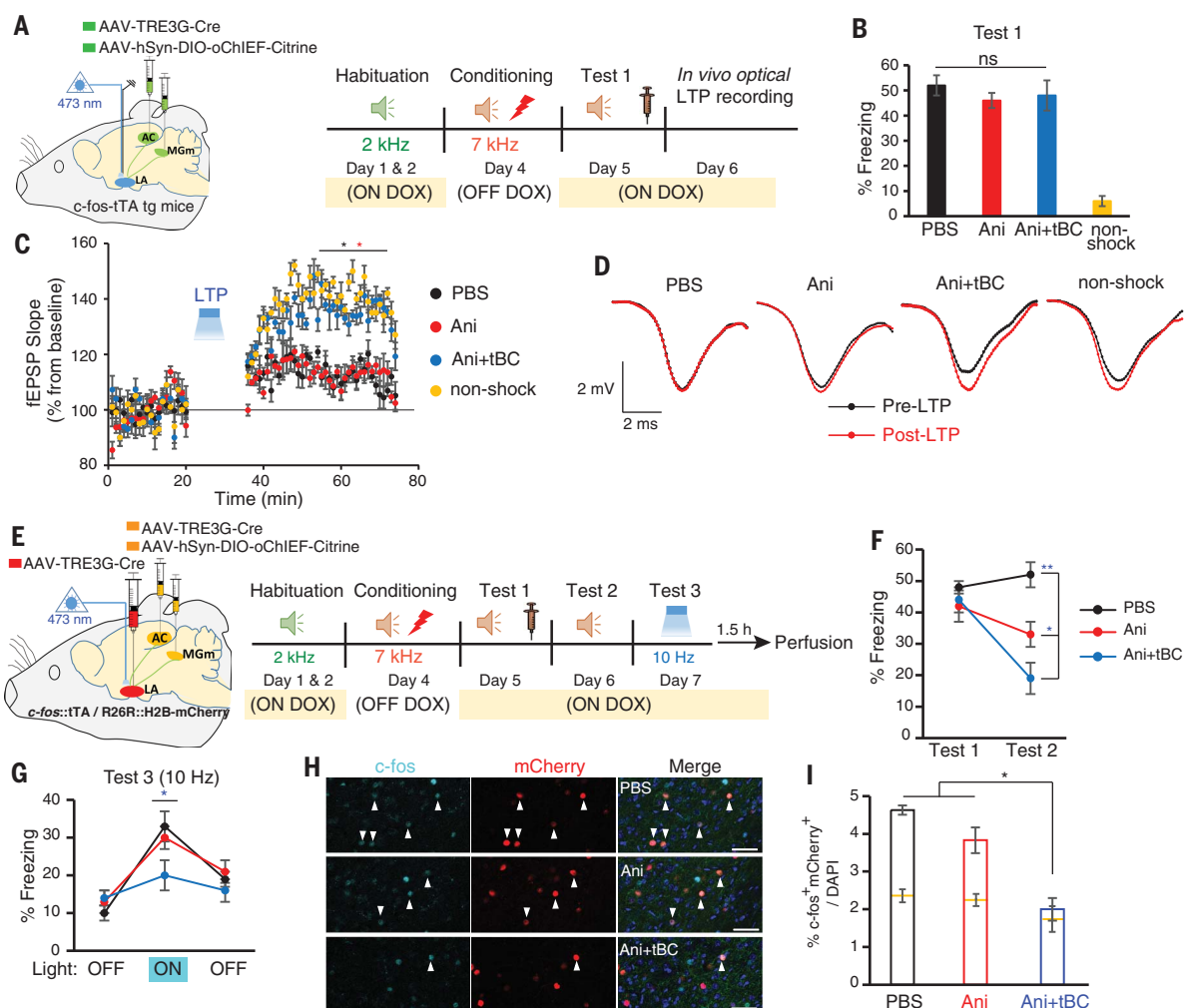
We then used the  $c-Fos::tTA$  transgenic mice to label the neural ensembles in the AC and MGm that were activated specifically during event 1 with  $oChIEF$  (Fig. 3F and fig. S6). After 5 hours ON DOX, mice were exposed to event 2 and then divided into two groups. The first group received PBS after event 1 memory retrieval and Ani+tBC after event 2 memory retrieval (gp1), whereas the second group received the opposite treatment (gp2). In gp1, memory of event 2 was erased by Ani+tBC (test 4), whereas memory of event 1 was

not affected (test 2). In contrast, in gp2, event 1 memory was erased (test 2), whereas event 2 memory was not affected (test 4; Fig. 3G). Moreover, optogenetic stimulation of the pre-synaptic terminals of the AC and MGm engram cells corresponding to event 1 memory induced a freezing response in gp1, but not in gp2, although in both groups, the LA neurons storing both associative memories underwent Ani+tBC treatment (Fig. 3H). These results reveal synapse-specific engram erasure and indicate that memories stored in the shared engram cells are synapse-specific and have different fates (Fig. 3I).

We then addressed the question of how each memory reserves its individual identity within the shared ensemble. We carried out a loss-of-function experiment using optical long-term depression (LTD) to depotentiate the synaptic

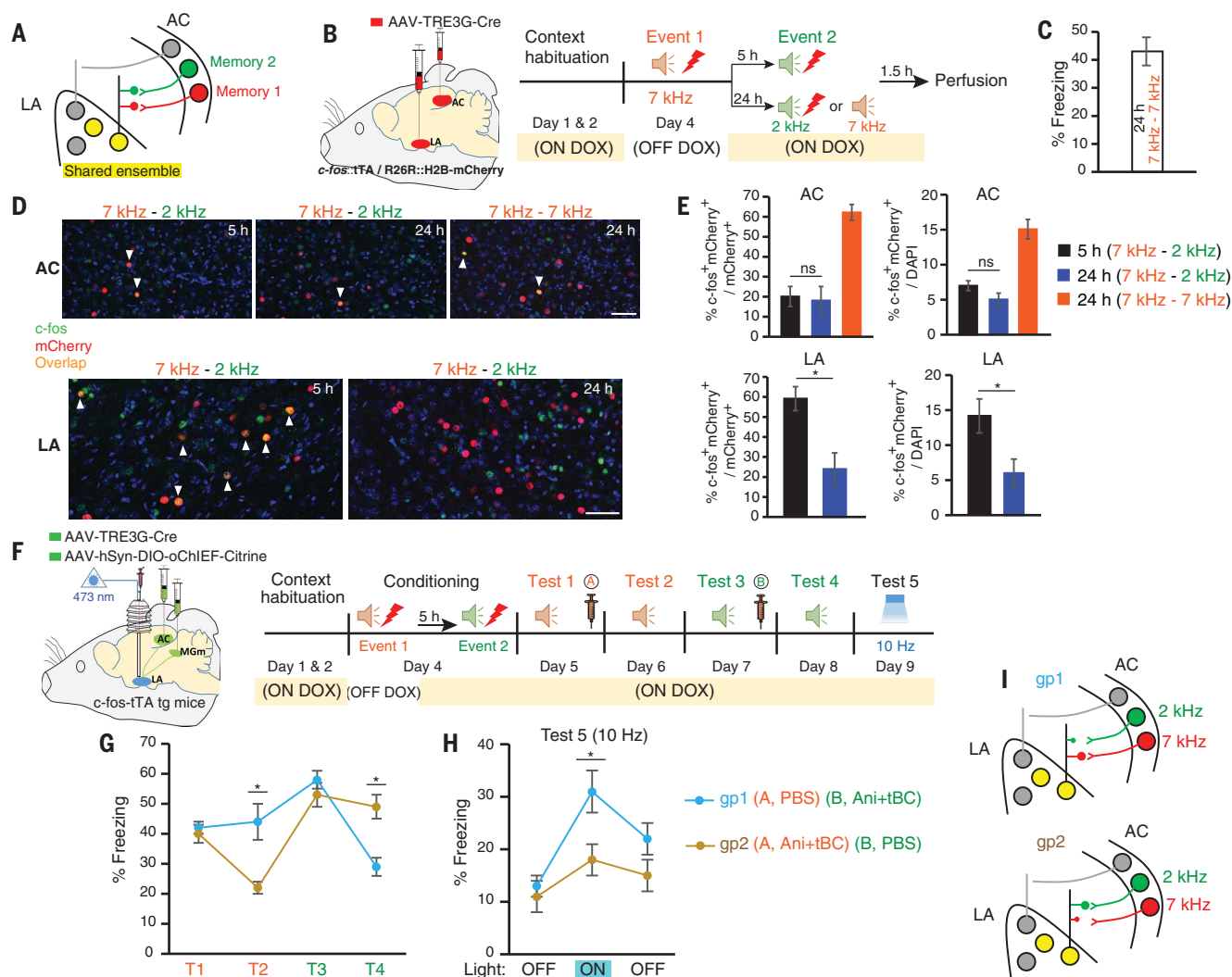
efficacy in synapses specific for event 1 memory (fig. S3, A and C, and Fig. 4, A and B). In comparison with a control group, mice that received LTD showed impairment in event 1 memory recall, but not in event 2 memory recall (Fig. 4C). Optogenetic stimulation to the terminals of the AC and MGm ensemble of event 1 memory triggered freezing in the control group, whereas it failed to trigger freezing in the LTD group, despite the fact that event 2 memory was intact (Fig. 4D). Thus, selective depotentiation of synaptic plasticity deconstructs the specific connectivity between engram assemblies, thereby erasing one memory without disrupting the other memory in the same population of neurons.

Last, a gain-of-function experiment was performed in which both memories were erased with Ani+tBC and then optical LTP was induced



**Fig. 2. Resetting of synaptic plasticity and functional connectivity between engram cell assemblies as neural correlates of complete amnesia.** (A) Left, labeling strategy. Right, experimental design for the LTP occlusion experiment. (B) Freezing level during test 1. (C) Average of in vivo field excitatory postsynaptic potential slope (normalized to baseline) before and after LTP induction (two-way repeated-measures ANOVA;  $n = 4$  mice per group). (D) Traces before (black) and after (red) optical LTP induction. (E) Left, labeling of engram cell assemblies in the AC, MGm,

and LA using double transgenic mice ( $c-Fos::tTA$ /R26R::H2B- $mCherry$ ) (18). Right, experimental design. (F) Freezing levels during tests 1 and 2 (one-way ANOVA). (G) Freezing levels during test 3 (one-way ANOVA). (H) Representative images showing  $c-Fos^+mCherry^+$  overlap in the LA, indicated by arrowheads. Blue, 4',6-diamidino-2-phenylindole (DAPI) staining. Scale bars, 50  $\mu m$ . (I)  $c-Fos^+mCherry^+$  overlap cell counts (one-way ANOVA;  $n = 4$  mice per group). Yellow lines represent chance level for each group. \* $P < 0.05$ ; \*\* $P < 0.01$ . Data are represented as mean  $\pm$  SEM.



**Fig. 3. Synapse-specific erasure of overlapping fear memories.** (A) Model for the neuronal ensemble in the LA and AC after two associative memories encoded with a 5-hour interval. Memories 1 and 2 respectively correspond to events 1 and 2. (B) Left, strategy to label engram cells in the AC and LA using double transgenic mice (*c-Fos::tTA/R26R::H2B-mCherry*) injected with AAV-TRE3G-Cre. Right, experimental design to check the overlapping ensembles between two associative memories that were encoded with different time intervals separating them. (C) Freezing level during the 7-kHz test session in the 7-kHz-7-kHz group. (D) Top, images for the overlapping ensembles, indicated by arrowheads, in the AC for different time intervals.

Bottom, same as top but in the LA. Blue, DAPI staining. Scale bars, 50  $\mu$ m. (E) Top, *c-Fos*<sup>+</sup>-*mCherry*<sup>+</sup> overlap cell counts in the AC (one-way ANOVA; *n* = 4 mice per group). Bottom, same as in top but in the LA (unpaired *t* test; *n* = 4 mice per group). (F) Design for the selective memory erasure experiment. (G and H) Freezing levels for gp1 and gp2 during 7- and 2-kHz tones before and after drug injection (G) and during light-off and light-on epochs (H) (unpaired *t* test; *n* = 10 mice per group). T1, test 1; T2, test 2; and so forth. (I) Model for selective erasure of either 7-kHz-tone fear memory (red) or 2-kHz-tone fear memory (green). Overlapping ensembles are in yellow. \**P* < 0.05. Data are represented as mean ± SEM.

in event 1 memory-specific synapses (Fig. 4, E and F). Mice that received the LTP protocol displayed higher freezing levels in response to the 7-kHz tone (test 5), whereas freezing responses to the 2-kHz tone (test 6) were unaffected (Fig. 4G).

Storing and distinguishing between several memories encoded in the same neurons are critically important for organizing unique memories. Our findings demonstrate that synapse-specific plasticity is necessary and sufficient for associative fear memory storage and that it guarantees uniqueness to the memory trace, pointing to plasticity as a substrate for the fear memory engram. This

perspective is consistent with a recent observation that LTP is selectively induced in specific auditory pathways after fear memory formation (20).

Engram cells retain a memory after anisomycin-induced amnesia, and synaptic plasticity is dispensable for memory storage (21). However, synaptic plasticity and functional connectivity between engram cell assemblies are indispensable for fear memory storage, because after LTD induction, the depressed synapses might be nonfunctional. Therefore, not only the natural cue, but also the optical stimulation of synapses between the engram cell assemblies failed to retrieve the memory. Furthermore, the engram network no longer

retained the associative fear memory after Ani + tBC-induced complete amnesia. The LTP occlusion experiment showed that synaptic potentiation persisted even 2 days after behavioral training in the PBS control group and that complete amnesia accompanied a reset of LTP. This further supports the idea that LTP is important for memory maintenance. The combined evidence suggests that synaptic plasticity can build a specific connectivity within the engram cell assemblies and that the functional connectivity is a simple reflection of the enhanced synaptic strength, rather than an independent mechanism for memory storage.



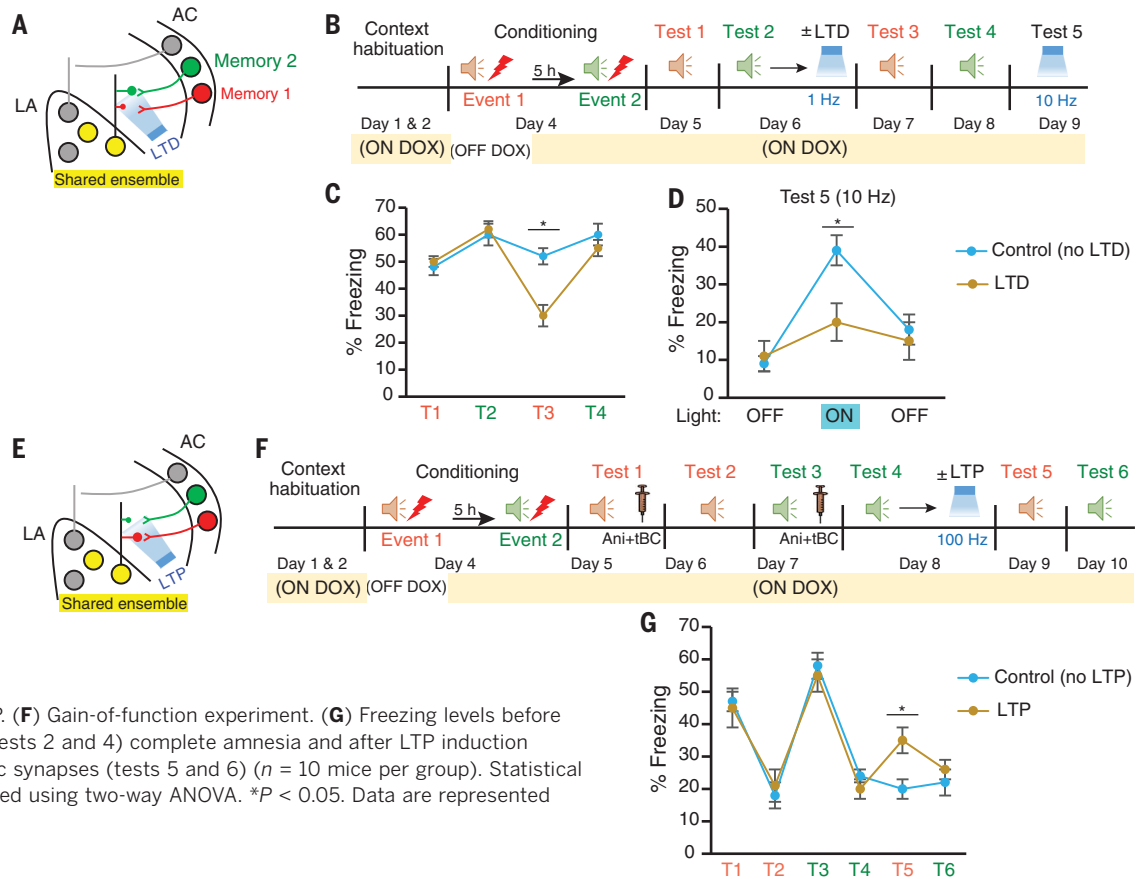
**Fig. 4. Engram-specific synaptic plasticity is crucial and sufficient for information storage and keeps the identity of the overlapping memories distinct.**

(A) Model for selective optogenetic targeting of synaptic plasticity with LTD. Memories 1 and 2 respectively correspond to events 1 and 2.

(B) Design of the loss-of-function experiment.

(C and D) Freezing levels in response to 7- and 2-kHz tones before and after optical LTD induction to event 1 memory-specific synapses (C) and in response to optical stimulation (D) ( $n = 10$  mice per group).

(E) Model for selective optogenetic targeting of synaptic plasticity with LTP. (F) Gain-of-function experiment. (G) Freezing levels before (tests 1 and 3) and after (tests 2 and 4) complete amnesia and after LTP induction to event 1 memory-specific synapses (tests 5 and 6) ( $n = 10$  mice per group). Statistical comparisons were performed using two-way ANOVA.  $*P < 0.05$ . Data are represented as mean  $\pm$  SEM.



This study uncovered the mechanism by which the brain can maintain the uniqueness of a massive number of associated memories stored in shared cell ensembles. Furthermore, we achieved selective and total erasure of a fear memory from an engram network without affecting other memories stored in the shared ensemble by resetting the plasticity in a synapse-specific manner. These findings lead to a better understanding of the mechanisms underlying memory storage and may give insight into therapeutic approaches to treating post-traumatic stress disorder.

#### REFERENCES AND NOTES

1. T. V. Bliss, G. L. Collingridge, *Nature* **361**, 31–39 (1993).
2. T. V. Bliss, T. Lomo, *J. Physiol.* **232**, 331–356 (1973).
3. M. Bocchio, S. Nabavi, M. Capogna, *Neuron* **94**, 731–743 (2017).
4. J. P. Johansen, C. K. Cain, L. E. Ostroff, J. E. LeDoux, *Cell* **147**, 509–524 (2011).
5. S. Nabavi *et al.*, *Nature* **511**, 348–352 (2014).
6. G. Neves, S. F. Cooke, T. V. Bliss, *Nat. Rev. Neurosci.* **9**, 65–75 (2008).
7. S. Tonegawa, M. Pignatelli, D. S. Roy, T. J. Ryan, *Curr. Opin. Neurobiol.* **35**, 101–109 (2015).
8. J. H. Han *et al.*, *Science* **323**, 1492–1496 (2009).
9. S. A. Josselyn, S. Köhler, P. W. Frankland, *Nat. Rev. Neurosci.* **16**, 521–534 (2015).
10. X. Liu *et al.*, *Nature* **484**, 381–385 (2012).

11. L. G. Reijmers, B. L. Perkins, N. Matsuo, M. Mayford, *Science* **317**, 1230–1233 (2007).
12. A. J. Silva, Y. Zhou, T. Rogerson, J. Shobe, J. Balaji, *Science* **326**, 391–395 (2009).
13. S. Tonegawa, X. Liu, S. Ramirez, R. Redondo, *Neuron* **87**, 918–931 (2015).
14. D. J. Cai *et al.*, *Nature* **534**, 115–118 (2016).
15. M. Nomoto *et al.*, *Nat. Commun.* **7**, 12319 (2016).
16. N. Ohkawa *et al.*, *Cell Rep.* **11**, 261–269 (2015).
17. A. J. Rashid *et al.*, *Science* **353**, 383–387 (2016).
18. J. Yokose *et al.*, *Science* **355**, 398–403 (2017).
19. T. Rogerson *et al.*, *Nat. Rev. Neurosci.* **15**, 157–169 (2014).
20. W. B. Kim, J.-H. Cho, *Neuron* **95**, 1129–1146.e5 (2017).
21. T. J. Ryan, D. S. Roy, M. Pignatelli, A. Arons, S. Tonegawa, *Science* **348**, 1007–1013 (2015).
22. P. Tovote, J. P. Fadok, A. Lüthi, *Nat. Rev. Neurosci.* **16**, 317–331 (2015).
23. M. Shehata, K. Inokuchi, *Rev. Neurosci.* **25**, 543–557 (2014).
24. M. Shehata, H. Matsumura, R. Okubo-Suzuki, N. Ohkawa, K. Inokuchi, *J. Neurosci.* **32**, 10413–10422 (2012).
25. M. Shehata *et al.*, *J. Neurosci.* **38**, 3809–3822 (2018).
26. C. C. Huang, C. C. Chen, Y. C. Liang, K. S. Hsu, *Int. J. Neuropsychopharmacol.* **17**, 1233–1242 (2014).
27. S. Park *et al.*, *Sci. Rep.* **6**, 31069 (2016).
28. E. Tsvetkov, W. A. Carlezon Jr., F. M. Benes, E. R. Kandel, V. Y. Bolshakov, *Neuron* **34**, 289–300 (2002).

#### ACKNOWLEDGMENTS

From the University of Toyama, we thank N. Ohkawa for his help in providing c-Fos:TA mice, Y. Saïtoh and M. Nomoto for their help with electrophysiology, and S. Tsujimura for maintenance of mice. We thank all members of the Inokuchi laboratory for discussion and

suggestions. We also thank M. Ito and N. Takino (Jichi Medical University, Japan) for their help with production of the AAV vectors.

**Funding:** This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas (“Memory dynamism”; JP25115002) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT); JSPS KAKENHI grant number 23220009; the Core Research for Evolutional Science and Technology (CREST) program (JPMJCR13W1) of the Japan Science and Technology Agency (JST); the Mitsubishi Foundation; the Uehara Memorial Foundation; and the Takeda Science Foundation (to K.I.). Additional support was provided by a Grant-in-Aid for young scientists from JSPS KAKENHI (grant number 25830007) to M.S. The Otsuka Toshimi Scholarship Foundation supported K.A. **Author contributions:** K.A., M.S., and K.I. designed the experiments. K.A., M.S., and K.I. wrote the manuscript. K.A., M.S., and K.C. performed the experiments. K.A., M.S., and K.I. analyzed the data. H.N. and M.M. produced and maintained transgenic mice. S.M. prepared AAVs. **Competing interests:** S.M. owns equity in a company, Gene Therapy Research Institution, that commercializes the use of AAV vectors for gene therapy applications. To the extent that the work in this manuscript increases the value of these commercial holdings, S.M. has a conflict of interest. **Data and materials availability:** All data are available in the main text or the supplementary materials.

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6394/1227/suppl/DC1  
Materials and Methods  
Figs. S1 to S7  
References (29–31)

21 February 2018; accepted 26 April 2018  
10.1126/science.aat3810

## HUMAN IMPACTS

# The influence of human disturbance on wildlife nocturnality

Kaitlyn M. Gaynor<sup>1\*</sup>, Cheryl E. Hojnowski<sup>1</sup>, Neil H. Carter<sup>2</sup>, Justin S. Brashares<sup>1</sup>

Rapid expansion of human activity has driven well-documented shifts in the spatial distribution of wildlife, but the cumulative effect of human disturbance on the temporal dynamics of animals has not been quantified. We examined anthropogenic effects on mammal diel activity patterns, conducting a meta-analysis of 76 studies of 62 species from six continents. Our global study revealed a strong effect of humans on daily patterns of wildlife activity. Animals increased their nocturnality by an average factor of 1.36 in response to human disturbance. This finding was consistent across continents, habitats, taxa, and human activities. As the global human footprint expands, temporal avoidance of humans may facilitate human-wildlife coexistence. However, such responses can result in marked shifts away from natural patterns of activity, with consequences for fitness, population persistence, community interactions, and evolution.

**T**he global expansion of human activity has had profound consequences for wildlife. Research has documented the effects of habitat destruction and defaunation on species and ecosystems (1), but the indirect or nonlethal pathways through which humans alter the natural world have largely eluded quantification. Human presence can instill strong fear in wild animals, which may adjust their activity to avoid contact with humans (2). As in natural predator-prey systems, such risk avoidance can have important nonlethal effects on animal physiology and fitness, affecting demography and triggering trophic cascades (3).

The study of fear effects on animals has focused mainly on spatial avoidance, propelled by rapid advances in wildlife tracking, remote sensing, and computational methods (2, 4). However, as the human footprint expands (5), fewer areas are available for animals to seek spatial refuge from people. In places where wild animals co-occur with humans, animals may minimize risk

by separating themselves in time rather than in space (6). Temporal partitioning is a common, even intrinsic, feature of ecological communities, shaping spatiotemporal patterns of predation and competition (6–8). Here we show that humans, as a diurnal apex “super predator” (9), are driving increases in nocturnal activity across diverse mammalian taxa.

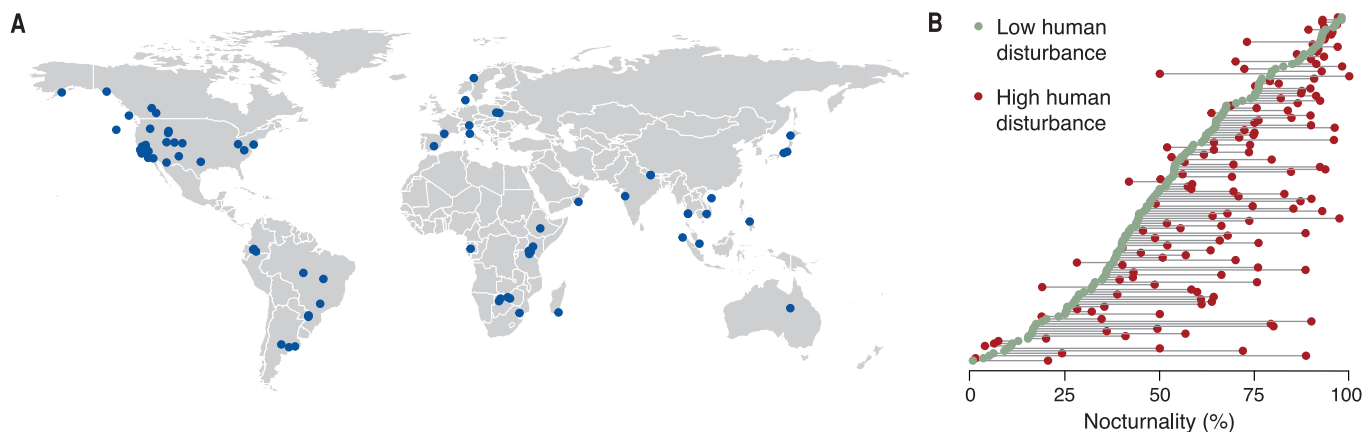
To quantify temporal shifts in wildlife activity in response to humans, we conducted a meta-analysis of published literature on the activity of mammals across gradients of human disturbance. Our dataset included 141 effect sizes for 62 mammal species, representing 21 families and nine orders, and spanned six continents (Fig. 1A). We restricted our analysis to medium- and large-bodied mammals [ $>1$  kg in body size (10)] because of their large space needs, potential for conflict with humans, and high levels of behavioral plasticity and because data on their 24-hour activity patterns were widely available. Within each study, we compared animal nocturnality

under conditions of low and high human disturbance. We classified areas, time periods, or individual animals as being associated with low or high disturbance on the basis of categorical descriptions of the study system or binned distance or elapsed time from an anthropogenic disturbance (tables S1 and S3).

For each species in each study, we calculated the risk ratio (RR) as a measure of effect size. We compared the percentage of activity that occurred at night (as measured by motion-activated cameras, telemetry devices, and direct observation) at sites or during seasons of high human disturbance ( $X_H$ ) with nighttime activity under low disturbance ( $X_L$ ), with  $RR = \ln(X_H/X_L)$ . A positive RR indicated a relatively greater degree of nocturnality in response to humans, and a negative RR indicated reduced nocturnality. We used meta-analytical random-effects models to estimate the overall effect of human disturbance on nocturnality and to compare responses across types of human disturbance, species traits, habitats, continents, and study methods. We also used multivariate models to explore the relative importance of these factors with an information-theoretic approach.

Our analysis revealed a marked increase in nocturnal activity. Overall, mammal nocturnality increased by a factor of 1.36 [95% confidence interval (CI), 1.23 to 1.51] in areas or time periods of high human disturbance relative to nocturnality under low-disturbance conditions. For example, an animal that typically split its activity evenly between the day and night would increase its proportion of nocturnal activity to 68% of total activity near human disturbance. Of the 141 effect sizes, 83% corresponded to an increase in nocturnality in response to humans (Fig. 1B). This finding

<sup>1</sup>Department of Environmental Science, Policy, and Management, University of California–Berkeley, Berkeley, CA 94720, USA. <sup>2</sup>Human-Environment Systems Research Center, Boise State University, Boise, ID 83725, USA.  
\*Corresponding author. Email: kgaynor@berkeley.edu

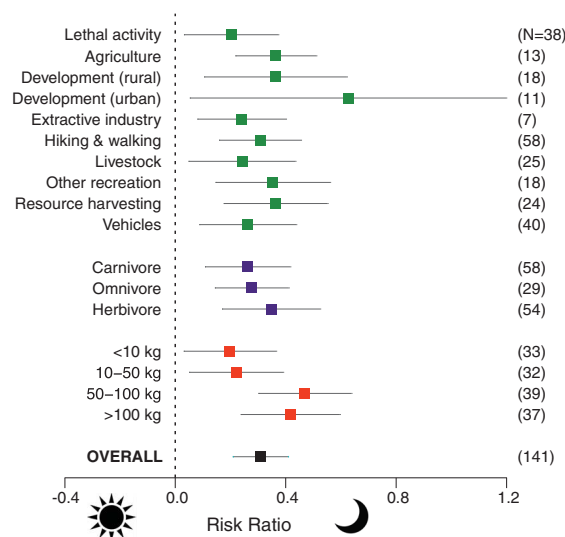


**Fig. 1. Mammals become more nocturnal to avoid humans throughout the world. (A)** Map illustrating the locations of the 76 studies included in the meta-analysis. **(B)** Paired measures of nocturnality (percentage of activity that occurs in the night) in areas of high human disturbance ( $X_H$ ) and

low human disturbance ( $X_L$ ), displayed for each species in each study ( $n = 141$  effect sizes, ordered from high to low  $X_L$ ). The relative change in nocturnality in response to human disturbance was used to calculate the effect size (RR) for the meta-analysis, where  $RR = \ln(X_H/X_L)$ .



**Fig. 2. Increase in large mammal nocturnality in relation to human activity types, trophic level, and body size.** Points represent the estimated overall effect size (with 95% CIs) for each category. Positive values indicate a relative increase in nocturnal activity in areas of higher human disturbance. The number of effect sizes in each category is indicated in parentheses (human activity categories were nonexclusive). Mammals exhibit a significant increase in nocturnal activity in response to all types of human activity, with similar patterns across trophic levels and body sizes. Back-transforming the overall (mean) RR (0.31; CI, 0.21 to 0.41) indicates that nocturnality increased by a factor of 1.36 (CI, 1.23 to 1.51) in response to human activity.



indicates a widespread increase in nocturnality among mammals living alongside people.

Our analysis spanned a wide range of human disturbances associated with diverse stimuli representing different levels of risk to wildlife, including lethal activities (e.g., hunting and retaliatory persecution), nonlethal activities (e.g., hiking and natural resource extraction), and human infrastructure (e.g., urban development, road construction, and agriculture). There was a significant increase in nocturnality in response to all forms of human presence (Fig. 2), signaling the robustness of our findings. Surprisingly, nonlethal human activities generated shifts in wildlife diel patterns similar to those from lethal activities (Fig. 2), suggesting that animals perceive and respond to humans as threats even when they pose no direct risk.

We expected temporal responses of wildlife to vary across species, given interspecific differences in biology such as variation in morphology and behavior. Body sizes of the 63 species analyzed ranged from 1.13 kg for the common opossum (*Didelphis marsupialis*) to >3,500 kg for the African elephant (*Loxodonta africana*). Mammals of all body size classes showed a strong response to human activity, although there was a slight trend toward a greater response among larger-bodied species (Fig. 2), perhaps because they are more likely to be hunted and harassed or because their space needs force them into more frequent contact with people. Across trophic levels, species also exhibited similar responses to human activity (Fig. 2). Even apex carnivores, which from an evolutionary perspective have typically faced little or no predation risk from other species, responded to humans by becoming more active at night. Species that are typically diurnal showed an increase in nocturnality, and even crepuscular and nocturnal species became more strongly nocturnal around humans (fig. S3A). Human activity increased wildlife nocturnality across continents and habitat types (fig. S3, B and C). The study method (camera trap, teleme-

try, or direct observation), human disturbance treatment (space or time), and continent were included as predictors of effect size in the best multivariate models (tables S5 and S6).

The absence of increased nocturnal activity by wildlife in many of the studies examined does not necessarily indicate an absence of human impact. Differences in age, sex, reproductive status, and personality may shape responses and mask patterns at the population level (11). For example, in one population of brown bears (*Ursus arctos*), human recreation induced temporal shifts among adult males that created additional daytime feeding opportunities for females, which are often otherwise outcompeted by males for access to resources (12). Alternatively, ecological and morphological constraints may limit behavioral plasticity, causing individuals to remain active during the daytime in the presence of humans, thereby incurring the cost of increased stress or energetically expensive antipredator behaviors (13). Animals living alongside humans in disturbed ecosystems may face additional constraints (e.g., limited food resources) that limit their ability to respond temporally (14). At the community level, strictly diurnal species may also entirely avoid areas of human activity, opening niches for more nocturnal competitors, including invasive species (15).

In addition to changing diel patterns of activity in response to human disturbance, wildlife species may alter temporal patterns of specific behaviors. During the day, animals often choose more protected habitats or microhabitats in areas of human disturbance (16) or more strongly avoid anthropogenic features like roads and buildings (17, 18). Some species also modify daytime and nighttime movement speed and tortuosity (19) and temporal patterns of antipredator behavior such as vigilance (20). Furthermore, in addition to shifting activity from the day to the night, animals often decrease their overall activity throughout the 24-hour period in response to human disturbance, spending more time rest-

ing and less time foraging or engaging in other fitness-enhancing behaviors (21).

We assert that fear of humans is the primary mechanism driving the increase in wildlife nocturnality, given its prevalence across activity types and the widespread evidence that mammals perceive and respond to risk from people (2, 4). In some of the studies included in our analysis, fear of humans may have interacted with other factors, such as food provisioning (through anthropogenic food sources such as livestock, crops, and food waste), to drive increased nocturnal activity (rather than spatial avoidance) by generalist species in areas of human disturbance (22, 23). Furthermore, nighttime light cues or increased visibility around permanent human infrastructure may also promote an increase in nocturnal activity in these areas (24). However, for animals wary of humans or more fearful of predators in lit areas, anthropogenic light may also be perceived as a source of risk and thus may limit the magnitude of a shift to nocturnality (25).

The global increase of nocturnality among wildlife in human-dominated areas demonstrates the high degree of behavioral plasticity of animals in a human-altered world, with great implications for ecology and conservation (Fig. 3). On the positive side, temporal partitioning may facilitate human-wildlife coexistence at fine spatial scales and effectively increase available habitats for species that are able to adjust (26). The separation of humans and wildlife in time, if not space, may also limit contact rates between people and dangerous animals and therefore reduce some forms of negative encounters between the two, such as disease transmission and attacks on people. In situations where humans pose a lethal threat to wildlife, increased nocturnality may be advantageous to individual animals and has been linked to increased probability of survival (27). In this case, increased mortality among more diurnal individuals may even drive selective pressure for behavioral plasticity and nocturnal activity (27).

Though human-wildlife coexistence may be a positive outcome of increased nocturnal activity of wildlife, this shift may also have negative and far-reaching ecological consequences. Humans may impose substantial fitness costs on individual animals, analogous to predation risk effects in predator-prey systems, in which costly antipredator behavior compromises prey reproduction and survival and alters trophic interactions (28). An increase in nocturnality may also eventually alter evolution through selection for morphological, physiological, and behavioral adaptations to nighttime activity. The human "super predator" has already been implicated in evolutionary changes through selective harvest (29), but as with other predators (3), the nonlethal effects of humans may have an even stronger influence on fitness and evolutionary trajectories.

Risk effects induced by a temporal response to human presence are expected to be particularly strong, as an increase in nocturnal activity can cause mismatches between morphology and environment for historically diurnal species. The

diel cycle provides a reliable set of environmental cues against which ecological and evolutionary processes play out (30). Behavior at different times of the diel cycle influences and is influenced by morphology [e.g., corneal size (37)], physiology

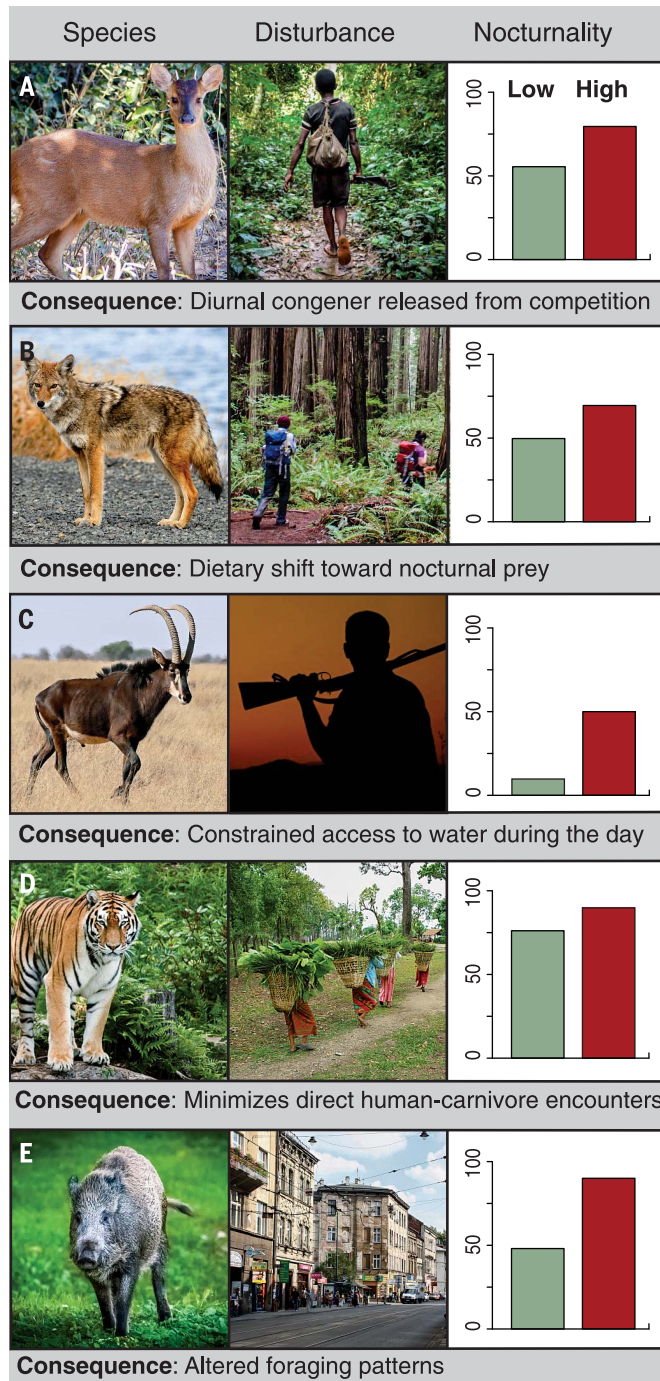
[e.g., opsin proteins (32)], and ecology [e.g., group living and predation risk (33)]. Although most mammals possess some sensory adaptations to nighttime activity due to nocturnal mammalian ancestors, many species have evolved traits that

optimize diurnal behavior (34, 35). When active at other times, diurnally adapted animals may suffer from reduced hunting and foraging efficiency, weakened antipredator strategies, disruption of social behavior, poor navigational capacity, and higher metabolic costs, all of which can compromise reproduction and survival (36, 37).

By altering typical activity patterns in some wildlife species, human disturbance initiates behaviorally mediated trophic cascades and transforms entire ecological communities (38, 39). Fear-based behavioral responses by apex predators to humans may diminish their ability to hunt and thus perform their ecological role at the top of a trophic web (40). Animals that are increasingly nocturnal may drastically alter their diets toward prey or forage that are more accessible at night, reshaping lower trophic levels (41). Predators may also abandon kills near human settlements in the daytime, resulting in increased overall predation rates (42). Human-induced increases in nocturnality among prey species can also increase their vulnerability to nocturnal predators (43). Differential responses to human disturbance among mammal species also alters patterns of predation and competition. Predators may increase nocturnality in response to humans, creating temporal human shields, in which a prey species will then decrease nocturnality near human disturbance to avoid predation, analogous to spatial human shields (44, 45). Alternatively, some prey species may instead seek out human-dominated areas at night to escape nocturnal predators that spatially avoid humans (46).

Human-induced change in diel activity is a growing field of inquiry, as indicated by the large number of studies in our meta-analysis (table S1). However, very few studies have examined the individual-, population-, or community-level consequences of these behavioral changes. Given the widespread nature of increased nighttime activity, there is ample opportunity and need to study not just the magnitude of this effect but also its consequences for individual fitness, species interactions, and natural selection. Additional research and synthetic analyses are also needed for nonmammalian taxa, which may also exhibit diel shifts in response to humans (47).

As research on the pattern and consequences of increased nocturnality advances, we urge that similar progress be made to incorporate knowledge of temporal dynamics into conservation planning. Currently, spatial ecology informs commonly used land-planning tools (48), but new tools are needed that explicitly address temporal interactions. Approaches may include diurnal "temporal zoning," analogous to spatial zoning, that would restrict certain human activities during times of the day when species of conservation concern are most active or when the likelihood of negative human-wildlife encounters is highest. Similar strategies already restrict human activity at certain times of the year, such as during breeding seasons (49). Systematic approaches to understanding and managing temporal interactions between humans



**Fig. 3. Case studies demonstrate the diverse consequences of human-induced increases in nocturnality.** (A) Red brocket deer (*Mazama americana*) and subsistence hunting, Atlantic Forest, Argentina (50). (B) Coyote (*Canis latrans*) and hiking, Santa Cruz Mountains, CA (51, 52). (C) Sable antelope (*Hippotragus niger*) and sport hunting, Hwange National Park, Zimbabwe (53). (D) Tiger (*Panthera tigris*) and forest product collection and farming, Chitwan National Park, Nepal (26). (E) Wild boar (*Sus scrofa*) and urban development, Cracow and Białowieża Forest, Poland (54). Green bars represent nocturnality (the percentage of total activity that occurs in the night) in areas of low human disturbance ( $X_L$ ), and red bars represent nocturnality in areas of high human disturbance ( $X_H$ ).



and wildlife can open up new domains for conservation in an increasingly crowded world.

## REFERENCES AND NOTES

1. R. Dirzo *et al.*, *Science* **345**, 401–406 (2014).
2. A. Frid, L. M. Dill, *Conserv. Ecol.* **6**, 11 (2002).
3. E. L. Preisser, D. I. Bolnick, M. E. Benard, *Ecology* **86**, 501–509 (2005).
4. M. A. Tucker *et al.*, *Science* **359**, 466–469 (2018).
5. O. Venter *et al.*, *Nat. Commun.* **7**, 12558 (2016).
6. N. Kronfeld-Schor, T. Dayan, *Annu. Rev. Ecol. Evol. Syst.* **34**, 153–181 (2003).
7. J. H. Carothers, F. M. Jakšić, *Oikos* **42**, 403–406 (1984).
8. D. B. Lesmeister, C. K. Nielsen, E. M. Schaubert, E. C. Hellgren, *Wildl. Monogr.* **191**, 1–61 (2015).
9. M. Clinchy *et al.*, *Behav. Ecol.* **27**, 1826–1832 (2016).
10. K. E. Jones *et al.*, *Ecology* **90**, 2648 (2009).
11. A. G. Hertel, J. E. Swenson, R. Bischof, *Behav. Ecol.* **28**, 1524–1531 (2017).
12. O. T. Nevin, B. K. Gilbert, *Biol. Conserv.* **121**, 611–622 (2005).
13. J. A. Gill, K. Norris, W. J. Sutherland, *Biol. Conserv.* **97**, 265–268 (2001).
14. J. Martin *et al.*, *Can. J. Zool.* **88**, 875–883 (2010).
15. C. M. Huijbers, T. A. Schlacher, D. S. Schoeman, M. A. Weston, R. M. Connolly, *Landsc. Urban Plan.* **119**, 1–8 (2013).
16. C. Dupke *et al.*, *Ecography* **40**, 1014–1027 (2016).
17. C. D. Morrison, M. S. Boyce, S. E. Nielsen, M. M. Bacon, *J. Wildl. Manage.* **78**, 1394–1403 (2014).
18. J. A. Stabach, G. Wittemyer, R. B. Boone, R. S. Reid, J. S. Worden, *Ecosphere* **7**, e01428 (2016).
19. A. Oriol-Cotterill, D. W. Macdonald, M. Valeix, S. Ekwanga, L. G. Frank, *Anim. Behav.* **101**, 27–39 (2015).
20. L. Sönninghansen *et al.*, *Ethology* **119**, 233–243 (2013).
21. N. van Doormaal, H. Ohashi, S. Koike, K. Kaji, *Eur. J. Wildl. Res.* **61**, 517–527 (2015).
22. J. P. Beckmann, J. Berger, *J. Zool.* **261**, 207–212 (2003).
23. M. Valeix, G. Hemson, A. J. Loveridge, G. Mills, D. W. Macdonald, *J. Appl. Ecol.* **49**, 73–81 (2012).
24. D. M. Dominoni, J. C. Borniger, R. J. Nelson, *Biol. Lett.* **12**, 20160015 (2016).
25. T. Longcore, C. Rich, *Front. Ecol. Environ.* **2**, 191–198 (2004).
26. N. H. Carter, B. K. Shrestha, J. B. Karki, N. M. B. Pradhan, J. Liu, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15360–15365 (2012).
27. M. H. Murray, C. C. St. Clair, *Behav. Ecol.* **26**, 1520–1527 (2015).
28. S. Creel, D. Christianson, *Trends Ecol. Evol.* **23**, 194–201 (2008).
29. C. T. Darimont *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 952–954 (2009).
30. K. J. Gaston, J. Bennie, T. W. Davies, J. Hopkins, *Biol. Rev. Cambridge Philos. Soc.* **88**, 912–927 (2013).
31. L. Schmitz, R. Motani, *Vision Res.* **50**, 936–946 (2010).
32. H. Zhao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8980–8985 (2009).
33. T. Stankowich, P. J. Haverkamp, T. Caro, *Evolution* **68**, 1415–1425 (2014).
34. R. Maor, T. Dayan, H. Ferguson-Gow, K. E. Jones, *Nat. Ecol. Evol.* **1**, 1889–1895 (2017).
35. M. I. Hall, J. M. Kamilar, E. C. Kirk, *Proc. R. Soc. London Ser. B* **279**, 4962–4968 (2012).
36. A. Sih, *Anim. Behav.* **85**, 1077–1088 (2013).
37. U. Tuomainen, U. Candolin, *Biol. Rev. Cambridge Philos. Soc.* **86**, 640–657 (2011).
38. E. E. Werner, S. D. Peacor, *Ecology* **84**, 1083–1100 (2003).
39. O. J. Schmitz, V. Krivan, O. Ovadia, *Ecol. Lett.* **7**, 153–163 (2004).
40. A. S. Ordiz, R. Bischof, J. E. Swenson, *Biol. Conserv.* **168**, 128–133 (2013).
41. A. Ordiz, S. Sæbo, J. Kindberg, J. E. Swenson, O. G. Støen, *Anim. Conserv.* **20**, 51–60 (2017).
42. J. A. Smith, Y. Wang, C. C. Wilmers, *Proc. R. Soc. London Ser. B* **282**, 20142711 (2015).
43. J. C. Kilgo, R. F. Labisky, D. E. Fritzen, *Conserv. Biol.* **12**, 1359–1364 (1998).
44. J. Berger, *Biol. Lett.* **3**, 620–623 (2007).
45. T. B. Muhly, C. Semeniuk, A. Massolo, L. Hickman, M. Musiani, *PLOS ONE* **6**, e17050 (2011).
46. A. Atickem, L. E. Loe, N. C. Stenseth, *Ethology* **120**, 715–725 (2014).
47. J. Burger, M. Gochfeld, *Condor* **93**, 259–265 (1991).
48. R. L. Pressey, M. Cabeza, M. E. Watts, R. M. Cowling, K. A. Wilson, *Trends Ecol. Evol.* **22**, 583–592 (2007).
49. C. L. Larson, S. E. Reed, A. M. Merenlender, K. R. Crooks, *PLOS ONE* **11**, e0167259 (2016).
50. M. S. Di Bitetti, A. Paviolo, C. A. Ferrari, C. De Angelo, Y. Di Blanco, *Biotropica* **40**, 636–645 (2008).
51. Y. Wang, M. L. Allen, C. C. Wilmers, *Biol. Conserv.* **190**, 23–33 (2015).
52. J. A. Smith, A. C. Thomas, T. Levi, Y. Wang, C. C. Wilmers, *Oikos* (2018).
53. W. G. Crosmarty, M. Valeix, H. Fritz, H. Madzikanda, S. D. Côté, *Anim. Behav.* **83**, 145–153 (2012).
54. T. Podgórski *et al.*, *J. Mammal.* **94**, 109–119 (2013).

## ACKNOWLEDGMENTS

We thank the authors of all studies included in this meta-analysis, especially those who provided additional data: A. Baker, P. Cruz, F. Dalerum, P. Diaz-Ruiz, T. Lynam, M. Murphy-Mariscal, D. Ngoprasert, J. Nix, V. Oberosler, M. Reilly, L. Rich, Y. Wang, R. Wheat, and T. Wronski. Special thanks to C. Burton, T. Forrester, W. McShea, R. Kays, R. Steenweg, and J. Whittington for thoughts on an earlier version of this study and to N. Schramm for data-processing assistance. E. A. Lacey, A. D. Middleton, M. E. Power, the Brashares Group at UC Berkeley, and three anonymous reviewers provided helpful feedback and edits. **Funding:** K.M.G. and C.E.H. were funded by the NSF-GRFP; N.H.C. was supported by IIA-1301792 from the NSF Idaho EPSCoR Program; and J.S.B. was funded in part by NSF-CNH 115057. **Author contributions:** All authors designed the research. K.M.G. analyzed the data and drafted the manuscript. All authors contributed comments and edits to the final paper. **Competing interests:** The authors have no competing interests to declare. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6394/1232/suppl/DC1  
Materials and Methods  
Figs. S1 to S4  
Tables S1 to S7  
References (55–130)

21 December 2017; accepted 7 May 2018  
10.1126/science.aar7121

# Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle

Lorenzo Caputi<sup>1</sup>, Jakob Franke<sup>1\*</sup>, Scott C. Farrow<sup>1</sup>, Khoa Chung<sup>1</sup>, Richard M. E. Payne<sup>1†</sup>, Trinh-Don Nguyen<sup>1</sup>, Thu-Thuy T. Dang<sup>1</sup>, Inês Soares Teto Carqueijeiro<sup>2</sup>, Konstantinos Koudounas<sup>2</sup>, Thomas Dugé de Bernonville<sup>2</sup>, Belinda Ameyaw<sup>1‡</sup>, D. Marc Jones<sup>1</sup>, Ivo Jose Curcino Vieira<sup>3</sup>, Vincent Courdavault<sup>2§</sup>, Sarah E. O'Connor<sup>1§</sup>

Vinblastine, a potent anticancer drug, is produced by *Catharanthus roseus* (Madagascar periwinkle) in small quantities, and heterologous reconstitution of vinblastine biosynthesis could provide an additional source of this drug. However, the chemistry underlying vinblastine synthesis makes identification of the biosynthetic genes challenging. Here we identify the two missing enzymes necessary for vinblastine biosynthesis in this plant: an oxidase and a reductase that isomerize stemmadenine acetate into dihydroprecondylocarpine acetate, which is then deacetoxyated and cyclized to either catharanthine or tabersonine via two hydrolases characterized herein. The pathways show how plants create chemical diversity and also enable development of heterologous platforms for generation of stemmadenine-derived bioactive compounds.

**A**nticancer drugs vincristine **5** and vinblastine **6** were discovered 60 years ago in *Catharanthus roseus* (Madagascar periwinkle). These compounds have been used for the treatment of several types of cancer, including Hodgkin's lymphoma, as well as lung and brain cancers. Much of the metabolic pathway (31 steps from geranyl pyrophosphate to vinblastine) has been elucidated (1–3). Here we report the genes encoding the missing enzymes that complete the vinblastine pathway. Two redox enzymes convert stemmadenine acetate **7** into an unstable molecule, likely dihydroprecondylocarpine acetate **11**, which is then desacetoxyated by one of two hydrolases to generate, through Diels-Alder cyclizations, either tabersonine **2** or catharanthine **3** scaffolds that are ultimately dimerized to yield vinblastine and vincristine (Fig. 1A). In addition to serving as the precursors for vincristine **5** and vinblastine **6**, tabersonine **2** and catharanthine **3** are also precursors for other biologically active alkaloids (4, 5) (Fig. 1B). The discovery of these two redox enzymes (precondylocarpine acetate synthase and dihydroprecondylocarpine acetate synthase), along with the characterization of two hydrolases (tabersonine

and catharanthine synthase), provides insight into the mechanisms that plants use to create chemical diversity and also enables production of a variety of high-value alkaloids.

Catharanthine **3** (iboga-type alkaloid) and tabersonine **2** (aspidosperma-type) scaffolds are likely generated by dehydration of the biosynthetic intermediate stemmadenine **1** to dehydrosecodine **9**, which can then cyclize to either catharanthine **3** or tabersonine **2** via a net [4+2] cycloaddition reaction (Fig. 2) (6–9). We speculated that the missing components were an enzyme with dehydration and cyclization function. We hypothesized that the unstable nature of the dehydration product dehydrosecodine **9** (7) would preclude its diffusion out of the enzyme active site, and we thus searched for an enzyme that could catalyze both dehydration and cyclization reactions.

Because the biosynthetic genes for vincristine **5** and vinblastine **6** are not clustered in the plant genome (10), we searched for gene candidates in RNA sequencing (RNA-seq) data (10) from the vincristine- and vinblastine-producing plant *C. roseus*. Two genes annotated as alpha/beta hydrolases were identified by a shared expression profile with previously identified vinblastine biosynthetic enzymes (fig. S1A and data S1). A dehydratase could facilitate the isomerization of the 19,20-exo-cyclic double bond of stemmadenine **1** to form *iso*-stemmadenine **8** (Fig. 2), which would then allow dehydration to form dehydrosecodine **9** and, consequently, catharanthine and tabersonine (8). Virus-induced gene silencing (VIGS) of herein-named tabersonine synthase (TS) and catharanthine synthase (CS) (Fig. 1 and figs. S2 and S3) in *C. roseus* resulted in a marked reduction of tabersonine **2** ( $P = 0.0048$ ) and catharanthine ( $P = 0.01$ ), respectively. These silencing experiments implicate CS and TS in

catharanthine **3** and tabersonine **2** biosynthesis in *C. roseus*. However, when CS and TS were heterologously expressed in *Escherichia coli* (fig. S4A) and tested for reactivity with stemmadenine **1** (fig. S5) or the acetylated form of stemmadenine **7** (fig. S6), in which spontaneous deformylation would be hindered (Fig. 2, figs. S5 and S6, and tables S4 and S5), no reaction was observed. Although TS and CS are known to be implicated in vinblastine biosynthesis (11), the substrates, and therefore the specific catalytic functions, remained elusive.

We used enzyme-assay guided fractionation in an attempt to isolate the active substrate for the TS and CS enzymes from various aspidosperma-alkaloid- and iboga-alkaloid-producing plants. We focused on *Tabernaemontana* plants that are known to accumulate more stemmadenine **1** intermediate relative to downstream alkaloids (12). These experiments demonstrated that TS and CS were always active with the same fractions (fig. S7), consistent with previous hypotheses (6) that both enzymes use the same substrate. However, attempts to structurally characterize the substrate were complicated by its rapid decomposition, and the deformylated product tubotaiwine **12** [previously synthesized in reference (13)] was the major compound detected in the isolated mixture [Fig. 2 and nuclear magnetic resonance (NMR) data in fig. S8]. Given the propensity for deformylation in these structural systems (14), we rationalized that tubotaiwine **12** could be the decomposition product of the actual substrate, which would correspond to *iso*-stemmadenine **8** (dihydroprecondylocarpine) or its protected form (dihydroprecondylocarpine acetate **11**) (Fig. 2). We surmised that a coupled oxidation-reduction cascade could perform a net isomerization to generate dihydroprecondylocarpine **8** (or dihydroprecondylocarpine acetate **11**) from stemmadenine **1** (or stemmadenine acetate **7**). This idea was initially proposed by Scott and Wei, who indicated that stemmadenine acetate **7** can be oxidized to precondylocarpine acetate **10**, after which the 19,20-double bond can then be reduced to form dihydroprecondylocarpine acetate **11**, which can then form traces of tabersonine **2** upon thermolysis (15). Similar reactions with stemmadenine **1** resulted in deformylation to form condylocarpine **13** (16). Therefore, we reexamined the RNA-seq dataset for two redox enzymes that could convert stemmadenine acetate **7** to dihydroprecondylocarpine acetate **11**.

We noted a gene annotated as reticuline oxidase that had low absolute expression levels but a similar tissue expression pattern to that of the TS gene (fig. S1B). The chemistry of reticuline oxidase enzymes (17) such as berberine bridge enzyme and dihydrobenzophenanthridine oxidase suggests that these enzymes are capable of C–N bond oxidation, which would be required in this reaction sequence (Fig. 2) (17). When this oxidase gene was silenced in *C. roseus*, a compound with a mass and <sup>1</sup>H NMR spectrum corresponding to semisynthetically prepared stemmadenine acetate **7** (the proposed oxidase substrate) accumulated, suggesting that this gene encoded the

<sup>1</sup>Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. <sup>2</sup>Université de Tours, EA2106 Biomolécules et Biotechnologies

Végétales, Parc de Grandmont 37200 Tours, France. <sup>3</sup>Laboratório de Ciências Químicas-UENF-Campus dos

Goytacazes-RJ, 28013-602, Brazil.

\*Present address: Centre of Biomolecular Drug Research, Leibniz University Hannover, Schneiderberg 38, 30167 Hannover, Germany.

†Present address: Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.

‡Present address: Department of Biology, University of Lancaster, Bailrigg, Lancaster LA1 4YW, UK.

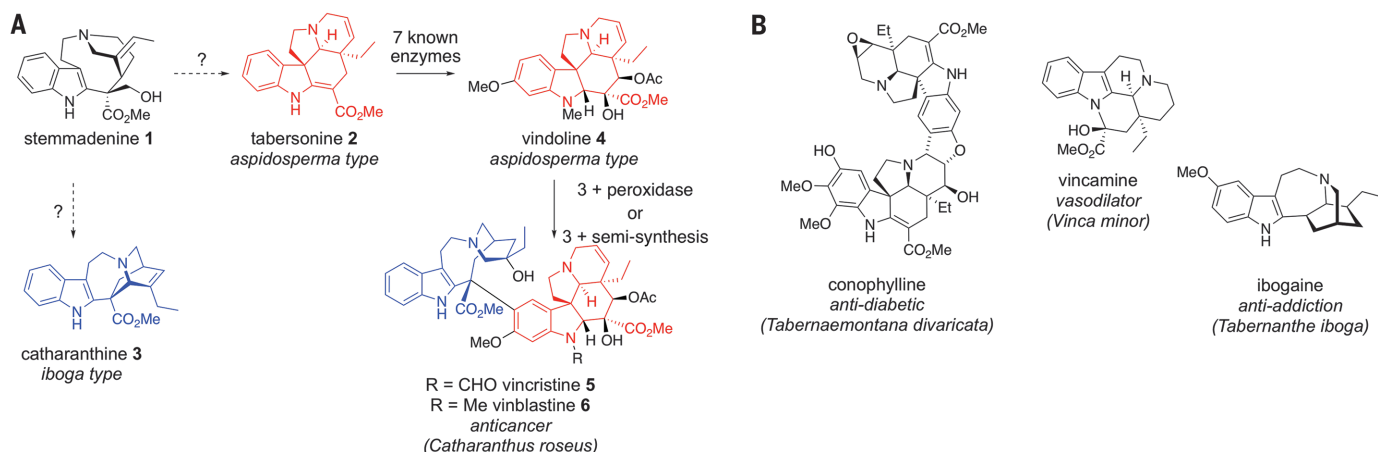
§Corresponding author. Email: sarah.oconnor@jic.ac.uk (S.E.O.); vincent.courdavault@univ-tours.fr (V.C.)



correct oxidase. We named this enzyme precondylocarpine acetate synthase (PAS) (figs. S9 to S11). Similarly, silencing of a medium-chain alcohol dehydrogenase, as part of an ongoing screen of alcohol dehydrogenases in *C. roseus* (14, 18, 19),

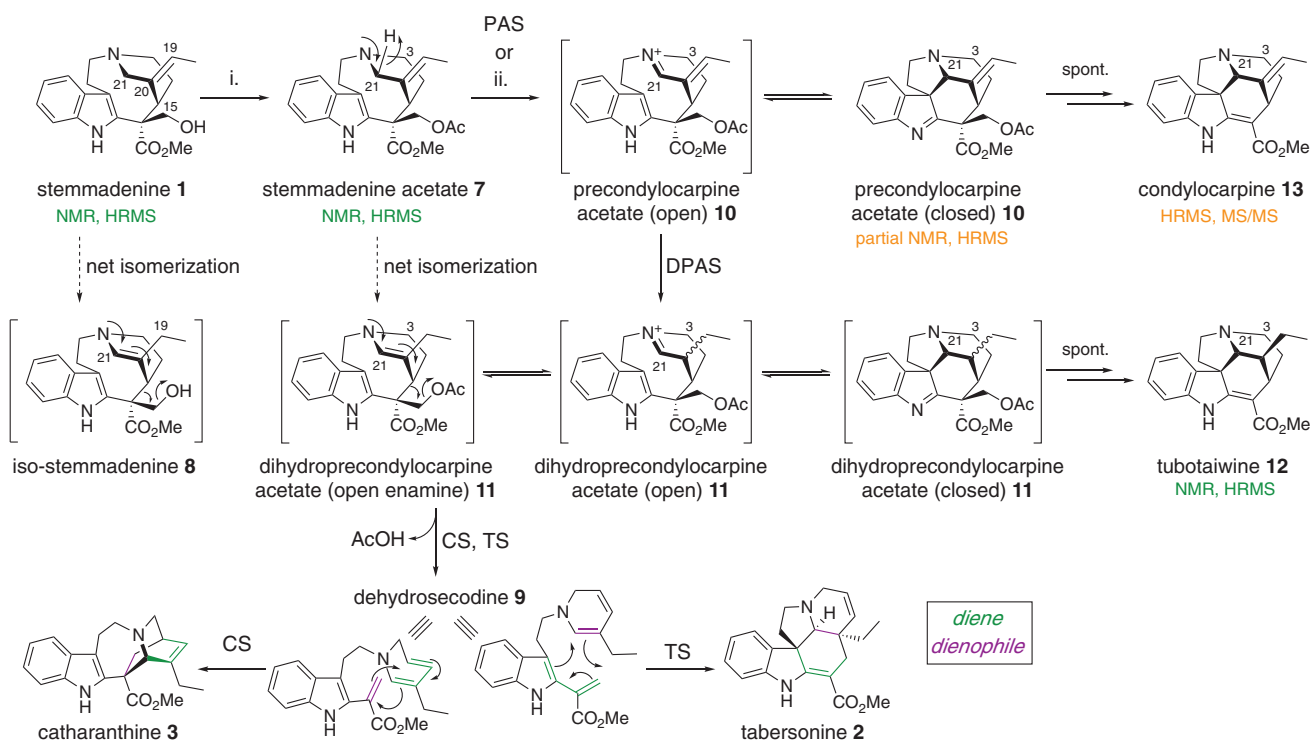
resulted in accumulation of a compound with a mass, retention time, and fragmentation pattern consistent with a partially characterized synthetic standard of precondylocarpine acetate **10** (the proposed substrate of the reductase) (figs. S12

to S14). This standard could be synthesized from stemmadenine acetate **7** using Pt and O<sub>2</sub> by established methods (8, 15, 20). With our small-scale reactions, yields were low and variable, and the product decomposed during characterization.



**Fig. 1. Vincristine and vinblastine biosynthesis.** (A) Vincristine **5** and vinblastine **6** are formed by dimerization from the monomers catharanthine **3** and vindoline **4** by a peroxidase (28) or chemical methods (29). The

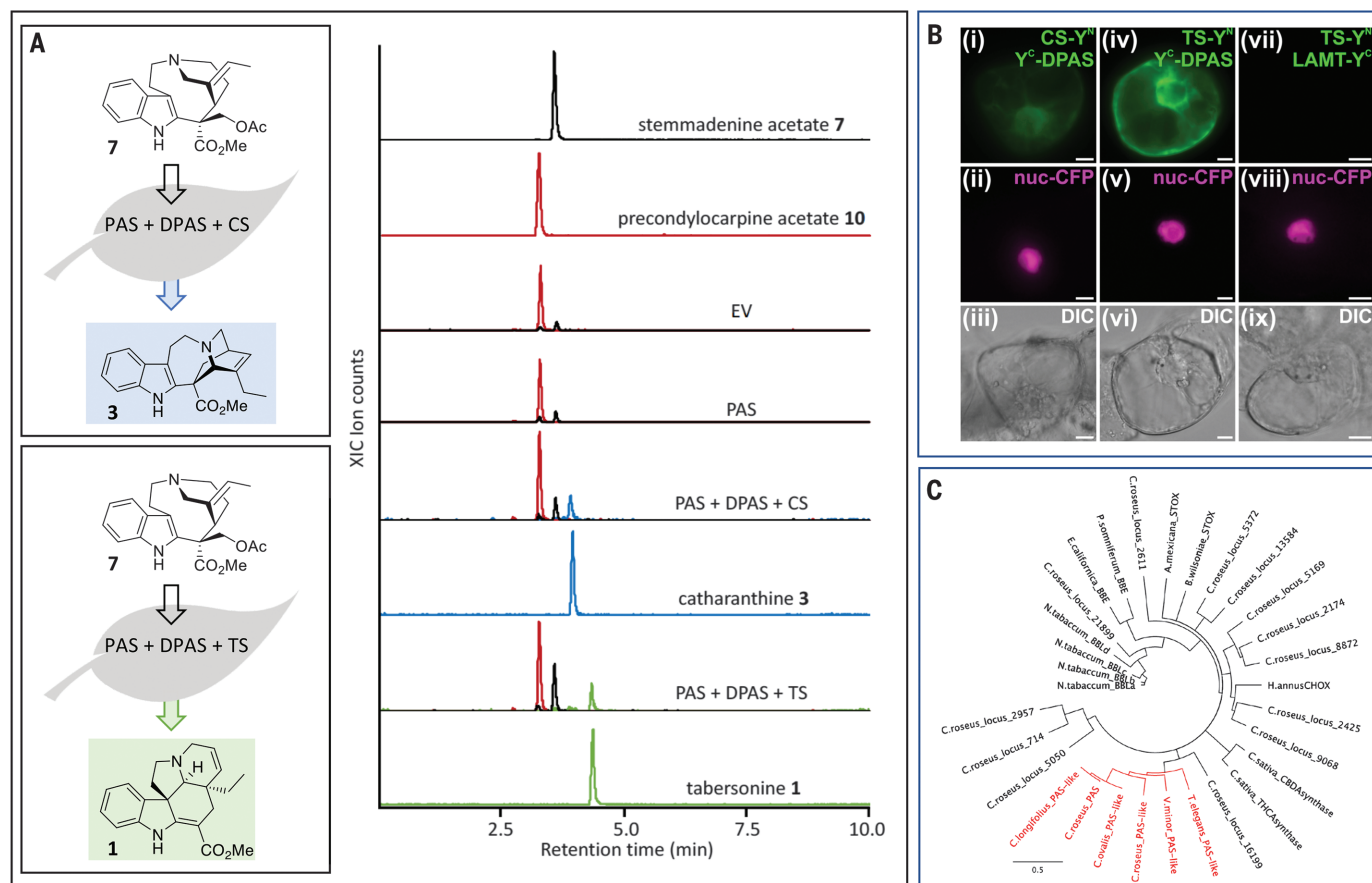
genes that convert tabersonine **2** to vindoline **4** have been identified (24). (B) Representative bioactive alkaloids derived from stemmadenine. Me, methyl; Et, ethyl.



**Fig. 2. Biosynthesis of catharanthine and tabersonine scaffolds.**

Stemmadenine acetate **7** {generated from stemmadenine **1** under condition i [Ac<sub>2</sub>O (excess) and pyridine (excess) at room temperature (r.t.), 4 hours, >99% yield]} undergoes an oxidation to form precondylocarpine acetate **10**. This reaction is catalyzed enzymatically by the reticuline oxidase homolog PAS or, alternatively, can be generated synthetically under condition ii [Pt (from 7.5 equivalents of PtO<sub>2</sub>), EtOAc, O<sub>2</sub> atmosphere, r.t., 10 hours, yields varied], as reported by Scott and Wei (8). Next, the open form of

precondylocarpine acetate **10** is reduced by the alcohol dehydrogenase DPAS. This reduced intermediate could not be isolated due to its lability but, on the basis of degradation product tubotaiwine **12**, is assumed to be dihydroprecondylocarpine acetate **11**. Dihydroprecondylocarpine acetate **11**, in the open form, can form dehydrosecodine **9** through the action of CS or TS to form catharanthine **3** or tabersonine **2**, respectively. Numbers within structures indicate different carbon atoms. spont., spontaneous; HRMS, high-resolution mass spectrometry; MS/MS, tandem mass spectrometry.



**Fig. 3. Biosynthesis of tabersonine 2 and catharanthine 3 from stemmadenine acetate 7 starting substrate.** (A) Reconstitution of tabersonine 2 and catharanthine 3 in *N. benthamiana* from stemmadenine acetate 7. Extracted ion chromatograms (XIC) for ions with mass/charge ratio ( $m/z$ ) 397.19 (stemmadenine acetate 7),  $m/z$  395.19 (precondylocarpine acetate 10), and  $m/z$  337.19 [catharanthine at retention time (RT) 4.0 min and tabersonine at RT 4.4 min] are shown. EV, empty vector. (B) Interaction of CS and TS with DPAS by bimolecular fluorescence complementation (BiFC) in

*C. roseus* cells. The efficiency of BiFC complex reformation reflected by the YFP fluorescence intensity highlighted that CS and DPAS exhibited weak interactions (i to iii), whereas TS and DPAS strongly interacted (iv to vi). No interactions with loganic acid methyltransferase (LAMT) were observed (vii to ix). Y<sup>N</sup>, YFP N-terminal fragment; Y<sup>C</sup>, YFP C-terminal fragment; nuc-CFP, nuclear cyan fluorescent protein; DIC, differential interference contrast. Scale bars, 10  $\mu$ m. **(C)** Phylogenetic relationship of PAS with other functionally characterized berberine bridge enzymes.

However, the limited two-dimensional NMR dataset was consistent with an assignment of precondylocarpine acetate **10**. Thus, we renamed this alcohol dehydrogenase dihydroprecondylocarpine synthase (DPAS). Collectively, these data suggest that PAS and DPAS act in concert with CS or TS to generate catharanthine **3** and tabersonine **2**.

To validate whether these enzymes produce catharanthine **3** and tabersonine **2**, we transiently coexpressed PAS, DPAS, and CS or TS in the presence of stemmadenine acetate **7** in *Nicotiana benthamiana*. These experiments illustrated the sequential activity of the newly discovered enzymes, whereby we observed formation of catharanthine **3** in plant tissue overexpressing PAS, DPAS, and CS or tabersonine **2** in plant tissue overexpressing PAS, DPAS, and TS, when the leaf was also co-infiltrated with stemmadenine acetate **7** (Fig. 3A). The presence of all proteins was validated by proteomics analysis (fig. S4B and data S2). Formation of precondylocarpine acetate **10** was observed

when stemmadenine acetate **7** was infiltrated into *N. benthamiana* in the absence of any heterologous enzymes (Fig. 3A), suggesting that an endogenous redox enzyme(s) of *N. benthamiana* can oxidize stemmadenine acetate **7**. Formation of **3** and **2** was validated by coelution with commercial standards, and formation of **10** was validated by coelution with the semisynthetic compound.

Purified proteins were required to validate the biochemical steps of this reaction sequence in vitro. Whereas CS, TS, and DPAS all expressed in soluble form in *E. coli* (fig. S4A), the flavin-dependent enzyme PAS failed to express in standard expression hosts such as *E. coli* or *Saccharomyces cerevisiae*. To overcome this obstacle, we expressed the native full-length PAS in *N. benthamiana* plants using a transient expression system (fig. S4B) in *Pichia pastoris* (fig. S4, C and D) and in Sf9 insect cells (fig. S4E). The presence of PAS was validated by proteomic data (data S3). Reaction of PAS from each of these

expression hosts with stemmadenine acetate **7** produced a compound that had an identical mass and retention time to our semisynthetic standard of precondylocarpine acetate **10** (figs. S15 and S16). The enzymatic assays with PAS protein derived from *P. pastoris* and Sf9 insect cells ensure that formation of the expected products is not the result of a protein contaminant found in the plant-expressed PAS protein. When the PAS proteins (from *N. benthamiana* and *P. pastoris*), along with stemmadenine acetate **7**, were combined with heterologous DPAS and CS, catharanthine **3** was formed, and when combined with DPAS and TS, tabersonine **2** was observed (figs. S17 and S18).

Semisynthetic precondylocarpine acetate **10** could be reacted with DPAS and TS or CS to yield tabersonine **2** or catharanthine **3**, respectively (fig. S19). In addition, a crude preparation of what is proposed to be dihydroprecondylocarpine acetate **11**, synthesized according to Scott and Wei (8), was converted to catharanthine **3** and



tabersonine **2** by the action of CS and TS, respectively (fig. S20). Reaction of PAS (purified from *N. benthamiana*) and DPAS with stemmadenine acetate **7** in the absence of CS or TS yielded a compound isomeric to tabersonine **2** and catharanthine **3**, suggesting that cyclization can occur spontaneously under these reaction conditions (fig. S17). As observed during attempts to purify the CS or TS substrate from *Tabernaemontana* plants, dihydroprecondylocarpine acetate **11** can also deformylate to form tubotaiwine **12**. Solvent and reaction conditions probably determine how the reactive dihydroprecondylocarpine acetate **11** decomposes.

PAS failed to react with stemmadenine **1**, indicating that the enzyme recognized the acetyl (Ac) group (fig. S21). Oxidation of stemmadenine **1** produced a compound with a mass consistent with that of the shunt product condylocarpine **13**. This identification was supported by comparison of the tandem mass spectrometry spectrum of **13** to that of the related compound tubotaiwine **12** (fig. S22). The transformation of stemmadenine **1** to condylocarpine **13** is known (15, 21). We therefore suspect that the acetylation of stemmadenine **1** is necessary to slow spontaneous deformylation after oxidation, and AcOH serves as a leaving group to allow formation of dehydrosecodine **9** (20). Acetylation also functions as a protecting group in the biosynthesis of noscapine in opium poppy (22).

The reactivity of the intermediates involved in the transformation of stemmadenine acetate **7** to catharanthine **3** or tabersonine **2** suggests that PAS, DPAS, and CS or TS should be colocalized because the unstable post-precondylocarpine acetate **10** intermediates may not remain intact during transport between cell types or compartments. Using yellow fluorescent protein (YFP)-tagged proteins in *C. roseus* cell suspension culture, we showed that PAS is targeted to the vacuole through small vesicles budding from the endoplasmic reticulum (ER), as was previously observed for the PAS homolog, berberine bridge enzyme (23) (fig. S23). This localization suggests that stemmadenine acetate **7** oxidation occurs in the ER lumen, ER-derived vacuole-targeted vesicles, and/or vacuole. In contrast, colocalization of DPAS, CS, and TS was confirmed in the cytosol (figs. S24 and S25). Bimolecular fluorescence complementation suggested preferential interactions between DPAS and TS (Fig. 3B and figs. S26 and S27). Such interactions may not only prevent undesired reactions on the reactive dihydroprecondylocarpine acetate **11** intermediate but may also control the flux of **11** into tabersonine **2**.

Homologs of PAS are used throughout benzylisoquinoline and pyridine alkaloid biosynthesis.

Certain PAS mutations characterize the enzymes found in aspidosperma-alkaloid- and iboga-alkaloid-producing plant clades (Fig. 3C). For instance, PAS lacks the His and Cys residues involved in covalent binding of the FAD cofactor (fig. S28). We anticipate that these aspidosperma-associated PAS homologs populate the metabolic pathways for the wide range of aspidosperma alkaloids found in nature. DPAS is a medium-chain alcohol dehydrogenase, a class of enzymes widely used in monoterpene indole alkaloid biosynthesis (14, 18, 19, 24). We hypothesize that CS and TS may retain the hydrolysis function of the putative ancestor hydrolase enzyme (25) to allow formation of dehydrosecodine **9** from dihydroprecondylocarpine acetate **11**. In principle, the formation of tabersonine **2** and catharanthine **3** is formed via two different modes of cyclization, and dehydrosecodine **9** can undergo two distinct Diels-Alder reactions (26) to form either catharanthine **3** or tabersonine **2** (Fig. 2).

Here we report the discovery of two enzymes—PAS and DPAS—along with the discovery of the catalytic function of two other enzymes—CS and TS (11)—that convert stemmadenine acetate **7** to tabersonine **2** and catharanthine **3**. Chemical investigations of this system (6–8, 15), coupled with plant DNA sequence data, enabled discovery of the last enzymes responsible for the construction of the tabersonine **2** or catharanthine **3** scaffolds. With the biosynthesis of stemmadenine acetate **7** (11), this completes the biosynthetic pathway for vindoline **4** and catharanthine **3**, compounds that can be used to semisynthetically prepare vinblastine. These discoveries allow the prospect of heterologous production of these expensive and valuable compounds in alternative host organisms, providing a new challenge for synthetic biology (27).

## REFERENCES AND NOTES

1. S. E. O'Connor, J. J. Maresh, *Nat. Prod. Rep.* **23**, 532–547 (2006).
2. T. Dugé de Bernonville et al., *Phytochemistry* **113**, 9–23 (2015).
3. S. Vonny, V. De Luca, in *Advances in Botanical Research* (Elsevier, 2013), pp. 1–37.
4. S. C. Benson et al., *Heterocycles* **84**, 135–155 (2012).
5. G. K. Jana, S. Paul, S. Sinha, *Org. Prep. Proced. Int.* **43**, 541–573 (2011).
6. E. Wenkert, *J. Am. Chem. Soc.* **84**, 98–102 (1962).
7. J. P. Kutney, Y. Kartou, N. Kawamura, B. R. Worth, *Can. J. Chem.* **60**, 1269–1278 (1982).
8. A. I. Scott, C. C. Wei, *Tetrahedron* **30**, 3003–3011 (1974).
9. H. Mizoguchi, H. Oikawa, H. Oguri, *Nat. Chem.* **6**, 57–64 (2014).
10. F. Kellner et al., *Plant J.* **82**, 680–692 (2015).
11. Y. Qu et al., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3180–3185 (2018).
12. N. Petitfrère, A. M. Morfaux, M. M. Debray, L. Le Men-Olivier, J. Le Men, *Phytochemistry* **14**, 1648–1649 (1975).
13. B. A. Dadson, J. Harley-Mason, *J. Chem. Soc. D* **1969**, 665b (1969).
14. E. C. Tatsis et al., *Nat. Commun.* **8**, 316 (2017).
15. A. I. Scott, C. C. Wei, *J. Am. Chem. Soc.* **94**, 8264–8265 (1972).
16. A. I. Scott, *Acc. Chem. Res.* **3**, 151–157 (1970).

17. N. G. H. Leferink, D. P. H. M. Heuts, M. W. Fraaije, W. J. H. van Berkel, *Arch. Biochem. Biophys.* **474**, 292–301 (2008).
18. A. Stavrinides et al., *Chem. Biol.* **22**, 336–341 (2015).
19. A. Stavrinides et al., *Nat. Commun.* **7**, 12116 (2016).
20. A. I. Scott, in *Recent Advances in Phytochemistry*, V. C. Runeckles, Ed. (Plenum Press, 1975), chap. 9, pp. 189–241.
21. M. El-Sayed, Y. H. Choi, M. Frédrich, S. Roytrakul, R. Verpoorte, *Biotechnol. Lett.* **26**, 793–798 (2004).
22. T. T. Dang, X. Chen, P. J. Facchini, *Nat. Chem. Biol.* **11**, 104–106 (2015).
23. A. Bock, G. Wanner, M. H. Zenk, *Planta* **216**, 57–63 (2002).
24. Y. Qu et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6224–6229 (2015).
25. N. Lenfant et al., *Nucleic Acids Res.* **41**, D423–D429 (2013).
26. E. M. Stocking, R. M. Williams, *Angew. Chem. Int. Ed.* **42**, 3078–3115 (2003).
27. A. Casini et al., *J. Am. Chem. Soc.* **140**, 4302–4316 (2018).
28. M. Sottomayor, A. Ros Barceló, *Protoplasma* **222**, 97–105 (2003).
29. H. Ishikawa et al., *J. Am. Chem. Soc.* **131**, 4904–4916 (2009).

## ACKNOWLEDGMENTS

We thank G. Saalbach of the John Innes Proteomics Centre for proteomics work, F. Kellner and F. Geu-Flores for early silencing results with CS and TS, B. Lichman for discussions, P. O'Connor for *Tabernaemontana* material, and R. Hughes and M. Franceschetti for preparing the modified TRBO vector. We also thank MRC PPU Reagents and Services for PAS expression in insect cells. **Funding:** This work was supported by grants from the European Research Council (311363), BBSRC (BB/J004561/1) (S.E.O.), and the Région Centre-Val de Loire, France (CatharSIS grant and BioPROPHARM project-ARD2020) (V.C.). J.F. acknowledges DFG postdoctoral funding (FR 3720/1-1) and T.-T.T.D. (ALTF 239-2015) and S.C.F. (ALTF 846-2016) acknowledge EMBO Long-Term Fellowships. K.K. acknowledges the French Embassy in Greece for a SSHN grant. R.M.E.P. was supported by a Ph.D. studentship from BBSRC. **Author contributions:** L.C. characterized all proteins, purified substrates, and coordinated all experiments; J.F. performed experiments with PAS and developed chemical mechanisms and synthetic strategy; S.C.F. helped with candidate selection and VIGS and performed the quantitative polymerase chain reaction experiments; K.C. performed all semisynthetic transformations; R.M.E.P. performed initial cloning and characterization of PAS; T.-D.N. expressed PAS in *P. pastoris*; T.-T.T.D. performed reconstitution in *N. benthamiana*; B.A. performed cloning and silencing experiments; D.M.J. helped to identify PAS and designed initial silencing constructs; I.J.C.V. purified stemmadenine; I.S.T.C., K.K., T.D.B., and V.C. performed localization and interaction studies; L.C. and S.E.O. wrote the manuscript; and S.C.F., J.F., K.C., and V.C. revised the manuscript. **Competing interests:** A patent on PAS and DPAS is currently being filed. **Data and materials availability:** Sequence data for PAS (MH213134), DPAS (KU865331), CS (MF770512), and TS (MF770513) are deposited in GenBank and are also provided in the supplementary materials. Transcriptome data used in this study have already been published (see supplementary materials and <http://medicinalplantgenomics.msu.edu/>). Cell lines used in this study will be made available upon request. Data supporting the findings of this study are available within the article and the supplementary materials.

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/1235/suppl/DC1](http://www.sciencemag.org/content/360/6394/1235/suppl/DC1)  
Materials and Methods  
Figs. S1 to S28  
Tables S1 to S7  
References (30–59)  
Data S1 to S3

5 March 2018; accepted 24 April 2018  
Published online 3 May 2018  
10.1126/science.aat4100

## NEUROSCIENCE

# Reactivation of recall-induced neurons contributes to remote fear memory attenuation

Ossama Khalaf, Siegfried Resch, Lucie Dixsaut, Victoire Gorden, Liliane Glauser, Johannes Gräff\*

Whether fear attenuation is mediated by inhibition of the original memory trace of fear with a new memory trace of safety or by updating of the original fear trace toward safety has been a long-standing question in neuroscience and psychology alike. In particular, which of the two scenarios underlies the attenuation of remote (month-old) fear memories is completely unknown, despite the impetus to better understand this process against the backdrop of enduring traumata. We found—chemogenetically and in an engram-specific manner—that effective remote fear attenuation is accompanied by the reactivation of memory recall-induced neurons in the dentate gyrus and that the continued activity of these neurons is critical for fear reduction. This suggests that the original memory trace of fear actively contributes to remote fear attenuation.

**T**raumatic memories develop after the experience of grave physical or psychological harm. Such memories are extraordinarily robust and difficult to treat, with an estimated lifetime prevalence of fear- and stress-related disorders of close to 29% (1). Despite these circumstances, little is known about how long-lasting (remote) fearful memories are stored and can be attenuated. In particular, it is unclear whether the attenuation of remote fearful memories represents an inhibition of the original trace of fear with a new memory trace of safety, a process termed extinction (2–4), or unlearning of the original fear memory trace toward safety following memory recall, termed reconsolidation-updating (5, 6).

To investigate this question, we visualized memory traces implicated in remote fear memory storage and attenuation by using double transgenic TetTag mice, which provide a *c-fos*-based labeling system of neuronal activity (7) (Fig. 1A). TetTag mice show no difference from wild-type mice in general anxiety and overall locomotion, making them ideally suited for studying acquired fear-related behaviors (fig. S1). We focused on the dentate gyrus (DG) because of its importance in the encoding (8), recall (9, 10), and attenuation (11) of contextual fear and because of the recently demonstrated implication of DG granule cells in fear memory retrieval by *c-Fos*-based engram studies (10, 12–14).

First, we identified the cellular ensemble participating in remote (28-day-old) fear memory storage. TetTag mice were contextually fear-conditioned while on doxycycline (DOX) and then taken off DOX 3 days prior to remote memory recall, which triggered lasting  $\beta$ -galactosidase

( $\beta$ -gal) expression in recall-induced neurons (Fig. 1, A and B). Immediately after memory recall, TetTag mice were put on DOX again, and 1 day later they entered a previously described spaced extinction paradigm (15), which successfully reduced their fearful memories (fig. S2). Second, we used endogenous *c-Fos* expression 1 hour after the last extinction trial to identify the cellular ensemble participating in memory attenuation per se (Fig. 1, A and B).

We then evaluated whether neurons activated by remote memory recall ( $\beta$ -gal<sup>+</sup> cells) became reengaged during extinction training (as indicated by *c-Fos*<sup>+</sup> cells) by calculating their reactivation rate, defined as the number of double-positive  $\beta$ -gal<sup>+</sup> *c-Fos*<sup>+</sup> cells normalized to the total number of  $\beta$ -gal<sup>+</sup> cells. In the DG, this reactivation rate was significantly higher than chance levels in the DG, unlike in hippocampal area CA3 (Fig. 1, B and C), which is of minimal importance for remote contextual fear retrieval (10). Furthermore, we found a positive correlation between the reactivation rate and the degree of fear memory attenuation (expressed as  $\Delta$  freezing, the difference in freezing between the recall and the last extinction session) in the DG but not in CA3 (Fig. 1D). Compared with a home cage control group, the CS-US group (the group exposed to the conditioned stimulus in association with the unconditioned stimulus) showed an increased recall-induced activation rate (the number of  $\beta$ -gal<sup>+</sup> cells normalized to the total number of cells) in both the DG and CA3, testifying to the specificity of  $\beta$ -gal<sup>+</sup> as a marker of remote memory recall (fig. S3). Likewise, the DG and CA3 did not differ in the extinction learning rate (the number of extinction-induced *c-Fos*<sup>+</sup> cells normalized to the total number of cells) (fig. S4), and neither did the activation or the extinction learning rate correlate with fear memory attenuation (fig. S5). These findings suggest that remote fear attenu-

ation is facilitated as more recall-induced neurons are reactivated by the extinction paradigm.

We independently verified these findings in a different paradigm of memory attenuation by using a second mouse line and a different technique to visualize neuronal ensembles. C57Bl6/J wild-type mice were contextually fear-conditioned as before (Fig. 1E), and 28 days later they entered a previously described massed extinction paradigm (15), which successfully attenuated their remote fear memories (fig. S6) to levels comparable to those obtained in the spaced extinction paradigm (fig. S7). Then we used the intracellular spatiotemporal characteristics of immediate early gene mRNA species and cellular compartment analysis of temporal activity by fluorescence in situ hybridization (catFISH) (16, 17) to identify the neuronal populations activated during remote fear memory recall (those displaying cytoplasmic *Homer1a* mRNA) and during massed extinction (those displaying nuclear *c-fos* mRNA) (Fig. 1, E and F). Similar to the results in the spaced extinction paradigm, the reactivation rate in the DG was significantly higher than chance levels and correlated positively with fear memory attenuation, in contrast to that in hippocampal area CA3 (Fig. 1, F to H). Compared with levels in the home cage control group, the recall-induced activation rate (the total amount of *Homer1a*<sup>+</sup> cells) in both areas was significantly elevated (fig. S8), indicating *Homer1a*<sup>+</sup> labeling to be specific for memory recall. Similarly, there was no difference in the extinction learning rate (the total amount of *c-fos*<sup>+</sup> cells) between the DG and CA3 (fig. S9), and neither the activation nor the extinction learning rate correlated with fear attenuation (fig. S10).

Next, to test whether the sustained activity of recall-induced neurons contributes to the behavioral manifestation of fear attenuation, we used the Daun02 inactivation method (18), which reduces neuronal excitability (19) by means of daunorubicin, a derivative converted from the prodrug Daun02 exclusively by  $\beta$ -gal (Fig. 2A). TetTag mice were fear-conditioned and  $\beta$ -gal<sup>+</sup> expression was triggered by remote recall as described above. Between training and recall, cannulas were stereotactically implanted in the DG so that Daun02 could be readily administered 90 min post-memory recall (Fig. 2B) to inhibit the activity of recall-induced neurons. Then the spaced extinction paradigm was performed. Compared with vehicle-treated animals, Daun02-treated TetTag mice showed impaired fear memory attenuation in the DG (Fig. 2C) despite displaying similar freezing at recall (fig. S11). This was accompanied by a reduced number of *c-Fos*<sup>+</sup> cells at the completion of memory attenuation (Fig. 2, D and E), consistent with the time course of 4-day-long neuronal ensemble inactivation after a single Daun02 injection in other brain areas (18, 20). In contrast, when administered in hippocampal area CA3, not characterized by reactivated recall-induced neurons after attenuation, Daun02 had no effect (fig. S12, A and B). Likewise, Daun02 had no effect in wild-type animals off DOX (fig. S12, C and D); in TetTag mice

Laboratory of Neuroepigenetics, Brain Mind Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

\*Corresponding author. Email: johannes.graef@epfl.ch



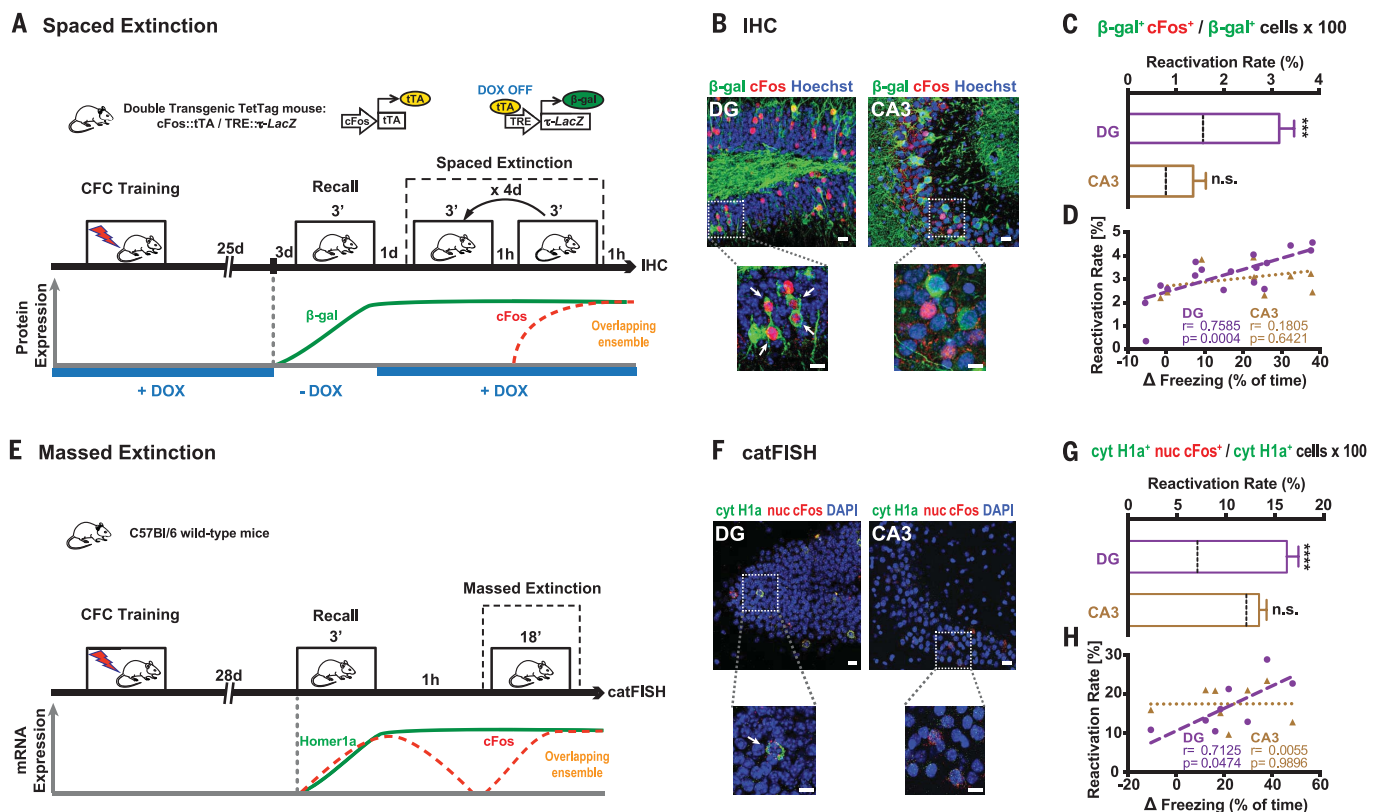
continuously on DOX (fig. S12, E and F); and in TetTag mice in which recall-induced neurons from a context B, different from the conditioning context A, were inhibited (fig. S12, G and H).

Lastly, we tested whether the continued activation of recall-induced neurons during the extinction paradigm facilitated remote fear memory attenuation. We used single transgenic c-Fos-tTA (tetracycline-controlled transactivator) mice and virally injected the excitatory DREADD (designer receptor exclusively activated by a designer drug) hM3Dq (AAV<sub>9</sub>-TRE::hM3Dq-mCherry) (Fig. 3A) to activate remote memory recall-induced ensembles by clozapine-N-oxide (CNO) administration. c-Fos-tTA animals were contextually fear-conditioned and stereotactically injected with the viral constructs in the DG between training and recall, and hM3Dq expression was triggered by remote recall as described above (Fig. 3B). To stim-

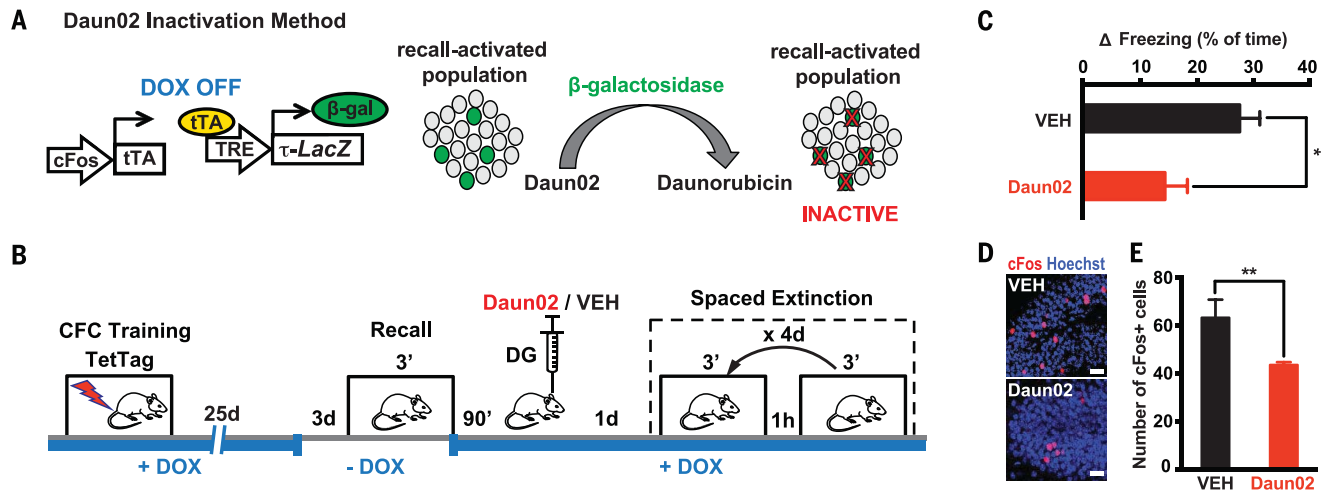
ulate the activity of the recall-induced ensemble throughout the fear attenuation procedure, mice were administered CNO intraperitoneally on each of the spaced extinction days (Fig. 3B). We found that in comparison with vehicle-treated animals and as corroborated by the same control groups described above (fig. S13), DG-injected CNO-treated mice displayed facilitated memory attenuation (Fig. 3C) despite showing similar freezing at recall (fig. S14). This facilitated fear memory attenuation was accompanied by an increased amount of extinction-induced c-Fos<sup>+</sup> cells among recall-induced mCherry<sup>+</sup> cells (Fig. 3, D and E).

Taken together, our data show that a subset of memory recall-induced neurons in the DG becomes reactivated after memory attenuation, that the degree of fear reduction positively correlates with this reactivation, and that consequently,

the continued activity of memory recall-induced neurons is critical for remote fear memory attenuation. Although other brain areas such as the prefrontal cortex (27) and the amygdala (22, 23) are likely to be implicated in remote fear memories and remain to be investigated, these results suggest that fear attenuation at least partially occurs in memory recall-induced ensembles through updating or unlearning of the original memory trace of fear. These data thereby provide the first evidence at an engram-specific level that fear attenuation may not be driven only by extinction learning, that is, by an inhibitory memory trace different from the original fear trace, as originally proposed by Pavlov (2) and experimentally supported for recent memories by others (3, 22, 24, 25). Rather, our findings indicate that during remote fear memory attenuation both mechanisms likely coexist (26, 27), albeit

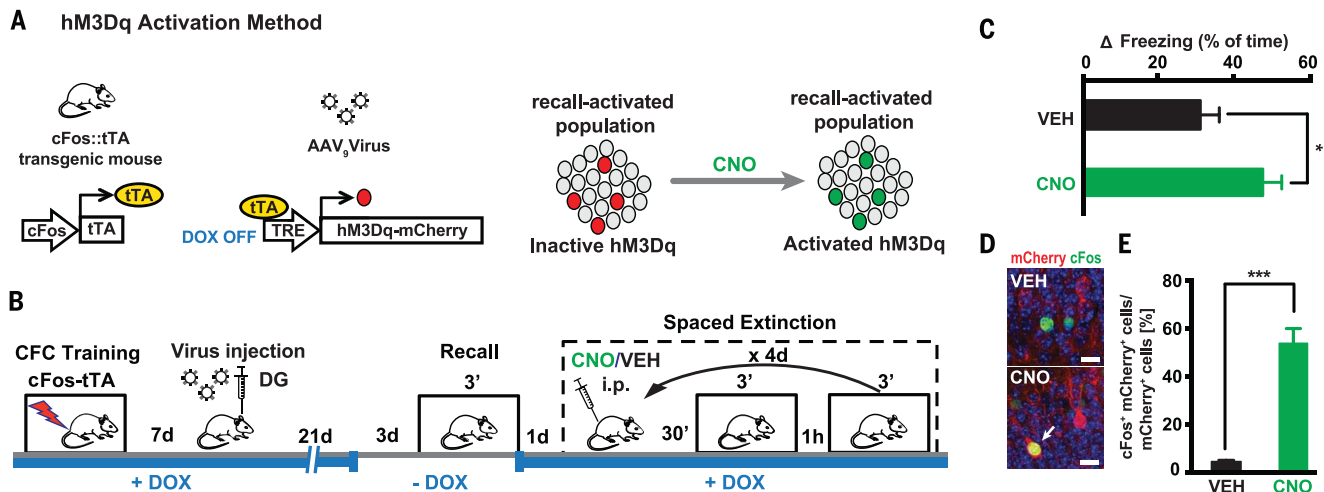


**Fig. 1. Remote fear attenuation is characterized by reactivation of memory recall-induced neurons in the DG but not CA3.** (A) Schematic representing the inducible double transgenic TetTag mouse (top) and the experimental design used in the spaced extinction paradigm (bottom). TRE, tetracycline response element; CFC, contextual fear conditioning; d, days; ', minutes; IHC, immunohistochemistry. (B) Confocal images of the DG and CA3 regions. Activated neurons at remote memory recall express β-gal (green), whereas those activated by the last extinction session express endogenous c-Fos (red). Scale bars, 20 μm. Smaller panels below show enlarged areas within the dotted squares. Arrows indicate recall-induced neurons reactivated by extinction training. Scale bars, 10 μm. (C) The reactivation rate shows a significant increase in the DG ( $n = 17$  animals) but not in CA3 ( $n = 9$  animals) over chance levels (dashed lines). Data are means  $\pm$  SEM compared by a two-tailed, unpaired  $t$  test. \*\*\* $P = 0.001$ ; n.s., not significant. (D) The reactivation rate in the DG (circles) but not that in CA3 (triangles) is positively correlated with the degree of fear memory attenuation ( $\Delta$  freezing). DG,  $n = 17$  animals, correlation coefficient  $r = 0.7585$ ,  $P = 0.0004$ ; CA3,  $n = 9$  animals,  $r = 0.1805$ ,  $P = 0.6421$ . (E) Schematic of the experimental design used in the massed extinction paradigm for wild-type mice. (F) Confocal images of the DG and CA3 regions. Activated neurons at remote memory recall express cytoplasmic *Homer1a* (cyt *H1a*) mRNA (green), whereas those activated during extinction training express nuclear (nuc) *c-fos* mRNA (red). Scale bars, 20 μm. Smaller panels below show enlarged areas within the dotted squares. The arrow indicates a representative recall-induced neuron reactivated by extinction training. Scale bars, 10 μm. (G) The reactivation rate shows a significant increase in the DG ( $n = 8$  animals) but not in CA3 ( $n = 8$  animals) over chance levels (dashed lines). Data are means  $\pm$  SEM compared by a two-tailed, unpaired  $t$  test (\*\*\*\* $P < 0.0001$ ). (H) The reactivation rate in the DG (circles) but not that in CA3 (triangles) is positively correlated with the degree of fear memory attenuation (expressed as  $\Delta$  freezing). DG,  $n = 8$  animals,  $r = 0.7125$ ,  $P = 0.0474$ ; CA3,  $n = 8$  animals;  $r = 0.0055$ ,  $P = 0.9896$ .



**Fig. 2. Loss-of-function of memory recall-induced neurons impairs remote fear attenuation.** (A) Schematic and (B) experimental design of the Daun02 inactivation method. (C) Compared with vehicle (VEH)-treated animals ( $n = 9$ ), Daun02-treated animals ( $n = 9$ ) show impaired fear memory attenuation ( $\Delta$  freezing) in the DG. Data are means  $\pm$  SEM compared by a

two-tailed, unpaired  $t$  test ( $*P < 0.05$ ). (D) Confocal images of the VEH- and Daun02-treated DG showing reduced c-Fos induction in Daun02-treated animals. Scale bars, 20  $\mu$ m. (E) Quantification of (D). VEH-treated animals,  $n = 3$ ; Daun02-treated animals,  $n = 6$ . Data are means  $\pm$  SEM compared by a two-tailed, unpaired  $t$  test ( $**P < 0.01$ ).



**Fig. 3. Gain-of-function of memory recall-induced neurons facilitates remote fear attenuation.** (A) Schematic and (B) experimental design of the hM3Dq-mediated activation method. i.p., intraperitoneally. (C) Compared with vehicle (VEH)-treated animals ( $n = 13$ ), CNO-treated animals ( $n = 16$ ) show facilitated fear memory attenuation ( $\Delta$  freezing). Data are means  $\pm$  SEM compared by a two-tailed, unpaired  $t$  test

( $*P < 0.05$ ). (D) Confocal images of the VEH- and CNO-treated DG. The arrow indicates an example of a recall-induced neuron (mCherry) reactivated by extinction training (c-Fos) in CNO-treated animals. Scale bars, 10  $\mu$ m. (E) Quantification of (D). VEH-treated animals,  $n = 4$ ; CNO-treated animals,  $n = 5$ . Data are means  $\pm$  SEM compared by a two-tailed, unpaired  $t$  test ( $***P < 0.001$ ).

with the importance of the continued activity of memory recall-induced neurons experimentally documented herein. Such activity may not only represent the capacity for a valence change in DG engram cells (28) but also be a prerequisite for memory reconsolidation, namely, an opportunity for learning inside the original memory trace (29). As such, this activity likely constitutes a physiological correlate sine qua non for effective exposure therapies against traumatic memories in humans: the engagement, rather than the suppression, of the original trauma (30). In future studies, this experimental approach may prove useful to assess boundary conditions for extinction versus reconsolidation (5, 31–33) or counter-

conditioning (34) in an engram-specific manner and, by extension, to gauge the efficiency of other interventions to reduce traumatic memories.

#### REFERENCES AND NOTES

- R. C. Kessler et al., *Arch. Gen. Psychiatry* **62**, 593–602 (2005).
- I. Pavlov, *Conditioned Reflexes* (Oxford Univ. Press, Oxford, 1927).
- M. E. Bouton, *Learn. Mem.* **11**, 485–494 (2004).
- K. M. Myers, M. Davis, *Mol. Psychiatry* **12**, 120–150 (2007).
- M.-H. Monfils, K. K. Cowansage, E. Klann, J. E. LeDoux, *Science* **324**, 951–955 (2009).
- D. Schiller et al., *Nature* **463**, 49–53 (2010).
- L. G. Reijmers, B. L. Perkins, N. Matsuo, M. Mayford, *Science* **317**, 1230–1233 (2007).
- V. Hernández-Rabaza et al., *Neurobiol. Learn. Mem.* **90**, 553–559 (2008).
- T. Kitamura et al., *Science* **356**, 73–78 (2017).
- C. A. Denny et al., *Neuron* **83**, 189–201 (2014).
- B. E. Bernier et al., *J. Neurosci.* **37**, 6359–6371 (2017).
- X. Liu et al., *Nature* **484**, 381–385 (2012).
- S. Ramirez et al., *Science* **341**, 387–391 (2013).
- S. Park et al., *Neuropsychopharmacology* **41**, 2987–2993 (2016).
- J. Gräff et al., *Cell* **156**, 261–276 (2014).
- J. F. Guzowski, B. L. McNaughton, C. A. Barnes, P. F. Worley, *Nat. Neurosci.* **2**, 1120–1124 (1999).
- A. Nonaka et al., *J. Neurosci.* **34**, 9305–9309 (2014).
- E. Koya et al., *Nat. Neurosci.* **12**, 1069–1073 (2009).
- M. Engeln et al., *Biol. Psychiatry* **79**, 354–361 (2016).
- J. M. Bossert et al., *Nat. Neurosci.* **14**, 420–422 (2011).
- P. W. Frankland, B. Bontempi, *Nat. Rev. Neurosci.* **6**, 119–130 (2005).
- C. Herry et al., *Nature* **454**, 600–606 (2008).
- C. Herry, J. P. Johansen, *Nat. Neurosci.* **17**, 1644–1654 (2014).
- S. Tronche, J. M. Sasaki, T. Tu, L. G. Reijmers, *Neuron* **80**, 1054–1065 (2013).
- N. C. Tronson et al., *J. Neurosci.* **29**, 3387–3394 (2009).



26. R. L. Clem, D. Schiller, *Trends Neurosci.* **39**, 340–351 (2016).
27. C. M. Alberini, *Front. Behav. Neurosci.* **5**, 12 (2011).
28. R. L. Redondo *et al.*, *Nature* **513**, 426–430 (2014).
29. R. G. M. Morris *et al.*, *Neuron* **50**, 479–489 (2006).
30. E. B. Foa, M. J. Kozak, *Psychol. Bull.* **99**, 20–35 (1986).
31. M. C. Inda, E. V. Muravieva, C. M. Alberini, *J. Neurosci.* **31**, 1635–1643 (2011).
32. E. Merlo, A. L. Milton, Z. Y. Goozée, D. E. Theobald, B. J. Everitt, *J. Neurosci.* **34**, 2422–2431 (2014).
33. Y. Dudai, *Annu. Rev. Neurosci.* **35**, 227–247 (2012).
34. A. Meulders, P. A. Karsdorp, N. Claes, J. W. S. Vlaeyen, *J. Pain* **16**, 1353–1365 (2015).

#### ACKNOWLEDGMENTS

We dedicate this work to O.K.'s father, Mohamed Salah El-Dien, and J.G.'s mother, Wilma, who both sadly passed away during its

completion. We thank the vector core at UNC, Chapel Hill, for providing the adeno-associated virus; S. Hausmann, O. Prat, and G. Urueña for assistance in histology and quantification; and J. V. Sanchez-Mut for comments on earlier versions of the manuscript. **Funding:** Research on remote memories in J.G.'s laboratory is supported by the Swiss National Science Foundation (31003A\_155898), the National Competence Center for Research "Synapsy," and the European Research Council (ERC-2015-StG 678832). J.G. is an MQ fellow (MQ15FIP100012) and a NARSAD independent investigator (through independent investigator award 24497). **Author contributions:** This study was planned and conceptualized by O.K. and J.G. O.K. carried out the experiments and analyzed data. S.R., V.G., and L.D. contributed to surgeries, histology, catFISH experiments, and quantification. L.G. contributed to mouse colony maintenance and the Daun02 method. The paper was written by O.K. and J.G.

and commented on by all authors. **Competing interests:** None. **Data and materials availability:** pAAV-TREtight-hM3Dq-mCherry is available from J.G. under a material agreement with Imperial College London. All data necessary to understand and assess the conclusions of this research are available in the supplementary materials.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/1239/suppl/DC1](http://www.sciencemag.org/content/360/6394/1239/suppl/DC1)  
Materials and Methods  
Figs. S1 to S14  
Reference (35)

12 January 2018; accepted 26 April 2018  
10.1126/science.aas9875

## STRUCTURAL BIOLOGY

## Near-atomic model of microtubule-tau interactions

Elizabeth H. Kellogg<sup>1,2\*</sup>, Nisreen M. A. Hejab<sup>2\*</sup>, Simon Poepsel<sup>1</sup>, Kenneth H. Downing<sup>2</sup>, Frank DiMaio<sup>3,4</sup>, Eva Nogales<sup>1,2,5†</sup>

Tau is a developmentally regulated axonal protein that stabilizes and bundles microtubules (MTs). Its hyperphosphorylation is thought to cause detachment from MTs and subsequent aggregation into fibrils implicated in Alzheimer's disease. It is unclear which tau residues are crucial for tau-MT interactions, where tau binds on MTs, and how it stabilizes them. We used cryo-electron microscopy to visualize different tau constructs on MTs and computational approaches to generate atomic models of tau-tubulin interactions. The conserved tubulin-binding repeats within tau adopt similar extended structures along the crest of the protofilament, stabilizing the interface between tubulin dimers. Our structures explain the effect of phosphorylation on MT affinity and lead to a model of tau repeats binding in tandem along protofilaments, tethering together tubulin dimers and stabilizing polymerization interfaces.

Microtubules (MTs) are formed by the assembly of  $\alpha\beta$ -tubulin dimers into protofilaments (PFs) that associate laterally into hollow tubes. MTs are regulated by MT-associated proteins (MAPs), including "classical" MAPs such as MAP-2, MAP-4, and tau that are critical to neuronal growth and function. Tau constitutes more than 80% of neuronal MAPs, stabilizes and bundles axonal MTs (1), and is developmentally regulated (2). Full-length adult tau is intrinsically disordered and includes a projection domain, an MT-binding region of four imperfect sequence repeats (R1 to R4), and a C-terminal domain (3) (Fig. 1A; the precise definition of the repeats has not always been consistent in the literature and the one displayed in the figure is justified by the structural findings in this study). Different repeats bind to and stabilize MTs (4, 5), with affinity and activity increasing with the number of repeats (5, 6). Neurodegenerative tauopathies, including Alzheimer's disease, develop when mutated (7, 8) or abnormally phosphorylated (9–11) tau loses affinity for MTs and forms filamentous aggregates called neurofibrillary tangles. Whereas we know the structure of amyloid tau fibrils (12), the physiological conformation of MT-bound tau remains controversial (13–17). Here we present atomic models of MT-bound tau by using a combination of single-particle cryo-electron microscopy (cryo-EM) and Rosetta modeling.

We used cryo-EM to visualize MTs in the presence of an excess of different tau constructs (figs. S1 and S2). The cryo-EM structure of dynamic

MTs (without stabilizing drugs or nonhydrolyzable guanosine 5'-triphosphate analogs) assembled with full-length tau (overall resolution of 4.1 Å) (fig. S2C) shows tau as a narrow, discontinuous density along each PF (Fig. 1B), following the ridge on the MT surface defined by the H11 and H12 helices of  $\alpha$ - and  $\beta$ -tubulin and adjacent to the site of attachment of the C-terminal tubulin tails (Fig. 1C) that are important for tau affinity (18, 19). This location is consistent with a previous, low-resolution cryo-EM study (16). To test an alternatively proposed tau-binding site on the MT interior (17), we added tau to preformed MTs or to polymerizing tubulin, both in the absence of Taxol, but never saw tau density on the MT luminal surface.

We also examined two N- and C-terminally truncated tau constructs including either all four repeats (4R) or just the first two (2R) (fig. S2, A and B) (we refer to the construct containing four repeats as 4R, whereas the sequence of the fourth repeat is referred to as R4). Both reconstructions (4.8 and 5.6 Å, respectively) (fig. S2C) and that of full-length tau were indistinguishable at the resolutions obtained. In all three cases, the length of the tau density corresponds to an extended chain of ~27 amino acids, with a weak connecting density that would accommodate another 3 to 4 residues, adding up to the length of one repeat (31 to 32 amino acids).

The lower resolution for tau (4.6 to 6.5 Å) than for tubulin (4.0 to 4.5 Å) in our reconstructions (fig. S3) may be due to substoichiometric binding (unlikely because of the excess of tau), flexibility, and/or differences between the repeats. Thus, we pursued the structure of a synthetic tau construct with four identical copies of R1 (R1 $\times$ 4) (Fig. 2) and reached an overall resolution of 3.2 Å (Fig. 2A and fig. S2C), with local resolution for tau between 3.7 and 4.2 Å (fig. S3). In this map the best-resolved region of tau approached the resolution of surface regions of tubulin. Again, tau appears as regularly spaced segments separated by more discontinuous density, as ex-

pected from the alternation of more tightly bound segments interspersed with more mobile regions (5, 20). A polyaniline model accommodating 12 residues was built into the best-resolved segment of the R1 $\times$ 4 tau density at the interdimer interface.

Given the lack of large side chains or secondary structure, we used Rosetta (21) as an unbiased approach to assess different tau repeat sequence registers (see methods in supplementary materials). A single register and conformation were favored, both energetically and on the basis of fit to density, and included amino acids 256 to 267 (VKSISKSTENLK), the most conserved segment among tau repeats (Fig. 2B). An alternative, more aggressive refinement protocol (22), which introduces more structural variability during the refinement procedure, converged to the same sequence and structure (fig. S4). This common solution corresponds to a 12-residue sequence contained within an 18-amino acid fragment that is sufficient to promote MT polymerization (4, 23). Furthermore, the inferred interactions are consistent with sequence conservation among classical MAPs beyond tau (fig. S5), with conserved residues contributing critical tubulin interactions within our model (Fig. 2C). Ser<sup>258</sup> and Ser<sup>262</sup> form hydrogen bonds with  $\alpha$ -tubulin Glu<sup>434</sup>. Phosphorylation of the universally conserved Ser<sup>262</sup> (fig. S5) strongly attenuates MT binding (24) and is a marker of Alzheimer's disease (25). Our structure now explains how its phosphorylation disrupts tau-tubulin interactions. Although Thr<sup>263</sup> is also positioned to hydrogen-bond with Glu<sup>434</sup>, hydrophobic substitutions are tolerated at this position (fig. S5), indicating that this interaction may not be as essential. The conserved Lys<sup>259</sup> is positioned to interact with an acidic patch on  $\alpha$ -tubulin formed by Asp<sup>424</sup>, Glu<sup>420</sup>, and Glu<sup>423</sup> (Fig. 2C). Ile<sup>260</sup>, conserved in hydrophobic character across all R1 sequences (fig. S5), is buried within a hydrophobic pocket formed by  $\alpha$ -tubulin residues Ile<sup>265</sup>, Val<sup>435</sup>, and Tyr<sup>262</sup> at the interdimer interface (Fig. 2C). Asn<sup>265</sup>, universally conserved among repeats of classical MAPs (fig. S5), forms a stabilizing intramolecular hydrogen bond within the type II'  $\beta$  turn formed by residues 263 to 266. Lastly, Lys<sup>267</sup> is positioned to interact with the acidic  $\alpha$ -tubulin C-terminal tail, and its basic character is conserved (Fig. 2C and fig. S5).

The tau density beyond this 12-residue stretch is very weak (Fig. 2A), indicating that residues 242 to 255 lack ordered interactions with tubulin. This region of R1 is rich in prolines, but the corresponding regions of R2 and R3 show a distinct, conserved hydrophobic pattern (Fig. 1A and fig. S5). As other tau repeats might form additional interactions with the MT surface, we also obtained a cryo-EM reconstruction by using a synthetic tau construct containing four copies of repeat R2 (R2 $\times$ 4). The R2 $\times$ 4-MT reconstruction, at an overall resolution of 3.9 Å (Fig. 3A and fig. S2C), was similar to R1 $\times$ 4, especially at the interdimer tubulin interface (Fig. 3A and fig. S6A), but had additional tau density along the surface of  $\beta$ -tubulin. We could model a backbone stretch of 27 residues into the R2 $\times$ 4 tau density, spanning three tubulin

<sup>1</sup>QB3 Institute and Department of Molecular and Cell Biology, University of California–Berkeley, Berkeley, CA 94720, USA.

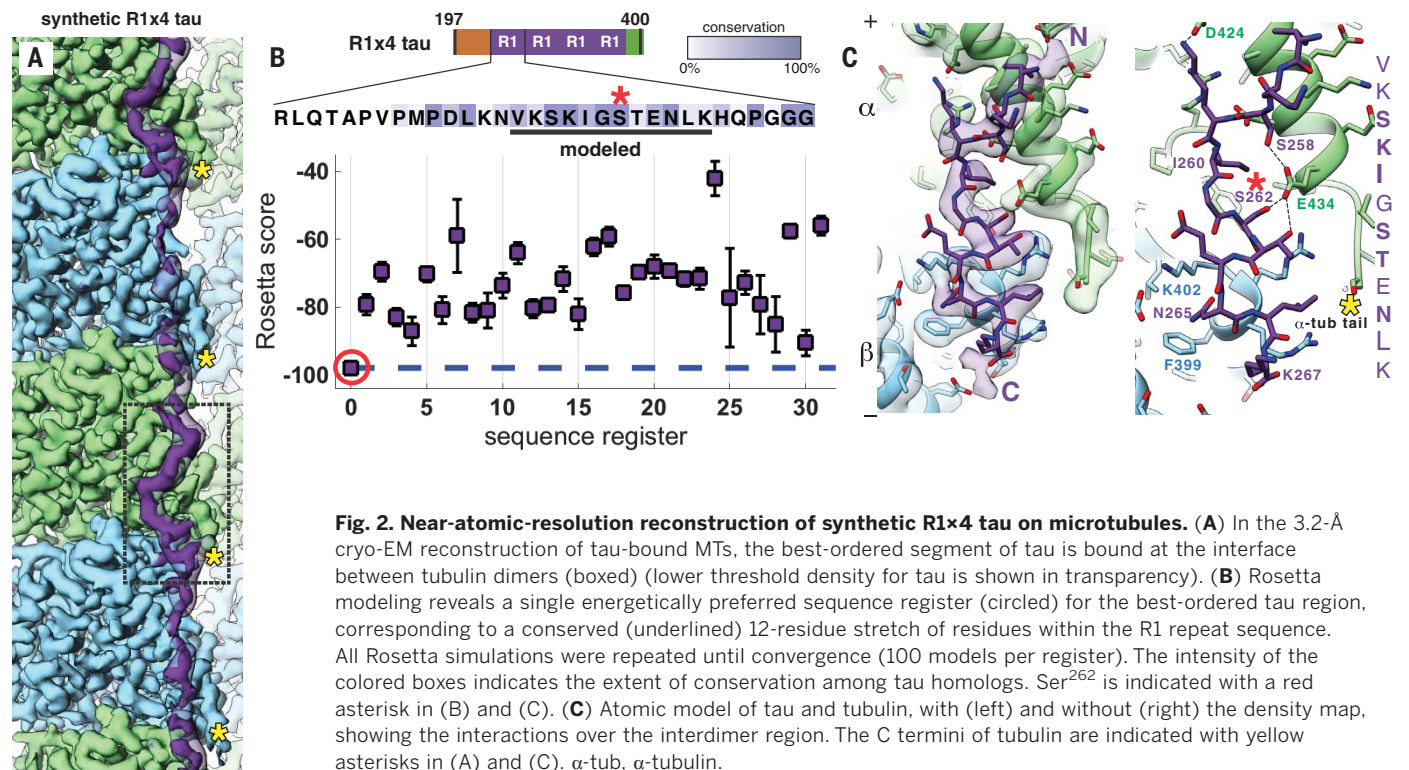
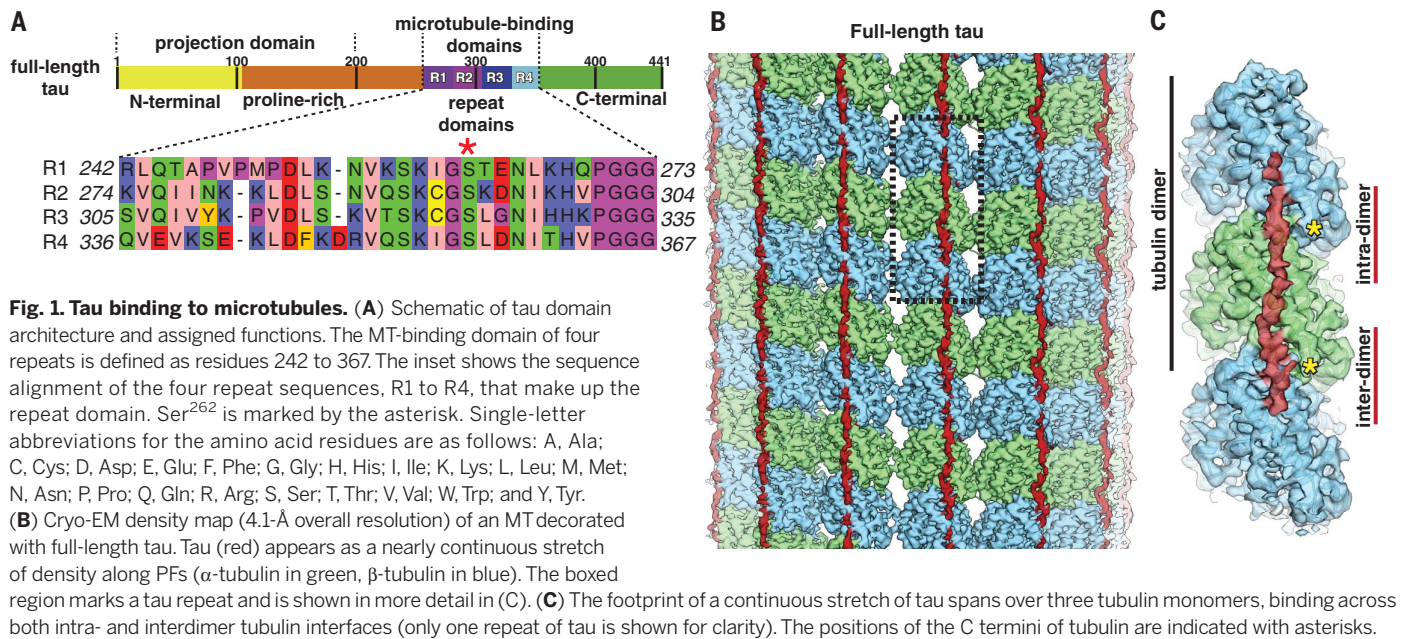
<sup>2</sup>Division of Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

<sup>3</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. <sup>4</sup>Institute for Protein Design, Seattle, WA 98195, USA. <sup>5</sup>Howard Hughes Medical Institute, University of California–Berkeley, Berkeley, CA 94720, USA.

\*These authors contributed equally to this work.

†Corresponding author. Email: enogales@lbl.gov



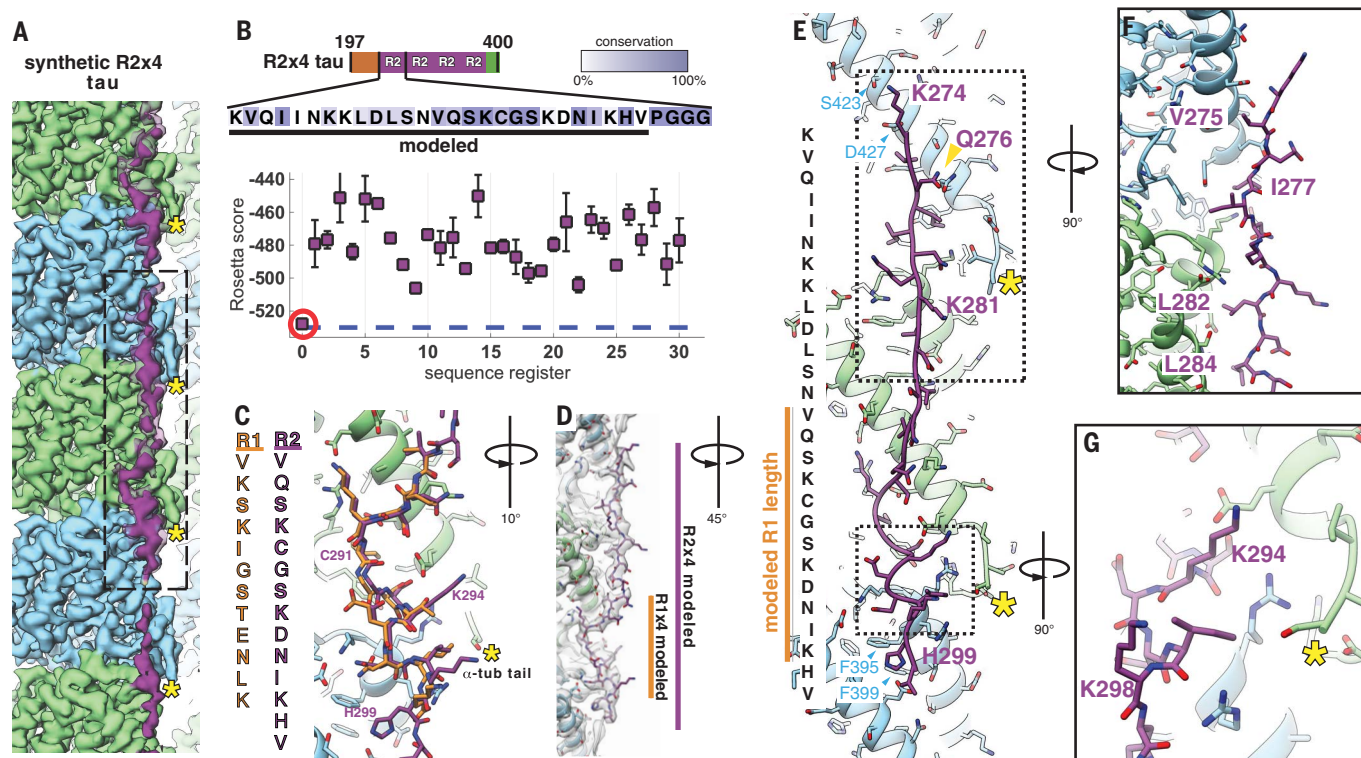


monomers, with a length close to that of the tubulin dimer repeat in the MT lattice (~80 Å). As we found for R1, we discovered a single, strongly preferred register and conformation for R2 with the use of Rosetta (Fig. 3B), regardless of the chosen simulation parameters. Our analyses of R1 and R2 resulted in equivalent sequence registers and virtually identical atomic models at the interdimer interface (Fig. 3, C and D), with conserved residues making critical contacts with tubulin. For two non-

conserved positions, Cys<sup>291</sup>(R2) versus Ile<sup>260</sup>(R1) and Lys<sup>294</sup>(R2) versus Thr<sup>263</sup>(R1) (Fig. 3C), the nature of the interactions is preserved [free cysteines demonstrate strong hydrophobic character (26), and Lys<sup>294</sup> likely interacts with the acidic C-terminal tail (Fig. 3, E and G)]. The identical sequence register and atomic details from two independent maps underscore the robustness of our solution and provide high confidence in the accuracy of our atomic models. The sequence

assignment is further supported by previous gold-labeling experiments on the binding of MAP-2 to MTs (fig. S6, B to D).

Our R2x4 model includes the peptide VQIINKK (not conserved in R1), which as an isolated peptide can bind MTs (6) and promote MT polymerization (23). This R2 peptide localizes to the intradimer interface and is sufficiently close to interact with the β-tubulin C-terminal tail (Fig. 3, E and F). Though tau and kinesin make distinct



**Fig. 3. High-resolution reconstruction of synthetic R2x4 tau on microtubules.** (A) The (3.9-Å) cryo-EM reconstruction of R2x4 is highly similar to that of R1x4 but reveals a longer stretch of ordered density for tau along the MT surface (lower threshold density for tau is shown in transparency). The boxed region marks the footprint of one R2 repeat. (B) Rosetta modeling supports a sequence register (circled) for the R2 sequence binding to tubulin equivalent to that for R1 shown in Fig. 2. All Rosetta simulations were repeated until convergence (100 models per register). (C) Major tau-tubulin interactions at the interdimer cleft are highly similar between the R1 (shown in orange for easier visualization) and R2 (purple) sequences. (D) Extending the model to account for the additional density reveals an almost entire repeat

of tau, spanning three tubulin monomers (centered on  $\alpha$ -tubulin and contacting  $\beta$ -tubulin on either side), with an overall length of  $\sim 80$  Å (the approximate length of a tubulin dimer). (E) Atomic model of the R2 repeat, shown along with the corresponding R2 sequence. The position of the previously studied gold-labeled residue in MAP-2 at the equivalent position according to homology (16) is indicated with an arrowhead and is in very good agreement with our model (see also fig. S6). Boxed-out regions show (F) the hydrophobic packing of tau residues on the MT surface (see also fig. S8) and (G) the positioning of two R2 lysines with potential interactions with the  $\alpha$ -tubulin acidic C-terminal tail. The C termini of tubulin are indicated with yellow asterisks in (A), (C), (E), and (G).

contacts with tubulin, their MT-binding sites partially overlap (fig. S7), which explains why tau binding interferes with kinesin attachment to MTs (27). Residues Val<sup>275</sup>, Ile<sup>277</sup>, Leu<sup>282</sup>, and Leu<sup>284</sup> (Fig. 3F) are buried against the MT surface and tolerate only conservative hydrophobic substitutions in R2, R3, and R4 (fig. S8). Lys<sup>274</sup> and Lys<sup>281</sup> are crucial for tau-MT binding (6). In our model, Lys<sup>274</sup> is close to an acidic patch on the MT formed by Asp<sup>427</sup> and Ser<sup>423</sup> in  $\beta$ -tubulin (Fig. 3, E and G), and Lys<sup>281</sup> is well positioned to interact with the  $\beta$ -tubulin C-terminal tail (Fig. 3E). The highly conserved His<sup>299</sup> in R2 is buried in a cleft formed by  $\beta$ -tubulin residues Phe<sup>395</sup> and Phe<sup>399</sup> (Fig. 3, C and E).

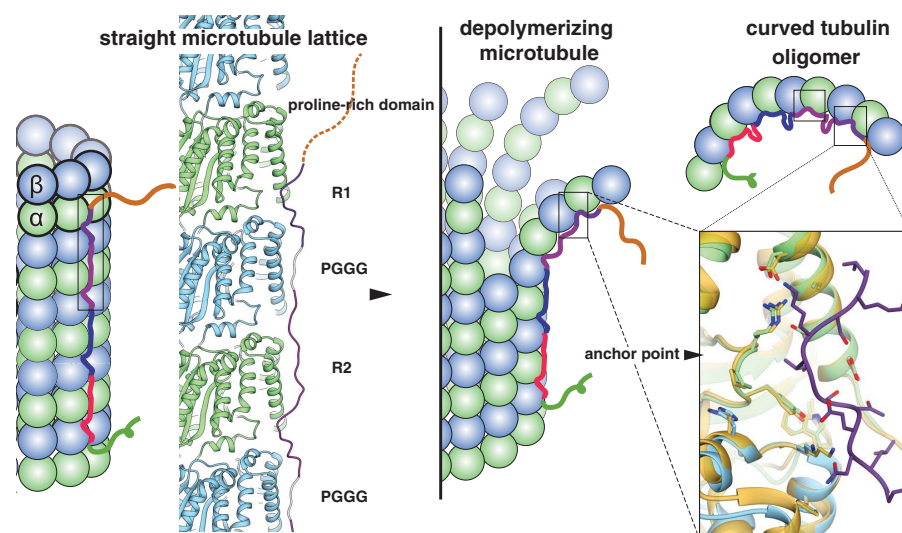
Although we could model most of the residues in a tau repeat, we could not clearly visualize the highly conserved PGGG motif, which would correspond to the region connecting the modeled segments. Figure S9 shows a model of consecutive R1 and R2 binding, on the basis of the atomic models of both repeats connected by an extended PGGG segment, placed into the full-length tau

experimental map. The similarity among repeats, especially R2 and R3, and the geometry of their binding to tubulin strongly support a tandem binding mode of the four repeats along a PF (fig. S8 shows the conserved character of hydrophobic interactions between these tau repeats and tubulin). Comparison of our MT-bound tau structures for R1 and R2 to that of fibrillary tau for R3 and R4 shows that though they are globally very different, they share some similarities in local structure, especially in the conserved, hydrophobic regions (fig. S10).

Our MT-tau structures lead to a model in which each tau repeat has an extended conformation that spans both intra- and interdimer interfaces, centered on  $\alpha$ -tubulin and connecting three tubulin monomers. Extensive modeling on two independently determined reconstructions complemented the experimental map and led to a proposed model of the repeat structure and MT interactions that is supported by sequence conservation and explains previous biochemical observations. The universally conserved Ser<sup>262</sup>, a

major site of phosphorylation, is critically involved in tight contacts with tubulin near a polymerization interface, explaining how its modification interferes with tubulin binding and MT stabilization. The major tau-binding site corresponds to the “anchor point” (28), a tubulin region that is practically unaltered during the structural changes accompanying nucleotide hydrolysis or in a comparison of assembled and disassembled states of tubulin (Fig. 4, right). This finding explains how tau promotes the formation of tubulin rings (29) and copurifies with MTs through tubulin assembly-disassembly cycles. Our structures suggest that all four tau repeats are likely to associate with the MT surface in tandem, through adjacent tubulin subunits along a PF. This modular structure explains how alternatively spliced variants can have essentially identical interactions with tubulin but different affinities according to the number of repeats present. The tandem binding of tau along a PF explains how tau promotes both MT polymerization and stabilization by tethering multiple





**Fig. 4. Model of full-length tau binding to microtubules and tubulin oligomers.** Our structural data lead to a model of tau interaction with MTs in which the four repeats bind in tandem along a PF. We did not observe strong density for the region that would correspond to the PGGG motif, which is modeled in gray for illustrative purposes and must be highly flexible. Tau binding at the interdimer interface, interacting with both  $\alpha$ - and  $\beta$ -tubulin, promotes association between tubulin dimers. The tau-binding site is also the location of the previously identified “anchor point” [rightmost box; orange is bent tubulin (PDB code 4I4T) (31), blue and green are straight tubulin (PDB code 3JAR) (28), and purple corresponds to the tightly bound region of a tau repeat]; thus tau-tubulin interactions are unlikely to change substantially with PF peeling during disassembly (center) or with binding to small, curved tubulin oligomers (top right).

tubulin dimers together across longitudinal interfaces (Fig. 4, left). This mode of interaction explains the reduced off-rate of tubulin dimers in the presence of tau (30) that results in reduced catastrophe frequencies (transitions from assembly to disassembly of MTs) and the loss of dynamics caused by tau and other classical MAPs. Our study does not discard the possibility that other structural elements within tau are involved in additional tubulin interactions, especially if engaging with the unstructured C-terminal tubulin tails. Such potential contacts may contribute to inter-PF and/or inter-MT interactions.

#### REFERENCES AND NOTES

1. C. Conde, A. Cáceres, *Nat. Rev. Neurosci.* **10**, 319–332 (2009).
2. K. S. Kosik, L. D. Orecchio, S. Bakalis, R. L. Neve, *Neuron* **2**, 1389–1397 (1989).

3. A. Himmler, D. Drechsel, M. W. Kirschner, D. W. Martin Jr., *Mol. Cell. Biol.* **9**, 1381–1388 (1989).
4. D. J. Ennulat, R. K. Liem, G. A. Hashim, M. L. Shelanski, *J. Biol. Chem.* **264**, 5327–5330 (1989).
5. K. A. Butner, M. W. Kirschner, *J. Cell Biol.* **115**, 717–730 (1991).
6. B. L. Goode, S. C. Feinstein, *J. Cell Biol.* **124**, 769–782 (1994).
7. M. Hutton et al., *Nature* **393**, 702–705 (1998).
8. M. Hong et al., *Science* **282**, 1914–1917 (1998).
9. A. C. Alonso, T. Zaidi, I. Grundke-Iqbal, K. Iqbal, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5562–5566 (1994).
10. A. C. Alonso, I. Grundke-Iqbal, K. Iqbal, *Nat. Med.* **2**, 783–787 (1996).
11. N. Gustke et al., *FEBS Lett.* **307**, 199–205 (1992).
12. A. W. P. Fitzpatrick et al., *Nature* **547**, 185–190 (2017).
13. D. Eliezer et al., *Biochemistry* **44**, 1026–1036 (2005).
14. X. H. Li, J. A. Culver, E. Rhoades, *J. Am. Chem. Soc.* **137**, 9218–9221 (2015).
15. H. Kadavath et al., *Angew. Chem. Int. Ed. Engl.* **54**, 10347–10351 (2015).

16. J. Al-Bassam, R. S. Ozer, D. Safer, S. Halpain, R. A. Milligan, *J. Cell Biol.* **157**, 1187–1196 (2002).
17. S. Kar, J. Fan, M. J. Smith, M. Goedert, L. A. Amos, *EMBO J.* **22**, 70–77 (2003).
18. M. F. Chau et al., *Biochemistry* **37**, 17692–17703 (1998).
19. L. Serrano, E. Montejo de Garcini, M. A. Hernández, J. Avila, *Eur. J. Biochem.* **153**, 595–600 (1985).
20. H. Kadavath et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7501–7506 (2015).
21. Y. Song et al., *Structure* **21**, 1735–1742 (2013).
22. R. Y. Wang et al., *eLife* **5**, e17219 (2016).
23. D. Panda, B. L. Goode, S. C. Feinstein, L. Wilson, *Biochemistry* **34**, 11117–11127 (1995).
24. J. Biernat, N. Gustke, G. Drewes, E. M. Mandelkow, E. Mandelkow, *Neuron* **11**, 153–163 (1993).
25. M. Hasegawa et al., *J. Biol. Chem.* **267**, 17047–17054 (1992).
26. N. Nagano, M. Ota, K. Nishikawa, *FEBS Lett.* **458**, 69–71 (1999).
27. B. Trinczek, A. Ebner, E. M. Mandelkow, E. Mandelkow, *J. Cell Sci.* **112**, 2355–2367 (1999).
28. R. Zhang, G. M. Alushin, A. Brown, E. Nogales, *Cell* **162**, 849–859 (2015).
29. F. Devred et al., *Biochemistry* **43**, 10520–10531 (2004).
30. B. Trinczek, J. Biernat, K. Baumann, E. M. Mandelkow, E. Mandelkow, *Mol. Biol. Cell* **6**, 1887–1902 (1995).
31. A. E. Prota et al., *Science* **339**, 587–590 (2013).

#### ACKNOWLEDGMENTS

We thank A. Chintangal and P. Tobias for computational support. We also thank P. Grob and D. Toso for EM support. We thank D. Dynerman and N. Grigorieff for help with implementing phase plate support in frealign. We thank B. Greber for help with initial model building.

**Funding:** This work was funded by a BWF collaborative research travel grant (008185) (E.H.K.) and NIHGM grants K99GM124463 (E.H.K.), GM123089 (F.D.), and GM051487 (E.N.). E.N. is a Howard Hughes Medical Institute Investigator. **Author contributions:** Author contributions were as follows: conceptualization, E.H.K., N.M.A.H., and E.N.; methodology, E.H.K., N.M.A.H., F.D., and E.N.; investigation, E.H.K., N.M.A.H., F.D., and E.N.; writing (original draft), E.H.K. and E.N.; writing (review and editing), all authors; funding acquisition, E.N.; resources, S.P.; and supervision, E.N. and K.H.D. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Atomic models are available through the Protein Data Bank (PDB) with accession codes 6CVJ (R1×4) and 6CVN (R2×4); all cryo-EM reconstructions are available through the EMDB with accession codes EMD-7520 (2R tau), EMD-7523 (4R tau), EMD-7522 (full-length tau), EMD-7769 (R1×4 tau), and EMD-7771 (R2×4 tau).

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6394/1242/suppl/DC1  
Materials and Methods  
Figs. S1 to S10  
Table S1  
References (32–46)

1 February 2018; accepted 30 April 2018  
Published online 10 May 2018  
10.1126/science.aat1780

## CYTOMETRY

# Ghost cytometry

Sadao Ota<sup>1,2,3,\*†</sup>, Ryoichi Horisaki<sup>3,4,\*</sup>, Yoko Kawamura<sup>1,2,\*</sup>, Masashi Ugawa<sup>1,\*</sup>, Issei Sato<sup>1,2,3,5</sup>, Kazuki Hashimoto<sup>2,6</sup>, Ryosuke Kamesawa<sup>1,2</sup>, Kotaro Setoyama<sup>1</sup>, Satoko Yamaguchi<sup>2</sup>, Katsuhito Fujiu<sup>2</sup>, Kayo Waki<sup>2</sup>, Hiroyuki Noji<sup>2,7</sup>

Ghost imaging is a technique used to produce an object's image without using a spatially resolving detector. Here we develop a technique we term "ghost cytometry," an image-free ultrafast fluorescence "imaging" cytometry based on a single-pixel detector. Spatial information obtained from the motion of cells relative to a static randomly patterned optical structure is compressively converted into signals that arrive sequentially at a single-pixel detector. Combinatorial use of the temporal waveform with the intensity distribution of the random pattern allows us to computationally reconstruct cell morphology. More importantly, we show that applying machine-learning methods directly on the compressed waveforms without image reconstruction enables efficient image-free morphology-based cytometry. Despite a compact and inexpensive instrumentation, image-free ghost cytometry achieves accurate and high-throughput cell classification and selective sorting on the basis of cell morphology without a specific biomarker, both of which have been challenging to accomplish using conventional flow cytometers.

Imaging and analyzing many single cells holds the potential to substantially increase our understanding of heterogeneous systems involved in immunology (1), cancer (2), neuroscience (3), hematology (4), and development (5). Many key applications in these fields require accurate and high-throughput isolation of specific populations of cells according to information contained in the high-content images. This raises several challenges. First, despite recent developments (6–10), simultaneously meeting the needs of high sensitivity, polychromaticity, high shutter speed and high frame rates, continuous acquisition, and low cost remains difficult. Second, ultrafast and continuous image acquisition subsequently requires computational image reconstruction and analysis that is costly in terms of both time and money (11). Given this, fluorescence imaging-activated cell sorting, for instance, has not been realized yet. Here we show that directly applying machine-learning methods to compressed imaging signals measured with a single-pixel detector enables ultrafast, sensitive, and accurate image-free (without image production), morphology-based cell analysis and sorting in real time, which we call "ghost cytometry" (GC).

In GC, as an object passes through a pseudorandom static optical structure, each randomly

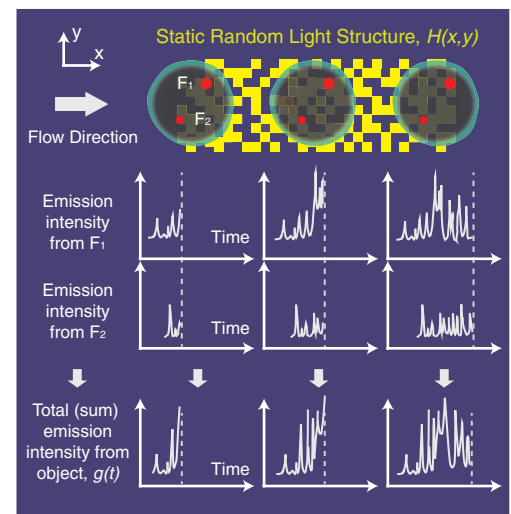
arranged spot in the structure sequentially excites fluorophores at different locations of the object (Fig. 1). These encoded intensities from each fluorophore are multiplexed and measured compressively and continuously as a single temporal waveform measured with a single-pixel detector (Fig. 1, bottom graphs), which, in this work, was a photomultiplier tube (PMT). Assuming the object is in constant unidirectional motion with velocity  $v$ , the signal acquisition is mathematically described as

$$g(t) = \iint H(x, y) I(x - vt, y) dx dy \quad (1)$$

where  $g(t)$  is the multiplexed temporal waveform,  $H(x, y)$  is the intensity distribution of the optical structure, and  $I(x, y)$  is the intensity

distribution of the moving object. Note that  $H$ , acting as a spatial encoding operator, is static, so that no scanning or sequential light projection is needed in GC. We designed a binary random pattern for the optical structure as a simple implementation (figs. S1 to S3). In the measurement process of GC, the object is convolved with the optical structure along the  $x$  direction, and the resultant signals are integrated along the  $y$  direction. In the compressive sensing literature, randomized convolutions are regarded as imaging modalities (12). Given Eq. 1 as a forward model, the image-reconstruction process amounts to solving the inverse problem. This solution can be iteratively estimated by minimizing an objective function that is computed by combinatorial use of the multiplexed temporal waveform,  $g(t)$ , and the intensity distribution of the optical structure,  $H$ . For sparse events in a regularization domain, we can reasonably estimate the moving object from the measured signal,  $g(t)$ , by adopting a compressed-sensing algorithm, which, in this work, was two-step iterative shrinkage/thresholding (TwIST) (13) (as detailed in the supplementary methods). This reconstruction process shares its concept with ghost imaging, in which the original image is computationally recovered after sequentially projecting many random optical patterns onto the object and recording the resultant signals with a single-pixel detector (14–19). Although ghost imaging has attracted considerable attention in the scientific community, the sequential projection of light patterns makes it slow and has hampered its practical use. Even when compressive sensing was used to reduce the time required for the light projections, the method was still slower than conventional arrayed-pixel cameras (18). By contrast, GC does not require any movement of equipment, and the speed of image acquisition increases with the object's motion, up to the high bandwidth of single-pixel detectors. The use of motion thus transforms slow ghost imaging

**Fig. 1. Schematic of the compressive sensing process in GC.** The relative motion of an object across a static, pseudorandom optical structure,  $H(x, y)$ , is used for compressively mapping the object's spatial information into a train of temporal signals.  $F_1$  and  $F_2$  are the representative fluorescent features in the object. According to the object's motion, the spatial modulation of  $H$  is encoded into the temporal modulation of emission intensity from each fluorophore in the object, and their sum  $g(t)$  is recorded with a single-pixel detector, as shown in the bottom graph. In the imaging mode of GC, the object's 2D image can be computationally reconstructed by a combinatorial use of the multiplexed temporal waveform,  $g(t)$ , and the intensity distribution of the optical structure,  $H$ . In the image-free mode of GC, directly applying machine-learning methods to the compressive temporal waveform yields high-throughput, highly accurate, image-free morphology-based cell classification. Schematics are not to scale.



<sup>1</sup>Thinkcyte Inc., 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan. <sup>2</sup>University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan. <sup>3</sup>Japan Science and Technology Agency, PRESTO, 4-1-8 Honcho, Kawaguchi-shi, Saitama 332-0012, Japan. <sup>4</sup>Department of Information and Physical Sciences, Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan. <sup>5</sup>RIKEN AIP, Nihonbashi 1-chome Mitsui Building, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan. <sup>6</sup>Japan Aerospace Exploration Agency, 6-13-1 Osawa, Mitaka-shi, Tokyo 181-0015, Japan. <sup>7</sup>ImPACT Program, Cabinet Office, Government of Japan, Chiyoda-ku Tokyo 100-8914, Japan.

\*These authors equally contributed to this work.

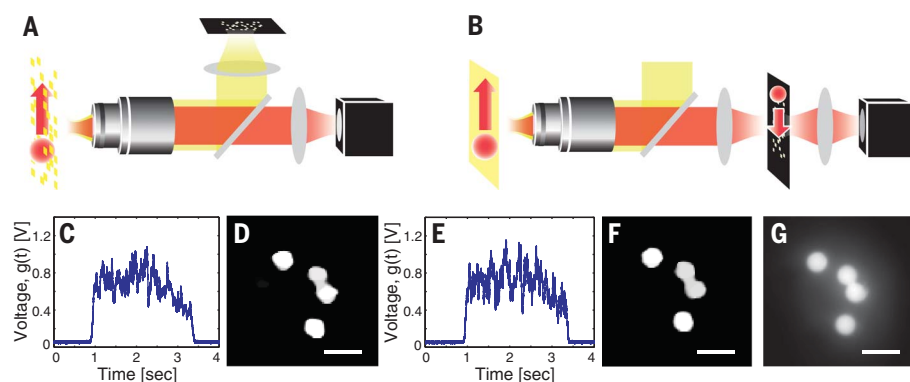
†Corresponding author. Email: sadaota@solab.rcast.u-tokyo.ac.jp



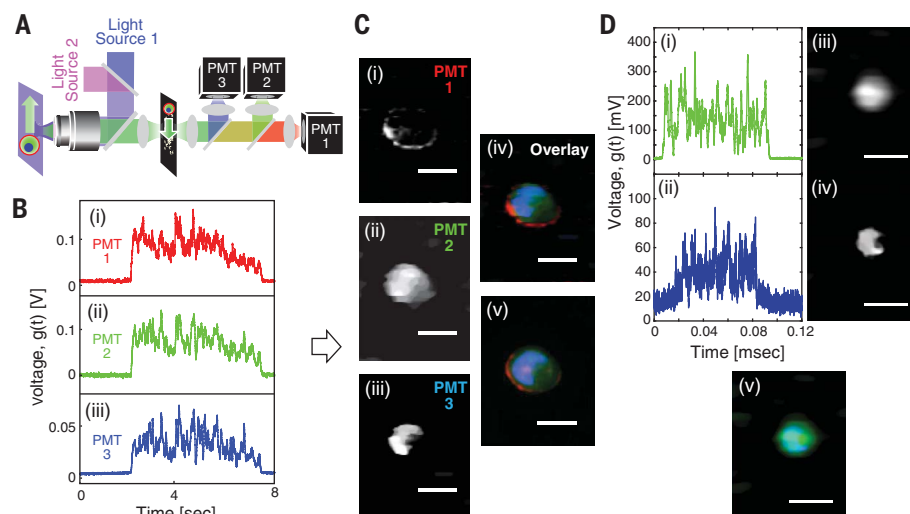
into a practical, ultrafast, and continuous imaging procedure—GC is 10,000 times faster than existing fluorescence ghost imaging (19–21).

As an experimental proof of concept of GC in the imaging mode, we imaged fluorescent beads mounted on a glass coverslip and moved them across a random pattern using an electronic translational stage (Fig. 2, A and B). The beads were kept in focus as they moved in the direction parallel to the row direction of  $H$ . Incorporating the random optical structure in the optical path before or after the sample is mathematically equivalent. This means that GC-based imaging can be experimentally realized by either random structured illumination (SI, shown in Fig. 2A) or random structured detection (SD, shown in Fig. 2B). These configurations correspond to computational ghost imaging (15) and single-pixel compressive imaging (22), respectively. The encoding operator  $H$  in Eq. 1 in the SI mode can be experimentally measured as the excitation intensity distribution in the sample plane; the operator  $H$  in the SD mode can be measured as the pointwise product of the excitation intensity distribution in the sample plane and the transmissibility distribution of the photomask in the conjugated plane between the sample and detector. We experimentally calibrated the exact operator,  $H$ , by placing a thin sheet of a fluorescent polymer in the sample plane and measuring its intensity distribution with a spatially resolving multipixel detector (fig. S1). A blue light-emitting diode (LED) was used as an excitation light source. During the object's motion, the PMT collected the photons that were emitted from the object as temporally modulated fluorescence intensity (Fig. 2, C and E, for the SI and SD modes, respectively). Figure 2, D and F, shows the computationally recovered fluorescence images of multiple beads for each waveform. For comparison, Fig. 2G shows the image acquired with an arrayed detector-based scientific complementary metal-oxide semiconductor (CMOS) camera. The morphological features of the beads are clear, validating GC imaging in both the SI and SD modes.

The simple design of the GC optics means that adding multiple light sources, dielectric mirrors, and optical filters enables multicolored fluorescence imaging with a single photomask (Fig. 3A). To validate GC for cell imaging, we stained MCF-7 cells (a human breast adenocarcinoma cell line) in three colors: The membranes, nuclei, and cytoplasm were stained with red (EpCAM PE-CF594), blue [4',6'-diamidino-2-phenylindole (DAPI)], and green [fixable green (FG)] dyes, respectively. Stained cells were mounted on a glass coverslip. We used a blue continuous-wave laser and ultraviolet LED light sources for exciting the fluorophores. We adopted the SD mode and experimentally estimated the operator,  $H$ , for each excitation light source (fig. S2). In the experiment, we moved the stage on which the coverslip was mounted and measured the temporal signals from each color channel using three PMTs, respectively (Fig. 3, A and B). Figure 3C, (i) to (iii), shows the computationally reconstructed fluorescence images for each color, clearly revealing the



**Fig. 2. Demonstration of motion-based compressive fluorescence imaging.** (A and B) Optical setups for the SI and SD modes, respectively. We moved aggregates of fluorescent beads, represented as a red sphere, on a glass coverslip (not shown) across the optical structure in the direction of the red arrow using an electronic translational stage. (C) In the SI mode, the beads go through the structured illumination according to the motion, resulting in the generation of the temporal waveform of the fluorescence intensity. (D) From this acquired temporal signal, a 2D fluorescence image is computationally reconstructed. (E) In the SD mode, the beads are illuminated by uniform light. A conjugate fluorescence image of the beads then passes through structured pinhole arrays, resulting in the generation of the temporal waveform. (F) From this temporal signal, a 2D fluorescence image is computationally reconstructed. (G) A fluorescence image of the same aggregated beads, acquired with an arrayed-pixel camera. Scale bars, 20  $\mu\text{m}$ .



**Fig. 3. Multicolor and high-throughput fluorescence cell imaging by GC.** (A) An optical setup for the multicolor motion-based compressive fluorescence imaging (SD mode). The setup utilizes a 473-nm-wavelength blue laser and a 375-nm-wavelength ultraviolet LED as excitation light sources coupled by a dichroic mirror (fig. S2) to create a relatively uniform illumination, shown as a purple rectangle through which a cell moves in the direction of the arrow. Cultured MCF-7 cells were used in the experiment, with membranes, cytoplasm, and nuclei fluorescently stained red, green, and blue, respectively. When the labeled cells moved with an electronic translational stage, their conjugated images passed through the optical encoder and generated temporal waveforms. The signals were then split into three-color channels of (i) red, (ii) green, and (iii) blue with dichroic mirrors, and finally recorded by different PMTs. (B) The representative traces recorded by the PMTs. (C) From the temporal signals for each of the three PMTs, fluorescence images of labeled cells were computationally recovered in (i) red, (ii) green, and (iii) blue, respectively. (iv) A pseudocolored multicolor image combined from (i), (ii), and (iii). (v) A multicolor fluorescence image acquired with an arrayed-pixel color camera. (D) Multicolor submillisecond fluorescence imaging of the cells under flow at the throughput rate above 10,000 cells/s. In the experiment, 488-nm-wavelength blue and 405-nm-wavelength violet lasers passed through diffractive optical elements to generate the random structured illumination to the cell stream (fig. S3, SI mode). (i) and (ii) show the green and blue fluorescence signals from the cytoplasm and the nucleus, respectively. From the temporal signals for each PMT, fluorescence images of the labeled cells were computationally recovered in (iii) green and (iv) blue, respectively. (v) The reconstructed multicolor fluorescence image. Scale bars, 20  $\mu\text{m}$ .

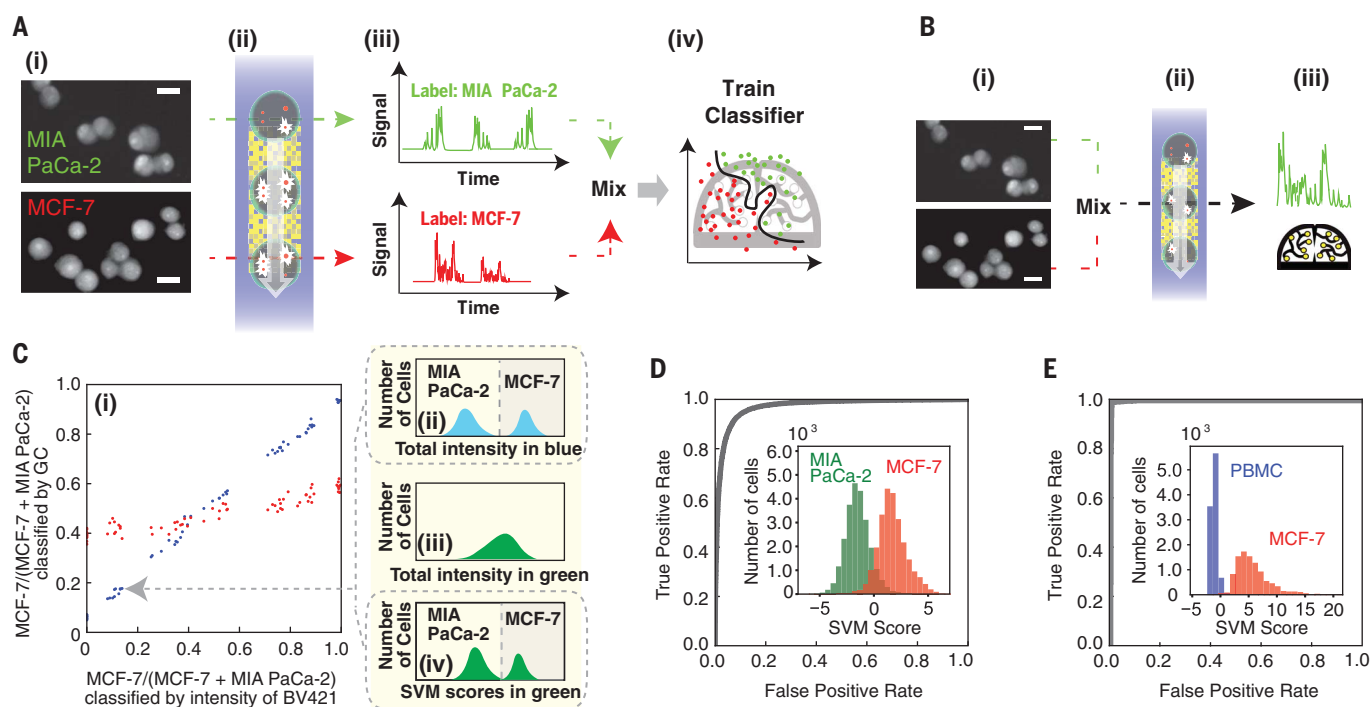
fine features of cellular membranes, cytoplasm, and nuclei. For comparison, Fig. 3C, (iv), shows an overlaid color image, and Fig. 3C, (v), shows a fluorescence image acquired with a conventional color camera (Wraycam SR 130, WRAYMER, Inc., Japan). The average of peak signal-to-noise ratios—which are calculated as eq. S3—of the red, green, and blue channels is 26.2 dB between the GC's reconstructed image, Fig. 3C, (iv), and the camera image, Fig. 3C, (v). These results demonstrate a good performance of multicolor GC imaging in delineating the morphological features of cells.

We also show that GC can achieve fast multi-color continuous fluorescence imaging of flowing cells. We used a flow cell assembly (Hamamatsu Photonics, Japan) for focusing a stream of flowing fluorescent cells in the three-dimensional

(3D) space, so that the cells are focused in the plane perpendicular to the flow that is aligned parallel to the length direction of the encoder  $H$ . Using diffractive optical elements that generate structured illuminations inside the flow cell (fig. S3), we performed continuous spread optical point scans of 100 pixels perpendicular to the flow, corresponding to the image size in the  $y$  direction in the computational reconstruction. Figure 3D, (i) and (ii), shows the temporal waveforms from each color channel of a single MCF-7 cell, with its cytoplasm labeled by FG and its nucleus labeled by DAPI. Fluorescence images were computationally reconstructed for each waveform in (iii) and (iv), respectively. Figure 3D, (v), shows the computationally reconstructed multicolor fluorescence image, with clearly resolved cellular morphological features. The cells were flowed at

a throughput higher than 10,000 cells/s, a rate at which arrayed-pixel cameras such as charge-coupled devices and CMOSs create completely motion-blurred images. The total input excitation intensities of the structured illuminations after the objective lens were  $\sim 58$  and  $\sim 14$  mW for 488- and 405-nm lasers, where those assigned to individual random spots were  $<43$  and  $<10$   $\mu$ W on average, respectively. By designing and adopting the appropriate diffractive optical elements, we created the light pattern with minimal loss, suggesting high sensitivity of GC imaging, calculated as the minimal number of detectable fluorophores (as detailed in the supplementary methods), close to single-molecule levels.

Using the object's motion across the static light pattern  $H$  for optical encoding and sparse sampling, we achieve blur-free, high-frame rate



**Fig. 4. High-throughput and highly accurate fluorescence image-free "imaging" cytometry by GC via direct machine-learning of compressive signals.** (A) The procedure for training a classifier model in GC. (i) Different but morphologically similar cell types (MCF-7 and MIA PaCa-2 cells) were fluorescently labeled: For both cell types, the cytoplasm was stained in green with FG, whereas the membranes of only the MCF-7 cells were stained in blue with BV421-EpCAM. Scale bars, 20  $\mu$ m. (ii) By separately flowing the different cell types through the encoding optical structure used in Fig. 3D at the throughput rate of  $>10,000$  cells/s, (iii) compressive waveforms of each cell type were collectively extracted from the temporally modulated signals of fluorescence intensity. (iv) A library of waveforms labeled with each cell type was used as a training dataset to build a cell classifier. A support vector machine model was used in this work. (B) Procedure for testing the classifier model. (i) The different types of cells were experimentally mixed at a variety of concentration ratios before analysis. (ii) When flowing the cell mixture through the encoder at the same throughput rate, (iii) we applied the trained model directly to the waveform for classifying the cell type. (C) In (i), blue data points are the concentration ratios of MCF-7 cells in each sample estimated by applying the trained SVM-based classification

directly on the waveforms of FG intensity (iv), compared with those obtained by measuring the total intensity of BV421 (ii). Red data points are the concentration ratios of MCF-7 cells estimated by applying the same procedure of SVM-based classification to the total intensity of FG (iii), which we obtained by integrating each GC waveform over time, compared with the results from measurement with BV421 (ii). Seventy samples were measured for a variety of concentration ratios, with each sample comprising 700 tests of randomly mixed cells. The image-free GC results shown with blue data points in (C) reveal a small RMSD of 0.046 from  $y = x$  and an AUC [ROC curve shown in (D)] of 0.971 over about 50,000 cells, even though the morphologies of these cells appear similar to the human eye. (D) Each point on the ROC curve corresponds to a threshold value applied to scores from the trained SVM model, wherein red and green colors in the histogram are labeled according to the intensity of BV421 (inset derived from eq. S4). By contrast, the red data points in (C) reveal inferior classification results, with a large RMSD of 0.289 and poor ROC-AUC of 0.596. (E) When classifying the model cancer (MCF-7) cells against a complex mixture of PBMCs, the ultrafast image-free GC recorded high values of AUC  $\sim 0.998$ , confirming its robust and accurate performance in a case of practical use.



imaging with high signal-to-noise ratio. The frame rate  $r$  and pixel scan rate  $p$  are defined as

$$r = v/(\text{width of } H + \text{width of } I) \quad (2)$$

$$p = v/(\text{single-spot size in } H) \quad (3)$$

where  $I$  is the final image and  $p$  is the inverse of the time taken for a fluorophore to pass over each excitation spot in  $H$ . First, compressive encoding reduces the number of sampling points, defined by the length of  $H$  and required for one frame acquisition, such that we effectively lower the required bandwidth for achieving high frame rates. This reduction is important, especially in ultrafast imaging with a small number of photons, because shot noise increases as the bandwidth increases. This feature allows us to effectively reduce the excitation power for the fluorescence signals to overcome noise. Second, at a sufficient signal-to-noise ratio, GC can take full advantage of the high bandwidth of single-

pixel detectors while  $H$  is temporally static, unlike in the other techniques that temporally modulate the excitation intensity before the object passes through a pixel of excitation light. Consequently, GC yields blur-free images, unless the pixel scan rate is at least two times greater than the bandwidth of the PMT. For example, for an  $H$  spot size of 500 nm and a PMT high bandwidth of 100 MHz, GC provides blur-free images, unless the flow rate surpasses at least 100 m/s.

Beyond a powerful imager, direct analysis of the GC's compressively generated signals enables high throughput and accurate classification of the cell's morphology at considerably lower computational cost, thus leading to the realization of ultrafast fluorescence "imaging"-activated cell sorting (FICS) and analysis. This can be achieved because compressive sensing in GC substantially reduces the size of the imaging data while retaining sufficient information for reconstructing the object image. Although human recognition is

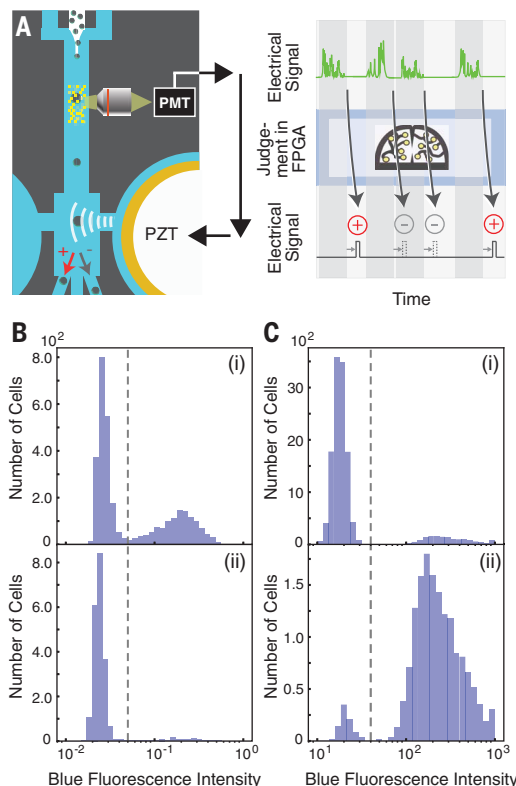
not capable of classifying the waveforms directly, machine-learning methods can analyze the waveforms without image recovery. Here we show that supervised machine learning directly applied to waveforms measured at the rate of ~10,000 cells/s classifies fluorescently labeled cells with high performance, surpassing that of existing flow cytometers and human image recognition (Fig. 4).

Our image-free GC consists of two steps: (i) training and (ii) testing a model of cell classification (Fig. 4, A and B). We first built the model based on the support vector machine (SVM) algorithm (23) by computationally mixing the waveforms of fluorescence signals from different cell types. This training data of waveforms was collected by experimentally passing each cell type separately through the optical encoder (Fig. 4A and table S5). We then tested this model by flowing experimentally mixed cells and classifying the cell types (Fig. 4B). Before the experiment, two different types of cells were cultured, fixed, and fluorescently labeled: MCF-7 cells and MIA PaCa-2 cells (a human pancreatic carcinoma cell line) [Fig. 4A, (i)]. For both cell types, the cytoplasm was labeled in green with FG for classification by image-free GC [Fig. 4A, (i)], whereas the membranes were labeled in blue with BV421-EpCAM (BV421 mouse anti-human CD326 clone EBA-1) only in MCF-7 cells. MIA PaCa-2 cells had only low autofluorescence in the blue channel, providing easily distinguishable contrast in this channel (fig. S5, A to C). We used this to validate the GC's classification that relied on similar cytoplasmic labeling in both cell types. Using the same GC imaging setup (fig. S3), we used both violet and blue continuous-wave lasers for exciting these fluorophores while a digitizer (M4i.4451, Spectrum, Germany) recorded the resultant signals from the PMTs. We developed a microfluidic system to spatially control the position of the stream of cells with respect to the random optical structure, corresponding to the cells' positions in the reconstructed images (fig. S4). Using this optofluidic platform, we collected waveforms of green fluorescence intensity from the cytoplasm for each cell type. From this training dataset, we then built SVM-based classifiers with no arbitrary feature extraction. To test this trained classifier, we introduced a series of solutions containing a combination of different cell types mixed at various concentration ratios. Each classifier then identified ~700 waveforms of the mixed cells as a single dataset and estimated the concentration ratios for each. We used a combination of MCF-7 and MIA PaCa-2 cells, so that the classification results could be quantitatively scored by measuring the total fluorescence intensity of BV421 at the membrane of MCF-7 cells [Fig. 4C, (ii), and fig. S5].

A plot of the concentration ratio of MCF-7 and MIA PaCa-2 cells measured by the blue fluorescence intensity versus that measured by applying the model to the green fluorescence waveform gives a line on the diagonal, with a small root mean square deviation (RMSD) of 0.046 from  $y = x$ . Using the BV421 measurement to evaluate the GC-based classification of ~49,000 mixed cells

### Fig. 5. Demonstration of machine learning-based FICS.

(A) A microfluidic device consists of three functional sites: A flow stream of cells is first focused by 3D hydrodynamic flow-focusing (top left), then experiences the random structured-light illumination (right), and finally arrives at a sorting area (bottom left). Upon sorting action, a PZT actuator driven by input voltages bends to transversely displace a fluid toward the junction for sorting the targeted cells into a collection outlet. For real-time classification and selective isolation of the cells (right), analog signals measured at PMTs are digitized and then analyzed at a FPGA in which we implemented the trained SVM-based classifier. When a classification result is positive, the FPGA sends out a time-delayed pulse that consequently actuates the PZT device. Experiments were performed at a throughput rate of ~3000 cells/s. The cytoplasm of all MIA PaCa-2, MCF-7, and PBMC cells, were labeled in green with FG. The membranes of MCF-7 cells in (B) and the cytoplasm of MIA PaCa-2 cells in (C) were labeled in blue with BV421-conjugated EpCAM antibodies and anti-pan cytokeratin primary and AF405-conjugated secondary antibodies, respectively. (B) Accurate isolation of MIA PaCa-2 cells against morphologically similar MCF-7 cells. GC directly classified the green fluorescence waveforms without image reconstruction. (i) is a histogram of maximum blue fluorescence intensity measured for the original cell mixture, showing a purity of 0.626 for the MIA PaCa-2 cells, whereas (ii) is a histogram for the same mixture after we applied FICS, showing a purity of 0.951. A dashed line corresponding to a threshold value of 0.05 was used to distinguish the populations of the two cell types. (C) Accurate isolation of model cancer (MIA PaCa-2) cells against a complex mixture of PBMCs. (i) is a histogram of maximum blue fluorescence intensity measured for the original cell mixture, showing a purity of 0.117 for the MIA PaCa-2 cells, whereas (ii) is a histogram for the same mixture after we applied FICS, showing a purity of 0.951. A dashed line corresponding to a threshold value of 40 was used to distinguish the populations of the two cell types.



gave an AUC [area under a receiver operating characteristic (ROC) curve] of 0.971 (Fig. 4D), confirming that cell classification by GC is accurate. Each point on the ROC curve corresponds to a threshold value applied to the score obtained from the trained SVM model (eq. S4), where red and green colors in the histogram are labeled according to the intensity of BV421 (Fig. 4D, inset, and fig. S5). To confirm that the high performance of GC is due to the spatial information encoded in the waveforms, we applied the same procedure of SVM-based classification to the total green fluorescence intensity obtained by integrating each GC waveform over time [Fig. 4C, (iii), and fig. S5D]. The results, shown as red data points in Fig. 4C, gave a poor ROC-AUC of 0.596 and a large RMSD of 0.289 from  $y = x$  and thus show little contribution of the total fluorescence intensity to the high performance of GC. In addition, by computing a simple linear fit (fig. S6), we confirmed that the SVM-based classification consistently retains its accuracy over a wide range of concentration ratios. Therefore, image-free GC is an accurate cell classifier even when the targeted cells are similar in size, total fluorescence intensity, and apparent morphological features and are present in mixtures at different concentrations. Indeed, in the absence of molecule-specific labeling, classifying such similar cell types has been a considerable challenge for existing cytometers and even for human recognition.

Besides classifying two cell types that share similar morphology, GC can accurately classify a specific cell type from a complex cell mixture at high throughput. Such technology is important, for example, in detecting rare circulating tumor cells in the peripheral blood of patients (24–27). Here we applied the workflow of the image-free GC for classifying model cancer cells (MCF-7) from peripheral blood mononuclear cells (PBMCs; Astarte Biologics, Inc.), a heterogeneous population of blood cells including lymphocytes and monocytes. Again, the cytoplasm of all the cells was labeled in green with FG for classification by image-free GC, whereas the membranes of only MCF-7 cells were labeled in blue with BV421-EpCAM to validate the GC classification result. We first trained the classifier SVM model by experimentally collecting the green fluorescence waveforms for labeled MCF-7 and PBMC cells and computationally mixing them. We then tested this model by flowing experimentally mixed cells and classifying their cell types one by one. All signals were measured at a throughput of greater than 10,000 cells/s by using the same experimental setup as used previously (fig. S3). For training the model, 1000 MCF-7 cells and 1000 PBMCs were used. For testing the model, 1000 cells from a random mixture of MCF-7 cells and PBMCs were used. After performing cross-validations 10 times, the AUC recorded 0.998 for the SVM-based classifier of the GC waveforms, and Fig. 4E shows one of the ROC curves, proving the ability of ultrafast and accurate detection of the specific cell type from the complex cell mixture.

Reducing the data size by compressive sensing and avoiding image reconstruction in GC shortens

the calculation time required for classifying the single waveform. By combining this efficient signal processing with a microfluidic system, we finally realized ultrafast and accurate cell sorting on the basis of real-time analysis of imaging data (Fig. 5A). Here we demonstrate the ability of FiCS to isolate a specific cell population from another population of morphologically similar cells as well as a complex cell mixture at high throughput and high accuracy. In a microfluidic device made of polydimethylsiloxane (Fig. 5A, left), we designed three functional sites (fig. S7A). First, a flow stream of cells was focused into a tight stream by a 3D hydrodynamic flow-focusing (28, 29) structure (fig. S7B). The cells then experienced the random structured-light illumination of GC and finally arrived at a junction where sorting occurs. A piezoelectric (PZT) actuator was connected to this junction through a channel directed in a direction perpendicular to the main flow stream (fig. S7C). For sorting action, this actuator, driven by an input voltage, bends and transversely displaces a fluid toward the junction to sort the targeted cells into a collection outlet. As shown in Fig. 5A and fig. S7D, for the real-time classification and selective isolation of the cells, their fluorescence signals, recorded as analog voltages by PMTs, were digitized by an analog-to-digital converter and then analyzed by a field-programmable gate array (FPGA) in which we implemented the SVM-based classifier in advance. When the FPGA classifies the cell of interest as positive, it sends out a time-delayed pulse that consequently drives the PZT actuator in the chip. The computation time in the FPGA for classifying each compressive waveform was short enough ( $<10 \mu\text{s}$ ) to enable reproducible sorting. Throughout the experiment, the width of GC's waveform was maintained for about 300  $\mu\text{s}$ , corresponding to a throughput of  $\sim 3000$  cells/s. After measuring the green fluorescence waveforms of positive and negative cells with their labels made by the maximum blue fluorescence intensity, we built the classifier model in a computer offline and implemented it in the FPGA. In the experiment, the cytoplasm of all the MIA PaCa-2 cells, MCF-7 cells, and PBMCs was labeled in green with FG; in addition, the membranes of MCF-7 cells in Fig. 5B and the cytoplasm of MIA PaCa-2 cells in Fig. 5C were labeled in blue with BV421-conjugated EpCAM antibodies and anti-pan cytokeratin primary and AF405-conjugated secondary antibodies, respectively.

We first show that integrated FiCS enables accurate isolation of MIA PaCa-2 cells from MCF-7 cells, which are similar in size, total fluorescence intensity, and apparent morphology. Two hundred waveforms of MIA PaCa-2 cells and 200 of MCF-7 cells were used for training the SVM model. When we mixed the two cell types and then measured their maximum blue fluorescence intensity with a homebuilt flow cytometer (analyzer), two distinct peaks corresponding to the two cell types appeared in the histogram [Fig. 5B, (i)]. After we applied the machine learning-driven FiCS to the same cell mixture by classifying the green fluorescence waveforms, we measured the maximum

blue fluorescence intensity of the sorted mixture in the same manner. As a result, the peak at stronger intensity, corresponding to MCF-7 cells, disappeared, and the purity of MIA PaCa-2 cells increased from 0.625 to 0.951 [compare Fig. 5B, (i) and (ii)]. We thus confirmed that, with just the use of cytoplasmic staining (FG) alone, which does not specifically label the targeted molecules, FiCS can recognize and physically isolate the apparently similar cell types on the basis of their morphologies with high accuracy and throughput.

Finally, we show that FiCS can accurately enrich MIA PaCa-2 cells against the complex mixture of PBMCs. Two hundred waveforms of MIA PaCa-2 cells and 200 of PBMCs were used for training the SVM model. When we mixed the two cell types and then measured their maximum blue fluorescence intensity with a homebuilt flow cytometer (analyzer), the peak at stronger intensity, corresponding to the population of MIA PaCa-2 cells, was relatively small [Fig. 5C, (i)]. After we applied FiCS to the same cell mixture by classifying the green fluorescence waveforms, we measured the maximum blue fluorescence intensity of the sorted mixture in the same manner. As a result, the purity of MIA PaCa-2 cells increased from 0.117 to 0.951 [compare Fig. 5C, (i) and (ii)]. We thus confirmed that FiCS can substantially enrich the model cancer cells against the background of a complex cell mixture without any specific biomarker at high accuracy and throughput.

Recent research has extensively used imaging flow analyzers for the detection and/or characterization of critical cells in various fields, including oncology, immunology, and drug screening (30, 31). GC's ability to notably increase the analysis throughput and selectively isolate cell populations according to high-content information in real time will lead to integration of the morphology-based analysis with comprehensive, downstream “-omics” analyses at single-cell levels. Beyond conventional image generation and processing that rely on the limited knowledge and capability of humans, we anticipate that the idea of applying machine-learning methods directly to compressive modalities will have broad applicability for the real-time application of high-quantity and high-dimensional data.

## REFERENCES AND NOTES

- O. Thauinat et al., *Science* **335**, 475–479 (2012).
- Y. Urano et al., *Nat. Med.* **15**, 104–109 (2009).
- E. A. Susaki et al., *Cell* **157**, 726–739 (2014).
- L. Samsel, J. P. McCoy Jr., *J. Immunol. Methods* **423**, 52–59 (2015).
- J. A. Knoblich, *Nat. Rev. Mol. Cell Biol.* **11**, 849–860 (2010).
- K. Goda, K. K. Tsia, B. Jalali, *Nature* **458**, 1145–1149 (2009).
- L. Gao, J. Liang, C. Li, L. V. Wang, *Nature* **516**, 74–77 (2014).
- E. D. Diebold, B. W. Buckley, D. R. Gossett, B. Jalali, *Nat. Photonics* **7**, 806–810 (2013).
- A. Orth, D. Schaak, E. Schonbrun, *Sci. Rep.* **7**, 43148 (2017).
- B. Guo et al., *PLOS ONE* **11**, e0166214 (2016).
- Y. Han, Y. Gu, A. C. Zhang, Y.-H. Lo, *Lab Chip* **16**, 4639–4647 (2016).
- J. Romberg, *SIAM J. Imaging Sci.* **2**, 1098–1128 (2009).
- J. M. Bioucas-Dias, M. A. T. Figueiredo, *IEEE Trans. Image Process.* **16**, 2992–3004 (2007).
- Y. Bromberg, O. Katz, Y. Silberberg, *Phys. Rev. A* **79**, 053840 (2009).



15. J. H. Shapiro, *Phys. Rev. A* **78**, 061802 (2008).
16. A. Gatti, E. Brambilla, M. Bache, L. A. Lugiato, *Phys. Rev. Lett.* **93**, 093602 (2004).
17. B. Sun *et al.*, *Science* **340**, 844–847 (2013).
18. O. Katz, Y. Bromberg, Y. Silberberg, *Appl. Phys. Lett.* **95**, 131110 (2009).
19. N. Tian, Q. Guo, A. Wang, D. Xu, L. Fu, *Opt. Lett.* **36**, 3302–3304 (2011).
20. M. Tanha, S. Ahmadi-Kandjani, R. Kheradmand, H. Ghanbari, *Eur. Phys. J. D* **67**, 44 (2013).
21. V. Studer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E1679–E1687 (2012).
22. M. F. Duarte *et al.*, *IEEE Signal Process. Mag.* **25**, 83–91 (2008).
23. V. N. Vapnik, *The Nature of Statistical Learning Theory* (Springer, 1995).
24. M. Takao, K. Takeda, *Cytometry A* **79A**, 107–117 (2011).
25. W. Sun *et al.*, *PLOS ONE* **8**, e75865 (2013).
26. B. M. Dent *et al.*, *Int. J. Cancer* **138**, 206–216 (2016).
27. Z. Liu *et al.*, *Sci. Rep.* **6**, 39808 (2016).
28. Y. Q. Wang, J. Y. Wang, H. L. Chen, Z. C. Zhu, B. Wang, *Microsyst. Technol.* **18**, 1991–2001 (2012).
29. S. H. Cho, C. H. Chen, F. S. Tsai, J. M. Godin, Y.-H. Lo, *Lab Chip* **10**, 1567–1573 (2010).
30. A. M. Khalil, J. C. Cambier, M. J. Shlomchik, *Science* **336**, 1178–1181 (2012).
31. M. Doan *et al.*, *Blood* **130**, 1437 (2017).
32. S. Ota *et al.*, Dataset for "ghost cytometry." Zenodo (2018).

#### ACKNOWLEDGMENTS

**Funding:** This work was mainly supported by JST-PRESTO, Japan, grant numbers JPMJPR14F5 to S.O., JPMJPR17PB to R.H., and JPMJPR1302 to I.S. and partially supported by funds of a visionary research program from Takeda Science Foundation, and the Mochida Memorial Foundation for Medical and Pharmaceutical Research. The work is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). S.Y., K.F., and K.W. are members of the Department of Ubiquitous Health Informatics, which is engaged in a cooperative program between the University of Tokyo and NTT DOCOMO, Inc.

**Author contributions:** S.O., R.H., Y.K., and M.U. contributed equally to this work. R.H. and S.O. conceived and designed the concepts, experiments, data analysis, and overall research. S.O., M.U., and K.H. developed the setups and performed experiments of optical imaging. R.H. developed algorithms for the imaging recovery and data analysis, and M.U., R.K., K.S., and S.O. modified and used them for the cell analysis with the strong support of I.S. Y.K. developed and performed microfluidic cell sorting with the support of R.K. and S.O. M.U. and S.O. developed and performed experiments of image-free GC analysis. S.O., S.Y., Y.K., K.F., I.S., and K.W. designed

the experiments of detecting cancer cells in blood. S.O., R.H., I.S., and H.N. supervised the work. S.O., R.H., Y.K., and M.U. wrote the manuscript with the input of the other authors. **Competing interests:** S.O., R.H., and I.S. are the founders and shareholders of Thinkcyte, Inc., a company engaged in the development of the ultrafast imaging cell sorter. S.O., R.H., Y.K., I.S., K.H., S.Y., K.F., and K.W. are inventors on patent applications submitted by the University of Tokyo and Osaka University covering the motion-based ghost imaging as well as image-free morphology analysis. **Data and materials availability:** Original measurement data and codes for analysis are available in the supplementary materials and are deposited in Zenodo (32), a repository open to the public.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/360/6394/1246/suppl/DC1](http://www.sciencemag.org/content/360/6394/1246/suppl/DC1)  
Materials and Methods  
Figs. S1 to S7  
References (33–40)  
Movie S1

25 February 2017; resubmitted 10 March 2018  
Accepted 14 May 2018  
10.1126/science.aan0096

# 2018 Cottrell Scholars

Through its Cottrell Scholar program, Research Corporation for Science Advancement nurtures outstanding teacher-scholars recognized for high-quality research as well as academic leadership skills. In addition to the 24 new scholars awarded \$100,000 each this year, three accomplished Cottrell Scholars have received RCSA's TREE Award (Transformational Research and Excellence in Education). The TREE Award celebrates the integration of outstanding research and science education at America's universities and colleges.

## 2018 Cottrell Scholar Awardees



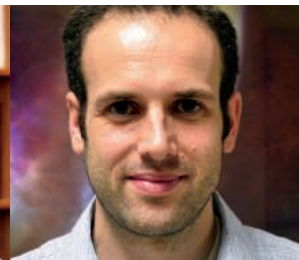
**ASHLEIGH E. BABER**  
Chemistry, James Madison University  
*Role of Surface Modifications on the Selectivity of Titania/Gold Inverse Model Catalysts*



**LOUISE CHARKOUDIAN**  
Chemistry, Haverford College  
*Capturing Transient Interactions of Biosynthetic Proteins to Access New Chemical Diversity*



**NATHANIEL CRAIG**  
Physics, University of California, Santa Barbara  
*New Approaches to the Hierarchy Problem and Undergraduate Education*



**CLAUDE-ANDRE FAUCHER-GIGUERE**  
Astronomy, Northwestern University  
*The Physics, Observational Signatures, and Consequences of Galactic Winds Driven by Active Galactic Nuclei*



**EVAN N. KIRBY**  
Astronomy, California Institute of Technology  
*Heavy Metal: Identifying the Sites of Nucleosynthesis*



**TIM KOWALCZYK**  
Chemistry, Western Washington University  
*Mechanism and Computational Design of Energy Storage and Release in Molecular Photoswitches*



**GARRET M. MIYAKE**  
Chemistry, Colorado State University  
*Design Principles of Strongly Reducing Visible-Light Organic Photoredox Catalysts*



**KATER MURCH**  
Physics, Washington University in St. Louis  
*Exploring Thermodynamics of Single Quantum Systems with Continuous Measurement*



**SEAN T. ROBERTS**  
Chemistry, University of Texas at Austin  
*Tracking Singlet Fission with Ultrafast Time-resolved Microscopy and A Focused Research Experience for Community College Students*



**EDUARDO ROZO**  
Physics, University of Arizona  
*Shining a Light on the Dark Energy*



**LISA M. RYNO**  
Chemistry, Oberlin College  
*Investigating the Impact of *flaA* Overexpression on the Formation and Composition of *E. coli* Biofilms at Different Temperatures*



**ALEXANDER M. SPOKOYNY**  
Chemistry, University of California, Los Angeles  
*Atomically-Precise Nanomaterials Based on Inorganic Clusters*

## 2018 TREE Awardees



**MARTIN GRUEBELE**  
Chemistry, University of Illinois, Urbana-Champaign  
*Cottrell Scholar since 1995*



**TERI W. ODOM**  
Chemistry, Northwestern University  
*Cottrell Scholar since 2005*



**GEORGE C. SHIELDS**  
Chemistry, Furman University  
*Cottrell Scholar since 1994*





**JARROD B. FRENCH**  
Chemistry, SUNY-Stony Brook University  
*Structural Dynamics of Photoactive  
Proteins, and Crowdsourcing Structural  
Biology*



**JOHN G. GIBBS**  
Physics, Northern Arizona University  
*Investigating Shape-Dependent  
Emergent Collective Behavior  
in Artificial Active Matter Systems*



**HAYK HARUTYUNYAN**  
Physics, Emory University  
*Generation and Dynamics of Hot  
Electrons in Metal-semiconductor  
Hybrid Nanosystems*



**A. MEREDITH HUGHES**  
Astronomy, Wesleyan University  
*The Last Gasp of Planet Formation:  
Gas and Dust in Debris Disks*



**KERSTIN NORDSTROM**  
Physics, Mount Holyoke College  
*Flow in Amorphous Systems:  
Understanding Dynamics Across Scales*



**STELLA OFFNER**  
Astronomy, University of Texas at Austin  
*Unveiling the Life Cycle of Stars  
through Cosmic Time and Enhancing  
Inclusivity in Astronomy*



**FEDERICO A. RABUFFETTI**  
Chemistry, Wayne State University  
*Crystallochemical Principles of  
Energy Management in Solid-State  
Photoluminescent Materials*



**CHAD M. RISKO**  
Chemistry, University of Kentucky  
*High Energy Density Metal Oxides  
for Energy Storage: In Silico  
Electrochemistry to Control Interface  
Chemistry*



**GRACE Y. STOKES**  
Chemistry, Santa Clara University  
*Thermodynamic Studies of  
Peptoid-Lipid Interactions*



**BRETT VANVELLER**  
Chemistry, Iowa State University  
*Surfing the Excited-state Energy  
Surface towards New Photo  
Reaction Strategies*



**ABIGAIL G. VIEREGG**  
Physics, University of Chicago,  
*An Interferometric Technique  
for Discovering the Highest Energy  
Neutrinos*



**LUISA L. WHITTAKER-BROOKS**  
Chemistry, University of Utah  
*Manipulating Spin and Charge Transport  
Dynamics in Organic-Inorganic  
Interfaces and Addressing the Disparity  
Between General Chemistry and Reality*

For additional information visit  
[www.rescorp.org](http://www.rescorp.org) or call 520.571.1111.



**RESEARCH CORPORATION**  
*for SCIENCE ADVANCEMENT*  
*A foundation dedicated to science since 1912.*

ADVERTISEMENT



## Institute for Advanced Study fosters spirit of collaboration

*Since its founding in 2015, the Institute for Advanced Study (IAS), based at the City University of Hong Kong (CityU), has offered academics and students an innovative, intellectual platform to encourage and foster groundbreaking science research.*

**U**nder the leadership of CityU president Way Kuo and vice president Jian Lu, the institute aspires to be an international hub of research excellence where science, technology, and engineering thought leaders can focus on finding solutions to pressing, real-world problems.

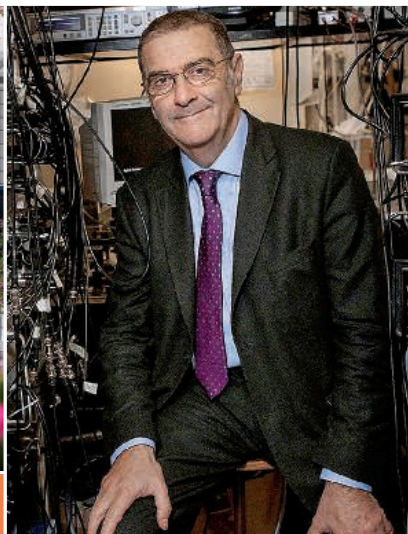
In the past three years, the institute's world-renowned, interdisciplinary faculty has grown from 10 to 27, and in March of this year, Nobel prize-winning physicist Serge Haroche joined IAS as its chairman. All senior faculty work with and mentor other research fellows and students at CityU.

"Under Professor Haroche's leadership, IAS will further its dual mission of serving as the perfect platform for spearheading the university's drive for excellence in interdisciplinary research that benefits society, and of building a regional and global intellectual center," explains Jacob Huang, executive director of IAS and chairman of the materials science and engineering department at CityU. Before joining IAS, Huang was vice-president of National Sun Yat-sen University in Taiwan and program manager of Taiwan's National Science Council.

The institute's model of learning and research is rooted in a spirit of collaboration and knowledge exchange. Since its founding, IAS has organized more than 40 lectures, workshops, and conferences. Over 80 internationally recognized scholars and academicians have been invited to exchange ideas at CityU, and more than 3,000 participants have attended IAS events.

According to Huang, IAS aspires to provide an opportunity for the best minds in science to pursue "curiosity-driven" research.

"We want to create an institute where the goals are to seek truth, to advance knowledge, and to better humanity—to conduct unfettered research based on free and deep thinking," says Huang. "A key part of the institute's work is to bring scientists together to share research achievements, exchange ideas, and establish relationships for future research projects."



Serge Haroche, IAS chairman



Jacob Huang, IAS executive director and chairman of materials science and engineering department, CityU

### Addressing key scientific issues

In June 2018, IAS will host the Materials Summit, where 10 top materials scientists will explore and discuss the future trends in advanced materials science and engineering. This event will run concurrently with the 14th International Conference on Nanostructured Materials (NANO 2018), and will take place on June 27.

In July 2018, the institute will bring together more than 20 scientists working in different areas of biomedical research for the IAS Extended Workshop on Genomes, Cells, and Mathematics. This seminar will provide a platform for international experts to exchange views on genome-sequencing studies and encourage attendees to develop a deeper understanding of gene function as it relates to chromosomal interactions and disease.

### Growing on the global stage

IAS also aims to raise the global profile of the science and engineering community at CityU and other academic institutions in Hong Kong and Greater China.

"Our chairman, Professor Haroche, is passionate about leading IAS to establish closer collaborations with leading academicians across the world," says Huang.

### Previous events: Nobel Science Week and Workshop on Gene Dynamics and Chromosomes

In November 2017, IAS organized its first Nobel Science Week, featuring three Nobel Laureates in chemistry and physics.

"The workshop provided a great opportunity for CityU students to meet Nobel Prize winners; it was aimed at boosting academic exchanges and inspiring the CityU community to pursue knowledge and innovative research," says Huang.

IAS also plans events that encourage conversations about how science relates to broader social issues. In August 2016, it organized the five-day Workshop on Gene Dynamics and Chromosomes, where experts in the fields of biology, chemistry, computer science, mathematics, and biomedical science delivered 16 lectures on recent research developments.

"Conversations ranged from the purely scientific to the commercial, legal, and social challenges in this field," says Huang.

### Prioritizing mentorships

In addition to real-time conferences, students and academics at IAS can find event details and videos posted on the institute's website. Students are also encouraged to work closely with senior staff to find inspiration for their research work.

"IAS is committed to fostering an environment that provides ample opportunities for mentorship. We encourage one-on-one meetings between senior staff and students, and also run regular lectures, workshops, activities, and events to stimulate original research," says Huang.





# Exceptional scientists wanted

## Present your work to the world

Are you a representative of the upcoming generation of thought leaders in your field? Together we look forward to your application for the new Sartorius & Science Prize for Regenerative Medicine & Cell Therapy.

## Apply now!

[www.passionforscience.com/prize](http://www.passionforscience.com/prize)



The Sartorius & Science  
Prize for Regenerative  
Medicine & Cell Therapy

Awarded by



sartorius

Science



## Translating big data: The proteomics challenge

Getting the most out of protein-related information depends on teamwork among scientists around the world, and that involves sharing large datasets. Simply passing big data back and forth is not a problem, however—the main obstacle is sharing that data in a way that other scientists can use it. Building software that can interpret information from different experiments and equipment remains complicated; likewise, exploring and analyzing large datasets from proteomics experiments even from one lab requires software that is most often developed in-house. **By Mike May**

**T**he roughly 20,000 protein-encoding genes in humans should make at least 20,000 proteins. However, modifications create more proteins—maybe many more. As of April 4, 2018, the Human Proteome Map included 30,057 proteins. Combining so many molecules with analytical technology such as mass spectrometry (MS), which explores many fine details, creates “big data.” The masses of complex information being uncovered about proteins are often so large that they require teams of scientists just to work on one dataset.

Yet despite the size and complexity of these datasets, sharing them is becoming expected. According to Joshua Coon, director of the NIH National Center for Quantitative Biology of Complex Systems at the University of **Wisconsin-Madison**, raw data from proteomics studies is usually available today in a database, or the authors will send it upon request.

“That was not the case 10 years ago, but attitudes have changed,” he says. The proteomics community—and increasingly the scientific community in general—realizes that data transparency improves the level of trust among researchers, even with those in different fields.

### Struggles of sharing

It’s easier than ever to produce large amounts of protein-related data—but it’s not always easy to share that data in the

most helpful way. “In a couple of days, a protein scientist can create a terabyte of data, which is hard to transfer or visualize,” says Gary Kruppa, vice president of business development for proteomics at **Bruker Daltonics** in Billerica, Massachusetts. “A month of data gets hard to even store.”

The challenge of sharing such copious amounts of data depends on the high number of possible approaches to doing so, and on the need to provide sufficient experimental and biological metadata. If a scientist just wants to share raw data from a proteomics experiment, along with a little background on what it represents and some results, “it’s very straightforward,” says Juan Antonio Vizcaino, proteomics team leader at the **European Bioinformatics Institute (EMBL-EBI)** in Cambridge, United Kingdom.

The challenges start to mount as more information is being shared among more scientists. Just dumping information into a database, for example, isn’t enough. “Someone has to pay attention to ensure that the data is of sufficient quality [for other scientists] to be able to do something with it,” says Andreas Huhmer, global marketing director for proteomics solutions at **Thermo Fisher Scientific** in San Jose, California. Plus, data cannot be easily uploaded to a database and retrieved unless it’s in a standardized format.

The method of analyzing this data also impacts the conclusions drawn from it. “There are countless ways to analyze proteomics data these days, leading to subjective interpretations of data,” explains Andrew Webb, acting head for the division of systems biology and personalized medicine at the **Walter and Eliza Hall Institute of Medical Research** in Parkville, Australia.

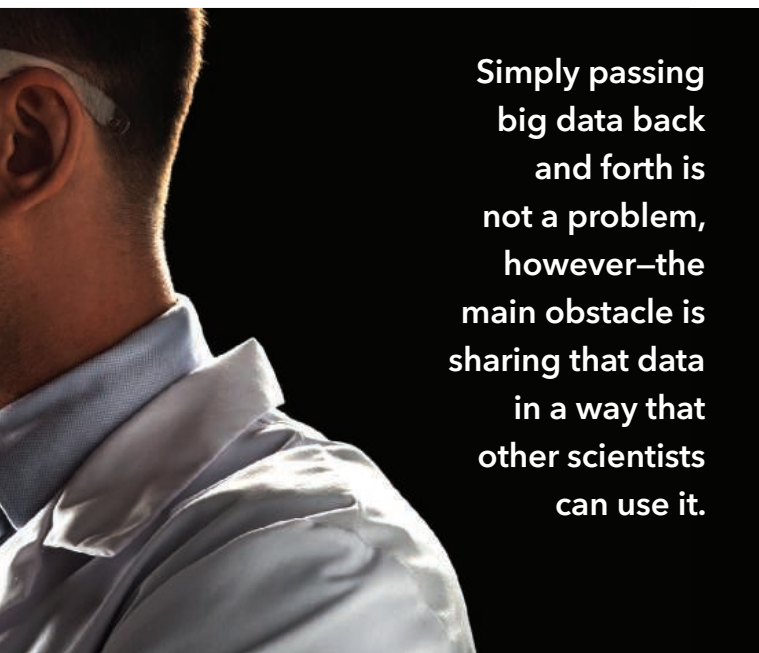
Other experts agree that analyzing data still challenges proteomics scientists. “How we effectively and efficiently turn raw data into something meaningful—even in one lab—is the first big hurdle,” notes James Langridge, director of health sciences at **Waters** in Manchester, United Kingdom.

Even when scientists agree on standardized formats for the data and ways to analyze it, more work lies ahead. For one thing, the

### Upcoming features

Genomics: Pharmacogenomics—September 28 ■ Animal Models: Better Mouse Models—October 12 ■ Tissue/Cell Culture: Expression Systems—November 16





data standards must be updated as needed. In addition, sharing even the biggest proteomics datasets will fall short. "To maximize the scientific knowledge that can be derived from proteomics datasets, that knowledge should be systematically integrated with its genomic counterparts—the genome and transcriptome," says Henry Rodriguez, director of the Office of Cancer Clinical Proteomics Research at the **U.S. National Cancer Institute (NCI)** in Rockville, Maryland. "By integrating proteomics with genomics—proteogenomics, a multi-'omics approach—the amount of new biological knowledge that can be derived will be greater than the sum of each individual 'omics part."

### Science of scale

The value of sharing big datasets from proteomics arises from the results they can potentially deliver, such as improvements in health care. For example, says Rodriguez, "Pharma could benefit by better understanding disease and, therefore, developing more effective drugs."

Likewise, proteomics tools can be combined with other tools, such as gene-editing technologies like CRISPR. "The ability to edit a biological system and look at the phenotype is really quite amazing," says Langridge. Tweaking the system with gene-editing tools and then analyzing the results will help scientists unravel the functions of specific proteins.

Some of today's biggest opportunities for sharing come from databases developed for that very purpose. One is EMBL-EBI's PRoteomics IDentifications (PRIDE) database. It includes proteomics data from more than 50 countries and over 8,400 datasets, representing nearly 80,000 assays for acquiring proteomics data—all adding up to about 400 terabytes.

The **Swiss Institute of Bioinformatics** in Lausanne developed neXtProt, which is also a protein knowledge-base. It includes entries on more than 20,000 proteins and nearly 200,000 posttranslational modifications.

"The most famous protein knowledge-base is UniProt, which is focused on more than just human [proteins]," says Vizcaino.

Databases like this can allow new kinds of science. "You can try to come up with ways to combine datasets produced by

different labs, or look for more innovative ways to analyze the data," Vizcaino says. "Usually, the analysis of proteomics data is answering one set of questions, but there could be other ways that the data can be analyzed." So, if someone comes up with a new way to explore existing data, the results could unveil new biological knowledge.

There is far more data about proteins to be determined. As Huhmer points out, "There are about 15,000 known families of proteins." Proteins within a family all have some structural similarity. According to Huhmer, scientists have investigated the structures of these families and measured about 4,500 of them directly with technology such as X-ray crystallography; they have determined another 4,500 structures through computer modeling (with high confidence in only about 1,000 of them); and they have no idea about the structure of the remaining 6,000 or so families.

Advances in technology keep giving scientists more proteomics data to handle. For example, Huhmer mentions that multiplexing label-free approaches to MS could be used to generate 1 million data points a day. In addition, combining MS with structural techniques—such as cryogenic electron microscopy, which can determine a protein's three-dimensional shape—could be used to "analyze some structures that are uncharacterized today," he says. "So, the evolution of technology is revealing more information about protein structure and driving more research in that space."

Even better, once a method reveals the structure of one member of a protein family, a computational approach can be used to unravel other protein structures in that family. "Then, the results grow exponentially," Huhmer explains. In fact, computation plays a wide role in advancing proteomics data and how scientists share it.

The right combinations of technology and research groups make it even easier for scientists to share proteomics data and to collaborate on projects. For example, the **Technical University of Munich (TUM)**, **JPT Peptide Technologies (JPT)** in Berlin, **SAP** in Walldorf, Germany, and Thermo Fisher Scientific created a consortium to help scientists translate proteomics data into advances in basic and medical research. The research data generated by this consortium will be freely available in an online database called ProteomeTools.

### Computing the connections

The example of the ProteomeTools consortium makes it clear that scientists and organizations need to create new ways to work together and share large proteomics datasets. Of course, with so much data being generated and so many possible connections between experiments and results, scientists are more focused than ever on developing new computational tools.

"I think artificial intelligence, machine learning, and deep learning are exciting areas in technology that encourage scientists to share big data," Rodriguez notes. "The reason is that [these technologies] require lots of data and, thus, nudge the research community to share big data in order to ensure their continued development."

Although Rodriguez appreciates the potential of these tools for making the connections that can generate new hypotheses to investigate, he adds, "We must never forget that it's not only about the technology. It's how people interpret, scrutinize, challenge, and agree or disagree with the analysis."

As Rodriguez explains, "These computational systems beckon the need for greater collaboration and for open data science that creates value in new ways." As examples of such ongoing and extensive collaborations, he points out three: the NCI's **Clinical Proteomic Tumor Analysis Consortium (CPTAC)** **cont. >**

## Featured participants

**Applied Proteogenomics Organizational Learning and Outcomes (APOLLO) network**  
proteomics.cancer.gov/programs/apollo-network

**Bruker Daltonics**  
www.bruker.com

**Clinical Proteomic Tumor Analysis Consortium (CPTAC) program**  
proteomics.cancer.gov/programs/cptac

**European Bioinformatics Institute**  
www.ebi.ac.uk

**International Cancer Proteogenome Consortium (ICPC)**  
proteomics.cancer.gov/programs/international-cancer-proteogenome-consortium

**JPT Peptide Technologies**  
www.jpt.com

**SAP**  
www.sap.com

**Swiss Institute of Bioinformatics**  
www.sib.swiss

**Technical University of Munich**  
www.tum.de/nc/en/homepage

**Thermo Fisher Scientific**  
www.thermofisher.com

**University of Wisconsin-Madison**  
www.wisc.edu

**U.S. National Cancer Institute**  
www.cancer.gov

**Walter and Eliza Hall Institute of Medical Research**  
www.wehi.edu.au

**Waters**  
www.waters.com

## Additional resources

**Human Proteome Map**  
www.humanproteomemap.org

**neXtProt**  
www.nextprot.org

**ProteomeTools**  
www.proteometools.org

**Proteomics IDentifications (PRIDE)**  
www.ebi.ac.uk/pride

Whether data is easily shared or not depends on its format. So, Bruker developed its trapped ion mobility spectrometry time-of-flight mass spectroscopy (timsTOF Pro MS/MS) platform to create a format that is available to anyone. "This instrument will generate lots of data, and we need to make it easy to work with," Kruppa says, adding that without this kind of data compatibility, even the most advanced computational tools will hit roadblocks when attempting to compare datasets.

## Seeing what is shared

At this point, one thing is clear: Proteomics scientists do not lack data. Instead, most of these scientists would probably agree with Coon, who says, "We are drowning in data."

Coon notes that the best results require collecting all the raw MS data from an experiment and processing it as a batch. "You want

the first and last samples to be collected and analyzed in the same way," he says.

Getting that done, especially the analytical part, often requires scientists to make their own tools. For instance, Coon hired a data scientist for two years to build a visualization tool. With datasets that combine proteomics, lipidomics, and metabolomics data, this research team needed a way to analyze and organize projects. So, Coon and his colleagues integrated their data viewer in a website.

"We first did this with a yeast project to let people use the data," Coon explains, "and now we create a site like that for every project." So, instead of giving visitors an 8,000-column Excel spreadsheet, Coon's viewer lets other scientists easily compare different samples. "They can make queries on the data very fast," he adds.

Although Coon says he hasn't seen many similar examples of his method, he found that it helps his team and other scientists extract useful biological information from a dataset, because they can interrogate it and compare samples and data points very quickly.

"Most labs that generate so much data have to figure out how to get from raw MS files to something useful, and they probably have their own tools," Coon notes. "There hasn't been a lot of industry leadership that everyone can use." He adds, "People don't value software as much as hardware."

To move forward in sharing large proteomics datasets, however, both hardware and software must continue to improve. Plus, scientists must maintain data quality. While "size is usually what leaps out at the mention of big data," Rodriguez says, "the content and quality of information that gets pulled out of big data is what I consider to be 'big' in terms of knowledge opportunities."

Older datasets can also help scientists move ahead in developing tools. For example, a new analytical tool can be tested on existing datasets and adjusted if needed. "Lots of scientists are working on new tools for analytical techniques involving artificial intelligence," Kruppa points out, "and these tools can be validated on old datasets as long as they can be easily shared."

program; the **Applied Proteogenomics Organizational Learning and Outcomes (APOLLO)** network, which is a collaboration between NCI, the Department of Defense, and the Department of Veterans Affairs; and the **International Cancer Proteogenome Consortium (ICPC)**, which "encourages data to be made available to the public through its 'Data-Sharing Pledge,'" notes Rodriguez.

To further the power of such collaborations, analytical platforms must include techniques that simplify data collection and sharing. Along these lines, Waters developed its SONAR, a data-independent method of acquiring tandem MS data. "The whole idea is that we can acquire proteomics data in a consistent fashion," Langridge says. For every sample tested, this system acquires a quantitative measure of the peptides and proteins. As Langridge explains, "Instead of just identifying the protein, it gathers its abundance across different runs." Moreover, the user makes no decisions before the run starts about the data to collect, because SONAR collects it all. "The challenge with targeted assays," Langridge says, "is that you're making decisions upfront about what you're going to focus on, but you often don't know if you'll have off-target effects or another biochemical pathway involved."

Gathering so much data and storing it in ways that scientists can share it and revisit it in the future should enhance the ongoing value of a dataset. "A lot of studies get published and never looked at again," Kruppa notes, "and that data can't be validated if it can't be shared easily." So, creating data-sharing tools allows both new and old results to be validated. "Plus, [these tools let you] analyze data from other scientists and do a more effective metacomparison of your study to others," he explains.

Older datasets can also help scientists move ahead in developing tools. For example, a new analytical tool can be tested on existing datasets and adjusted if needed. "Lots of scientists are working on new tools for analytical techniques involving artificial intelligence," Kruppa points out, "and these tools can be validated on old datasets as long as they can be easily shared."

Mike May is a freelance writer and editor in Texas. You may reach him at [mike@techtyper.com](mailto:mike@techtyper.com).

1254





### Laser Scanner

The Sapphire Biomolecular Imager is a next-generation laser scanner featuring near infrared (NIR) and visible wavelengths (red, green, blue) fluorescent, true chemiluminescent, and phosphor imaging down to

10- $\mu$ m resolution. This flexibility and performance is achieved by the unique combination of four solid-state laser diodes and three different detection methods: photomultiplier tubes, avalanche photodiodes, and a cooled CCD camera. Coupled with easy-to-use software, the Sapphire allows for rapid image capture, analysis, and quantification across a wide range of proteomics techniques (quantitative NIR Western blots, visible fluorescent imaging, chemiluminescent Western blots, and phosphor imaging), all within a single compact footprint.

#### Azure Biosystems

For info: 925-307-7127

[www.azurebiosystems.com](http://www.azurebiosystems.com)

### Cloud-Connected Pipette and Electronic Lab Notebook

Gilson Connect is a cloud-connected platform that powers a product line of Bluetooth-enabled, smart liquid-handling devices designed to help scientists achieve verifiable science. It includes TRACKMAN Connected, an all-in-one kit that includes a tablet with the PipettePilot tracker application, and PIPETMAN M Connected, a Bluetooth-enabled, smart electronic pipette. Gilson Connect devices can record and track pipette performance in real-time and transmit data to sciNote, a free, open-source electronic lab notebook. Researchers can check their pipetting data to detect any errors resulting from improper use of the pipette, thereby improving traceability and experiment reproducibility. Environmental sensors connected wirelessly to this tablet also enable researchers to monitor environmental conditions that can affect pipetting accuracy. By partnering with sciNote, Gilson Connect allows researchers to store data and consolidate records in a secure location that is easily accessible and shareable, effectively eliminating lost data, a significant contributor to the irreproducibility crisis.

#### Gilson

For info: 800-445-7661

[www.gilson.com](http://www.gilson.com)

### Multiplex ELISA Array

Quantibody is an array-based, multiplex ELISA system for simultaneous quantitative measurement of multiple cytokines, growth factors, proteases, soluble receptors, and other proteins in a wide variety of sample types. It combines a high specificity and sensitivity of ELISA with the high throughput of the glass chip-based array. With this platform, only 50  $\mu$ L of sample is needed for quantification of up to 40 proteins in quadruplicate (0.5 mL for the 400-target Quantibody array), making this array 80 times more efficient than traditional ELISA.

#### RayBiotech

For info: 888-494-8555

[www.raybiotech.com/quantibody-multiplex-elisa-array](http://www.raybiotech.com/quantibody-multiplex-elisa-array)

### Orthogonal Platform

Postnova Analytics has integrated its AF2000 MultiFlow Field-Flow Fractionation (FFF) platform with its SC2000 modular multidetector size-exclusion chromatography (SEC) system, creating an advanced orthogonal design that allows both FFF and SEC separations to be carried out on the same sample within one instrument. The result is enabling labs worldwide to generate more useful information from even complex proteins, antibodies, polypeptides, and polysaccharides. As both separation techniques in the Postnova 2000 Series FFF/SEC system can use the same optimized detector set—ultraviolet, refractive index, and multiangle light scattering—the system's enhanced applications potential is achieved with minimal redundancy, using an automatic switching value and software upgrade kit. The system can be run with different eluents, pH, and wide-ranging temperature conditions, making it ideal for the characterization of natural or synthetic macromolecules.

#### Postnova Analytics

For info: 801-521-2004

[www.postnova.com](http://www.postnova.com)

### Modular Membrane Chromatography

Sartorius Stedim Biotech has expanded its range of single-use membrane chromatography solutions with Sartobind cassettes, which enable large-scale, bind-and-elute membrane chromatography for large proteins such as viruses and virus-like particles, protein conjugates, and blood factors. Sartobind cassettes offer the same flow path, bed heights (4 mm and 8 mm), and void-volume ratios as Sartobind capsules, and are compatible with Q, S, STIC PA (primary amine), and phenyl ligands. Multiple cassettes, each with 0.8-L or 1.6-L membrane volume, can be set up in three different stainless-steel holders, resulting in maximum membrane volumes of 20 L, 50 L, or 100 L. Pressure-flow performance and the shape of breakthrough curves are identical to those found in the smaller capsule sizes, independent of the number of cassettes used. Setup takes only minutes, even at manufacturing scale, whether for capture applications or flow-through removal of process contaminants.

#### Sartorius Stedim Biotech

For info: +49-(0)-551-308-0

[www.sartorius.com](http://www.sartorius.com)

### Lab Automation Scheduling Software

Cellario system control software allows integration with a wide variety of components and liquid handlers, making it easy to add or subtract them from their automated platforms. This unique lab scheduling software architecture enables HighRes Biosolutions to tightly integrate Cellario with your laboratory information management system (LIMS) or other internal software. If needed, HighRes can even create custom laboratory scheduling software for your more complex processes and assays (or lab data output capturing requirements). Total system power loss can be overcome in just a few clicks. Our instrument device driver software is tailor-made for each device we integrate, and can detect and report most errors back to Cellario—which can then be configured to automatically react to them. Cellario protocols can be modified on the fly, even in the middle of a run, using the Active Plate Editor feature.

#### HighRes Biosolutions

For info: 781-932-1912

[highresbio.com/software/cellario.php](http://highresbio.com/software/cellario.php)

Electronically submit your new product description or product literature information! Go to [www.sciencemag.org/about/new-products-section](http://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.

# HOW FAR WILL YOUR ESSAY TAKE YOU?

Apply for the *Science* & SciLifeLab Prize for Young Scientists — an annual prize awarded to early-career scientists. The prize is presented in four categories: Cell and Molecular Biology, Genomics and Proteomics, Ecology and Environment, and Translational Medicine.

The winners will have their essays published by *Science*, win up to USD 30,000 and be invited to a week in Sweden to attend the award ceremony. Get ready for a life-changing moment in your scientific career.

**SCIENCEPRIZE.SCILIFELAB.SE**



*Thut och Alice  
Wallenbergs  
Stiftelse*

**Science**  
MAAS

SciLifeLab



# Consistency Is

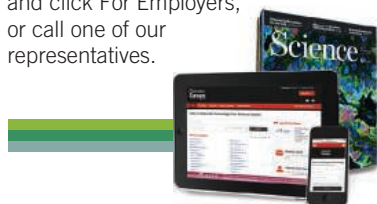
**Get Results The First Time  
And Every Time  
With R&D Systems® Antibodies**

Learn more | [rndsystems.com/antibodies](https://rndsystems.com/antibodies)

# Science Careers

## SCIENCE CAREERS ADVERTISING

For full advertising details, go to [ScienceCareers.org](http://ScienceCareers.org) and click For Employers, or call one of our representatives.



### AMERICAS

+1 202 326-6577  
+1 202 326-6578  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

### EUROPE, INDIA, AUSTRALIA, NEW ZEALAND, REST OF WORLD

+44 (0) 1223 326527  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

### CHINA, KOREA, SINGAPORE, TAIWAN, THAILAND

+86 131 4114 0012  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

### JAPAN

+81 3-6459-4174  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

## CUSTOMER SERVICE

### AMERICAS

+1 202 326-6577  
**REST OF WORLD**  
+44 (0) 1223 326528

[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

**ScienceCareers**

FROM THE JOURNAL SCIENCE AAAS

[ScienceCareers.org](http://ScienceCareers.org)

Max Planck Institute of Immunobiology and Epigenetics

Max-Planck-Institut für Immunbiologie und Epigenetik



MAX-PLANCK-GESellschaft

The Max Planck Institute of Immunobiology and Epigenetics in Freiburg, Germany invites applications from outstanding scientists for the position of the

## Director

A Max Planck Director pursues a bold long-term research agenda in her/his own Department, and leads the Institute jointly with fellow Directors. Applicants should have achieved internationally recognized distinction in their scientific achievements. Ideally, their work complements or extends the activities of the current departments and independent research groups (see <http://www.ie-freiburg.mpg.de>).

Chief criteria for a directorship are scientific and scholarly excellence and an innovative and unique research programme. We are looking for the strongest and most creative scientists worldwide from all career stages, and we offer an excellent international research environment, outstanding infrastructure, and long-term scientific funding.

The Max Planck Society strives for gender equity and welcomes applications from all backgrounds. We actively encourage applications from persons with disabilities.

Applications in PDF format due by **July 31<sup>st</sup>, 2018** must consist of a letter of intent, a curriculum vitae, a publication list, reprints of three to five selected publications, a concept statement outlining the structure and long-term, visionary research agenda of the proposed Department, and names of three referees.

Preliminary inquiries and applications should be sent to the email address below, and will be held in strictest confidence.

Thomas Boehm, M.D., Managing Director, Max Planck Institute of Immunobiology and Epigenetics, Stuebeweg 51, 79108 Freiburg, Germany, email: [managingdirector@ie-freiburg.mpg.de](mailto:managingdirector@ie-freiburg.mpg.de)

## POSITIONS OPEN

### University of South Carolina Tenure-Track Assistant Professor

The Department of Pathology, Microbiology, and Immunology at the University of South Carolina's School of Medicine invites applications for a tenure-track ASSISTANT PROFESSOR position. The successful candidate is expected to develop a strong extramurally funded research program complementing current faculty research interests (<http://pmi.med.sc.edu/>), and participate in teaching. The department is currently ranked in the top 15 among Pathology departments in the nation in NIH funding, and hosts several NIH-funded Research Centers including the Center for Complementary and Alternative Medicine as well as the Center of Biomedical Research Excellence on Dietary Supplements and Inflammation. The department and Centers provide excellent mentoring and research opportunities for junior faculty. Candidates must have a Ph.D. or equivalent, and at least 3 years of postdoctoral research experience. Competitive salary and startup funds are available. Please submit curriculum vitae and a statement of research and teaching interests with names of 3 references to Dr. Mitzi Nagarkatti, Chair, Department of Pathology, Microbiology, and Immunology, University of South Carolina School of Medicine, Columbia, SC 29208 or apply to website: <http://uscjobs.sc.edu/postings/19584>. The search will start immediately and will continue until the position is filled. *USC Columbia is an EOAA Employer and encourages applications from women and minorities and is responsive to the needs of dual career couples.*

## Post Your Jobs

- 1,877,103 unique job seekers
- 250,657 job applications in 2016

**ScienceCareers**



**Download Free Career  
Advice Booklets!**

[ScienceCareers.org/booklets](http://ScienceCareers.org/booklets)



**ScienceCareers**

FROM THE JOURNAL SCIENCE AAAS



# Tianjin Polytechnic University

## The Third "Tianjin scholars"

## International Youth Forum announcement

(2018.7.11-2018.7.14)

### Requirements

Applicants should possess a PhD degree granted by prestigious overseas universities or granted by domestic famous universities and have at least 2 years of overseas research experience. Preferred applicants are winner of "Thousand Youth Talents Plan", "The national youth talent support program", "National Excellent Youth Fund", "Young scholars of the Chang Jiang River", or excellent Young Scholars with high-level. The young outstanding talents with the intention of our positions or replying on our national Thousand Talents program, Tianjin Thousand Talents program, Tianjin Distinguished Professor, "Heavenly Youths" support program and other excellent young talents with great development potential.

### Research Fields

Textiles, Materials, Chemical Engineering, Environmental Sciences, Mechanics, Computer Science, Electronics, Information Science, Biomedical Sciences, Electrical Science, Mathematics, Physics, Optics, Economic Management, Law, Journalism, Philosophy, Foreign Language

Application Method

Please send your CV to email: [rsc@tjpu.edu.cn](mailto:rsc@tjpu.edu.cn)



*The sponsors will cover all costs during this activity and provide high subsidies of international flight for successful applicants.*



東南大學  
SOUTHEAST UNIVERSITY

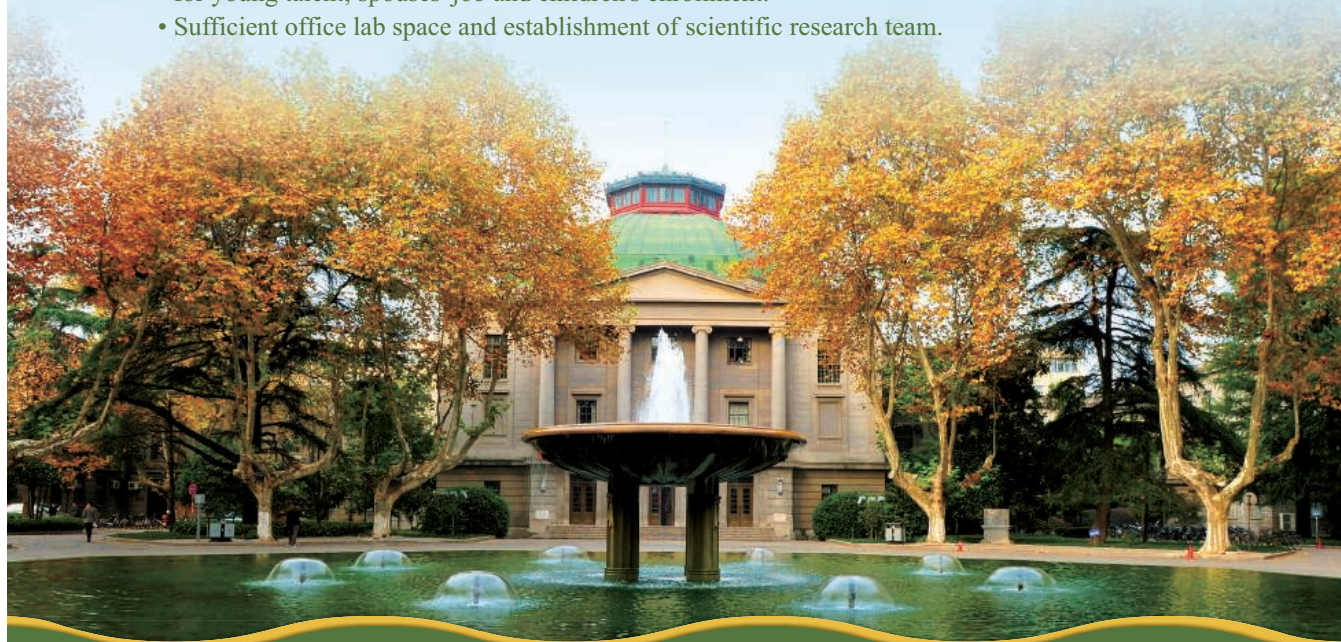
# Recruiting talents at home and abroad in **Southeast University**

*Southeast University (SEU)*, a Class A first-class university building university, was ranked 7th in mainland of China, 23th of the world in USNews2017 engineering college rankings of the world. Southeast University warmly welcomes the outstanding talents at home and abroad to apply for:

- 1. Chang Jiang Scholars Program for Distinguished Professor**
- 2. 1000 Plan Program for Young Talent in 2018**

## **Treatment:**

- Senior professional title: level 2 for Chang Jiang Scholars, senior title for young talents;
- Competitive annual salary: RMB 700,000 for Chang Jiang Scholars, RMB 500,000 for young talents;
- Generous scientific and research fund: RMB 2 to 3 million for Chang Jiang Scholars, RMB 2 to 5 million for young talents;
- Monetary subsidy for house purchasing: RMB 4 million for Chang Jiang Scholars, RMB 3 million for young talent; spouses' job and children's enrollment.
- Sufficient office lab space and establishment of scientific research team.



**For more information, please contact:** Southeast University Talent Service Office

**Contact:** Mrs. Liu, Mr. Yin, Mrs. Liao

**Tel:** + 86-25-83793301, + 86-25-52090253, + 86-25-52090251

**E-mail:** rcb@pub.seu.edu.cn

**Fax:** + 86-25-83793001

**Website:** <http://rsc.seu.edu.cn/2016/0613/c3575a161483/page.htm>





# Northeast Forestry University

## Invites Excellent Talents at Home and Abroad to Apply for "Recruitment Program of Global Young Experts"

Established in 1952 and located in Harbin – the beautiful "Ice City", Northeast Forestry University is in national "211 Project" directly under the Ministry of Education of the People's Republic of China and key construction projects of "Advantage Discipline Innovation Platform". It is a multidisciplinary university integrating agriculture, science, industry, economics, management, culture, law, medicine and art with forestry science as its advantage and forestry engineering as its feature.

### I. Recruitment of high-level talents

#### Relevant support policies for talents

NEFU vigorously implements the "5211" talent introduction plan, and provides a guarantee of talent for realizing the grand goal of building NEFU into a world-class forestry university and comprehensively promoting the construction of the "Double Tops" university.

And the level and treatment of imported talents (house purchase subsidy and personal emolument are both pre-tax) are as follows:

#### 1. Leading Talents or Teams

Academicians of the Chinese Academy of Sciences, academicians of the Chinese Academy of Engineering, and foreign academicians of famous overseas academic institutions; Or someone less than 55 years old who has obtained important scientific research results recognized by domestic and foreign counterparts, and has quite strong competitiveness and wide academic influence in the industry.

**Remuneration:** it depends on the talents' condition

#### 2. Distinguished Young Scholars

Scholars with a PhD who have published high-level academic articles in the top academic journals in the field; Outstanding young scholars at home and abroad who have the potential to be awardees of talent programs such as Recruitment Program of Global Experts, National Science Fund for Distinguished Young Scholars, Changjiang Scholars Program, Young overseas high-level talents introduction plan, and National special support program for high-level talents, etc.; Scholars with a title at or above the level of associate professor from well-known overseas universities, or researchers with the same title as those from overseas well-known research academy (institute); The age generally is under 40 years old, especially distinguished ones can be relaxed to 45 years old.

**Remuneration:** Appointment as professor; Annual salary system, starting from RMB 300, 000 a year; Subsidies for house purchase of RMB 300, 000 to 500, 000; Start-up funds for scientific research in natural science and acquisition expenses of equipment range from RMB 2, 000, 000 to RMB 3, 000, 000; Start-up funds for scientific research in humanities and social science range from RMB 300, 000 to RMB 600, 000.

#### 3. Excellent Young Scholars

Scholars with a PhD who have published high-level academic articles in the influential academic journals in the field; Outstanding young scholars at home and abroad who have the potential to be awardees of projects such as Provincial Science Fund for Distinguished Young Scholars, Longjiang Scholars and Provincial Outstanding Young and Middle-aged experts, etc.; Generally under 35 years old, especially outstanding ones can be relaxed to

40 years old.

**Remuneration:** Appointment as professor or associate professor; Implementation of the national wage and school allowance standard; House purchase subsidies are from RMB 150, 000 to RMB 300, 000; Start-up funds for scientific research in natural science and acquisition expenses of equipment range from RMB 300, 000 to RMB 1, 000, 000; Start-up funds for scientific research in humanities and social science range from RMB 100, 000 to RMB 300, 000.

#### 4. Young Backbone

A doctorate holder who is from a well-known university at home or abroad or who has been engaged in teaching or scientific research in a well-known academic institution at home or abroad for more than 3 years. Young backbone who has published high-level academic articles in the outstanding academic journals in the field, who has the ability to obtain the National Natural Science Foundation or the National Foundation for Philosophy and Social Sciences, and has great development potential in academic and scientific research, etc. The age is generally under 35 years old.

**Remuneration:** Appointment as associate professors or lecturers; Implementation of the national wage and school allowance standards; House purchase subsidies are from RMB 80, 000 to RMB 150, 000; Start-up funds for scientific research in natural science and acquisition expenses of equipment range from RMB 150, 000 to RMB 300, 000; Start-up funds for scientific research in humanities and social science range from RMB 50, 000 to RMB 100, 000.

#### 5. Excellent Young Teachers

The excellent young teacher whose moral quality is high, with outstanding achievements and good academic development potential. The first degree should be a full-time undergraduate degree from a well-known university, and applicants should possess a doctoral candidate education background and a degree at a non-equivalent education level; A non-foreign language teacher should be proficient in a foreign language, and a foreign language teacher should reach level 8 of a professional foreign language. In principle, doctoral education background and degree are required and the age is generally under 35 years old.

**Remuneration:** The position of appointment shall be executed in accordance with the relevant documents; Implementation of the national wage and school allowance standards; House purchase subsidies are RMB 50, 000; Start-up funds for scientific research in natural science range from RMB 50, 000 to RMB 150, 000; Start-up funds for scientific research in humanities and social science range from RMB 30, 000 to RMB 50, 000.

NEFU recruits the flexible talents, and the treatment of flexible talents depends on the circumstances.

### Contact Information

#### Contact Person:

Ni Songyuan (Tel: +86-451-82191327)

Zhang Chunlei (Tel: +86-451-82190165)

Please submit your resume to: [nefumoe@163.com](mailto:nefumoe@163.com)

### II. Recruitment Program of Global Young Experts

#### Recruitment conditions

1. The applicant shall be in the field of natural science or engineering technology, and the age shall not exceed 40 years old;
2. Doctoral degree and more than 3 years of overseas scientific research work experience;
3. Having formal teaching or scientific research positions in overseas well-known universities, scientific research institutions or R & D institutions of well-known enterprises;
4. Top talents among their peers in the scientific research field, and has the potential to become the academic or technical leader in the field.
5. Applicants shall return to China and work on a full-time basis after application. In general, they should not work full time in the country; if they have already worked in the country, they should return to China within one year.

Candidates with outstanding achievements may break the limits of age and length of employment to be introduced.

#### Remuneration and Conditions

The following treatments are provided to those selected under "Recruitment Program of Global Young Experts":

1. Appointment as professors and doctoral tutors;
2. Supporting the establishment of academic teams;
3. The central government will provide the one-off subsidy of RMB 500, 000 per person.
4. The central government will provide financial support of RMB 1, 000, 000 to 3, 000, 000 for scientific research, NEFU will provide the same amounts of scientific research funds.
5. The minimum annual salary provided by NEFU is RMB 400, 000, and house purchase subsidies is RMB 500, 000.
6. NEFU provides a talent apartment;
7. To assist in solving the issues of spouse employment and the enrollment of children.

### Contact information

#### Contact Person:

Zhang Ran (Tel: +86-0451-82192070)

Li Guoliang (Tel: +86-0451-82190195)

Please submit your resume to: [nefubgb@nefu.edu.cn](mailto:nefubgb@nefu.edu.cn)



## Yale Cancer Biology Institute

### Assistant, Associate, or Full Professors

Yale Cancer Biology Institute, part of the Yale Comprehensive Cancer Center (YCC) at Yale University School of Medicine, invites applications from basic and physician scientists for Assistant, Associate, or Full Professor positions in the tenure track. Appointment at Yale is available in a number of departments, and rank will be commensurate with experience and accomplishments. Applicants must have an M.D. and/or Ph.D. or equivalent degree and have demonstrated excellence in research and education. The successful candidates will have experience in the field of cancer biology, with specific target areas being cancer metabolism, bioinformatics, cancer genetics and genomics, epigenetics, chemical biology, tumor microenvironment, and tumor immunology. Responsibilities include establishing a vigorous and independently funded research program in cancer biology, supervising and mentoring students with diverse backgrounds, and contributing to the graduate and medical school educational missions. We seek individuals with strong records of independent creative accomplishments, who will interact productively with colleagues within the Yale Cancer Biology Institute, across the Yale West Campus, the Yale Comprehensive Cancer Center, and the University generally – taking advantage of unique opportunities to translate basic science from the Institute into clinical practice.


The Yale Cancer Biology Institute is one of 6 newly-formed multi-disciplinary research institutes on Yale's burgeoning West Campus, which is self-contained but conveniently linked to Yale's nearby New Haven campuses and the Yale Cancer Center. Additional West Campus assets include the Systems Biology, Chemical Biology, Nanobiology, and Microbial Sciences Institutes – as well as state-of-the-art core facilities in imaging, small molecule discovery, genome analysis, electron microscopy, physical biochemistry, mass spectrometry, and other technologies. Linkage to these Institutes and technologies provides unique opportunities for interdisciplinary research in the Yale Cancer Biology Institute that is simultaneously poised for translation through close involvement with the YCC and Smilow Cancer Hospital at Yale-New Haven.

Please submit a letter describing qualifications, along with a CV, a two-page summary of current and proposed research, and three letters of reference to the following Interfolio website: <http://apply.interfolio.com/50196>

Informal inquiries may be submitted to **Drs. Mark Lemmon and Joseph Schlessinger**, Co-Directors, Yale Cancer Biology Institute, at [cancerbiologyinstitute@yale.edu](mailto:cancerbiologyinstitute@yale.edu). Consideration of applications will begin **August 31st, 2018** and continue until the positions are filled.

*Yale University is an Equal Opportunity/Affirmative Action Employer. We seek candidates who embrace and reflect diversity in the broadest sense. Yale values diversity among its students, staff, and faculty and strongly welcomes applications from women, persons with disabilities, protected veterans, and underrepresented minorities.*

## ScienceCareers

FROM THE JOURNAL SCIENCE 

Follow us for jobs,  
career advice  
and more!



@ScienceCareers



/ScienceCareers



Science Careers

[ScienceCareers.org](http://ScienceCareers.org)

The 2019-2020  
Fulbright U.S. Scholar  
Competition closes  
August 1, 2018.

► For more information on recent  
program innovations, including  
flexible, multi-country opportunities,  
please visit: [www.iie.org/cies](http://www.iie.org/cies)

Fulbright  
SCHOLAR PROGRAM



Charitable Foundation

The EGL Charitable Foundation  
invites you to apply to the

### Gruss Lipper Post-Doctoral Fellowship Program

#### Eligibility

- Israeli citizenship
- Candidates must have completed PhD and/or MD/PhD degrees in the Biomedical Sciences at an accredited Israeli University/Medical School or be in their final year of study
- Candidates must have been awarded a postdoctoral position in the U.S. host research institution.

Details regarding the fellowship are available  
at [www.eglcfc.org](http://www.eglcfc.org)

Application Deadline is October 12, 2018



# Molecular Microbiology and Immunology

School of Medicine

University of Missouri Health

## Department Chair

The University of Missouri (MU) School of Medicine (SOM) invites nominations and applications for the Department Chair of Molecular Microbiology and Immunology (MMI). MMI has a long history of excellence in basic research in microbial pathogenesis and immunology, and has consistently been one of the most productive departments not only within the SOM, but also on the entire MU campus. The research environment at MU is comprehensive, combining Colleges of Medicine, Veterinary Medicine, Agriculture, Engineering, Arts and Sciences, and others on one campus, which fosters the development of interdisciplinary scientific interactions that enhance both research and training opportunities for faculty and students alike. Indeed, MU MMI share faculty with three departments and an interdisciplinary Life Sciences Center, exemplifying our collaborative approach to research. MU has supported the development of state-of-the-art Research Core Facilities, providing instrumentation and high-throughput capabilities in genomics, mass spectrometry, imaging (confocal, light and electron microscopy), transgenic animal generation (particularly mouse, rat and pig), live animal micro-PET and micro-CT imaging, and structural biology (NMR, X-ray crystallography, membership in a synchrotron beamline consortium, and access to Krios-level cryoEM instrumentation). In addition, the campus houses the recently completed Laboratory for Infectious Disease Research (LIDR) which provides modern BSL3/ABSL3 containment research space and animal holding facilities for the investigation of highly infectious organisms and human select agents. Our faculty direct a robust graduate program and are internationally known for their research activities. MU, the flagship campus of the University of Missouri System, is located in Columbia, midway between St. Louis and Kansas City. Columbia is a vibrant, family-oriented community consistently ranked among the top small cities to live in the nation. For additional information, visit <https://medicine.missouri.edu/departments/molecular-microbiology-immunology> and <https://columbiamochamber.com/columbia/>.

**Responsibilities:** The Department Chair is responsible for directing all aspects of the Department. We seek a dynamic individual who will provide vision, leadership and guidance for academic and research programs; recruit and retain a diverse world-class faculty, high-caliber students, postdoctoral fellows and staff; develop resources to strengthen the long-term viability, efficiency, and effectiveness of the department's research and teaching programs; promote collaborative research within the department and across MU, fostering interactions amongst basic science investigators, clinical faculty and clinician-scientists; administer the departmental budget; support/lead external development opportunities; be an advocate for the department; and mentor faculty at different stages of their career. The successful candidate will be expected to maintain a strong extramurally funded research program and to participate in departmental teaching.

**Required Qualifications:** The successful candidate will have the following qualifications or attributes: an earned doctorate in an appropriate discipline; academic accomplishments that merit appointment at the rank of Professor with tenure; a world-class scientist with a strong history of extramural funding and an established record of excellence in teaching and research in an area that complements departmental strengths; demonstrated effective leadership, management and communication with faculty, staff, students, and other administrators; commitment to innovative strategies in diversity and inclusion initiatives as it applies to faculty, staff and students; familiar with national trends in undergraduate, graduate, and medical education; passionate about improving education for all students; a strong collaboration record and willingness to collaborate with others; and a willingness to adopt a transparent, open-minded, respectful and approachable leadership style.

**Preferred Qualifications:** Prior administrative experience or leadership at successive levels of responsibility; an understanding and commitment to the mission of an AAU Research 1 university and the School of Medicine mission; experience with external stakeholders at the state, federal and international levels; commitment to shared governance, collegiality, and continual professional development for faculty and staff; commitment to leading departmental discussions regarding departmental vision and strategic plans with a particular emphasis on translational research strategies; desire to lead developmental activities and cultivate alumni and donor relations.

**Application:** To apply, visit <http://hrs.missouri.edu/find-a-job/academic/> (Job ID 26711) and submit: (1) Cover letter; (2) Curriculum vitae; (3) A two-page summary of research accomplishments and future plans; (4) A narrative of administrative and leadership philosophy, including a statement describing efforts to increase diversity and inclusion; (5) A narrative of teaching philosophy; and (6) Names and contact information of four references (*Candidate will be notified prior to references being contacted*). Review of applications will begin on August 1, 2018 and will continue until the position is filled. For information, contact **Dr. Ronald J. Korthuis**, Chair of Search Committee ([korthuisr@health.missouri.edu](mailto:korthuisr@health.missouri.edu)).

**Benefit Eligibility:** This position is eligible for University benefits. The University offers a comprehensive benefits package, including medical, dental and vision plans, retirement, and educational fee discounts. For additional information on University benefits, please visit the Faculty & Staff Benefits website at <http://www.umsystem.edu/totalrewards/benefits>

**Diversity Commitment:** The University of Missouri is 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.

### An Equal Opportunity/Access/Affirmative Action/Pro Disabled and Veteran Employer

To request ADA accommodations, please call the Disability Inclusion and ADA Compliance Manager at 573-884-7278.



## WORKING LIFE

By Sharon Ramos Goyette

# Hitting the wall

I landed my dream job: a tenure-track position at a primarily undergraduate institution near my hometown where I would develop a new neuroscience major. I entered that position the way one enters a marriage: expecting it to last forever, assuming I would give it everything I had, hoping that—while it would not always be easy—it would be worth it. Soon, though, something seemed amiss. It felt kind of like sexism—but not exactly. Whatever it was, I experienced it from both women and men, from the department chair to the administrative assistant. It was only after many years and a career upheaval that I learned there was a legal term to describe it.

Relatively early on, an administrator asked me to meet with every member of the department to tell them about my teaching schedule. It was odd. None of us knew anyone else's schedules, and mine wasn't particularly unusual, just Monday-Wednesday-Friday. Was this just a normal part of being a junior faculty member? If it was, then why didn't I know the schedule of a colleague who started with me? I had just returned from maternity leave.

One morning, my 2-year-old son had a serious fall and I needed to take him to the emergency room. When I got to campus the next day, a colleague stopped by. He told me that, at lunch, he had heard about what happened and had said to everyone there, "Hey, if I had a dollar for every time I fell on my head, I'd be a millionaire."

Maybe he thought it was funny. I saw it as suggesting that I had foolishly overreacted and shouldn't have let my son's injury take me away from my academic work. I had seen colleagues go out of their way to cover one another's classes when they needed to take a pet to the vet, go to the dentist, or attend their Friday jazz practice. Why was the reaction to me so different?

Later, when the department chair position opened up, I was not invited to apply. I had created and directed one of the most popular science majors on campus and had been unanimously granted tenure. I was also pregnant with my third child at the time.

The sustained effort necessary to succeed in the face of this behavior consumed inordinate energy. I learned to seek out colleagues from other departments and look into administrative guidelines when something didn't feel right. I spent a lot of time trying to anticipate where the next challenge would come from. As I gathered my



***"I felt like my career had been derailed, and I didn't understand exactly why."***

materials to apply for promotion to full professor, I thought, "I cannot do this anymore." I handed in my resignation. Initially my provost did not accept it. However, another administrator said, "It's time. It's time to be home with your children," and then circulated an email to that effect. I felt like my career had been derailed, and I didn't understand exactly why, except that—as this last comment made clear—it was related to the fact that I'm a mother.

When I regained the mental space to reflect on what happened, and after much research, I finally came upon the term "maternal wall bias," now broadened to "family responsibilities discrimination." I learned that some of what I experienced reflected attribution bias, when everyday actions—

taking a child to the doctor, choosing one teaching schedule over another—are interpreted through negative stereotypes of parents, and particularly mothers, as scientists. I learned about leniency bias, when the same work standard is applied more strictly to parents. And I learned that the more children one has, the more severe the bias. It was both validating and maddening to confirm why I had been so marginalized.

I cannot change what happened to me. But I can share what I learned so that both nonparents and parents will be aware that this kind of discrimination is not only unacceptable—it's illegal. After all, this is our responsibility as scientists: When we learn something that may fundamentally impact the future of our fields, we share it. ■

*Sharon Ramos Goyette is currently a visiting lecturer in psychology at Bridgewater State University in Massachusetts. Send your career story to [SciCareerEditor@aaas.org](mailto:SciCareerEditor@aaas.org).*