

A bridge between Indigenous  
communities and geneticists *p. 1304*

Waking up dormant  
cancer cells *pp. 1314 & 1353*

Global killer whale  
decline *p. 1373*

# Science

\$15  
28 SEPTEMBER 2018  
[sciencemag.org](http://sciencemag.org)

AAAS



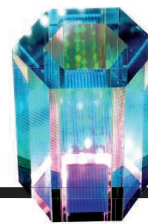
SPECIAL ISSUE

## SHAPING THE BODY

How gene expression  
regulates development

# CONTENTS

28 SEPTEMBER 2018 • VOLUME 361 • ISSUE 6409



1300

Computing with  
captive atoms

## SPECIAL SECTION

# GENES IN DEVELOPMENT

### INTRODUCTION

**1330** Forces behind form

### REVIEWS

**1332** Chromatin plasticity: A versatile landscape that underlies cell fate and identity *T. Yadav et al.*

**1336** Dynamic DNA methylation: In the right place at the right time *C. Luo et al.*

**1341** Developmental enhancers and chromosome topology *E. E. M. Furlong and M. Levine*

**1346** RNA modifications modulate gene expression during development *M. Frye et al.*

### ON THE COVER



In balloon art, the sculptor exerts precise forces at specific sites to make identifiable body parts. Likewise, choreographed molecular

mechanisms, albeit at a complex and tightly controlled level, enable a few seemingly identical starting cells to transform into a whole animal. This special issue highlights the regulatory mechanisms controlling gene expression, depicted in the background here as a heat map. See page 1330. Graphic: FOREAL

SEE ALSO ► PERSPECTIVE P. 1310 ► RESEARCH ARTICLE P. 1354 ► REPORTS PP. 1377 & 1380  
► RESEARCH ARTICLE BY T. GERBER *ET AL.* 10.1126/science.aag0681

## NEWS

### IN BRIEF

**1294** News at a glance

### IN DEPTH

**1298** CLASSIC WOLF-MOOSE STUDY TO BE RESTARTED ON ISLE ROYALE

Wolf airlift will reboot world's longest running predator-prey study, ended by climate change, inbreeding *By C. Mlot*

**1299** NSF ISSUES SEXUAL HARASSMENT POLICY AS NIH PROMISES ACTION

Funding agencies face challenge of sanctioning grantees found guilty without disrupting academic research *By J. Mervis and J. Kaiser*

**1300** ATOMIC ARRAYS POWER QUANTUM COMPUTERS

Neutral atoms rise as dark horse qubit candidate *By S. Chen*

**1302** 'OLD' GENOME EDITORS MIGHT TREAT MITOCHONDRIAL DISEASES

In mice, disrupting mutant DNA appears to relieve its effects *By M. Leslie*

**1303** COOL PAINT JOB FIGHTS SOLAR WARMTH

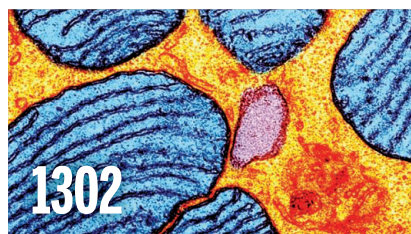
New material could reduce cost of cooling buildings by reflecting light and shedding heat *By R. F. Service*

### FEATURE

**1304** BRIDGING THE GAP

In a bid to ease a history of mistrust, a summer workshop trains Indigenous scientists in genomics *By L. Wade*

► PODCAST



## INSIGHTS

### PERSPECTIVES

**1308** THE MAKING OF A PLANKTON TOXIN

A combined genetic and chemical study reveals the biosynthetic steps involved in making the toxin domoic acid

*By G. Pohnert et al.*

► REPORT P. 1356

**1310** HOX GENES AND BODY SEGMENTATION

An ancient gene cluster controls the formation of repetitive body parts in a sea anemone *By D. Arendt*

► GENES IN DEVELOPMENT SECTION P. 1330;  
REPORT P. 1377

**1311** NO STRAND LEFT BEHIND

Histone chaperones direct how epigenetic information is inherited in dividing cells *By K. Ahmad and S. Henikoff*

► REPORTS PP. 1386 & 1389

**1313** ABOVE AND BELOW THE MAYA FOREST

Advanced remote sensing technology raises questions about settlement and land use *By A. Ford and S. Horn*

► RESEARCH ARTICLE P. 1355

**1314** HOW DORMANT CANCER PERSISTS AND REAWAKENS

Insights reveal possible avenues to prevent metastasis *By J. A. Aguirre-Ghiso*

► RESEARCH ARTICLE P. 1353

**1316** ELECTRO-OPTIC COMBS RISE ABOVE THE NOISE

Electro-optic modulation of light can have a precision equivalent to one optical-field cycle *By V. Torres-Company*

► REPORT P. 1358

### POLICY FORUM

**1317** CANCER PREVENTION: MOLECULAR AND EPIDEMIOLOGIC CONSENSUS

Research in many fields emphasizes the value of prevention *By M. Song et al.*

### BOOKS ET AL.

**1319** STORIES OF THE SANGUINE

A meandering history of blood tackles transfusions, taboos, and trauma *By S. E. Lederer*



# CONTENTS



## 1311, 1386, & 1389

Distributing histones in cell division

28 SEPTEMBER 2018 • VOLUME 361 • ISSUE 6409

### 1320 ME AND YOU AND EVERYONE WE KNOW

From the Big Bang to our early ancestors, a philosopher probes human origins and identity *By M. A. Goldman*

### LETTERS

#### 1322 EDITORIAL EXPRESSION OF CONCERN

*By J. Berg*

#### 1322 LACK OF SCIENCE SUPPORT FAILS BRAZIL

*By K. R. Zamudio et al.*

#### 1323 RELATING ADDICTION AND PSYCHIATRIC DISORDERS

*By J. Vink and A. Schellekens*

#### 1324 SPECIES DEFINITIONS SHAPE POLICY

*By E. G. Ritchie et al.*

#### 1324 TECHNICAL COMMENT ABSTRACTS

## RESEARCH

### IN BRIEF

**1350** From *Science* and other journals

### RESEARCH ARTICLES

#### 1353 CANCER

Neutrophil extracellular traps produced during inflammation awaken dormant cancer cells in mice

*J. Albregues et al.*

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

[dx.doi.org/10.1126/science.aao4227](https://doi.org/10.1126/science.aao4227)

► PERSPECTIVE P. 1314

#### 1354 EPIGENOMICS

Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci

*V. Onuchic et al.*

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

[dx.doi.org/10.1126/science.aar3146](https://doi.org/10.1126/science.aar3146)

► GENES IN DEVELOPMENT SECTION P. 1330



#### 1355 ARCHAEOLOGY

Ancient lowland Maya complexity as revealed by airborne laser scanning of northern Guatemala *M. A. Canuto et al.*

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

[dx.doi.org/10.1126/science.aau0137](https://doi.org/10.1126/science.aau0137)

► PERSPECTIVE P. 1313

### REPORTS

#### 1356 BIOSYNTHESIS

Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom

*J. K. Brunson et al.*

► PERSPECTIVE P. 1308

#### 1358 ULTRAFAST OPTICS

Ultrafast electro-optic light with subcycle control *D. R. Carlson et al.*

► PERSPECTIVE P. 1316

#### 1363 ORGANIC CHEMISTRY

Interrupted carbonyl-olefin metathesis via oxygen atom transfer *J. R. Ludwig et al.*

#### 1369 ORGANIC CHEMISTRY

Arylsulfonylacetamides as bifunctional reagents for alkene aminoarylation *T. M. Monos et al.*

#### 1373 PERSISTENT CHEMICALS

Predicting global killer whale population collapse from PCB pollution *J.-P. Desforges et al.*

► PODCAST

#### 1377 EVO-DEVO

An axial Hox code controls tissue segmentation and body patterning in *Nematostella vectensis* *S. He et al.*

► PERSPECTIVE P. 1310; GENES IN

DEVELOPMENT SECTION P. 1330

#### 1380 SINGLE-CELL GENOMICS

Joint profiling of chromatin accessibility and gene expression in thousands of single cells

*J. Cao et al.*

► GENES IN DEVELOPMENT SECTION P. 1330

#### MOLECULAR BIOLOGY

**1386** A mechanism for preventing asymmetric histone segregation onto replicating DNA strands *C. Yu et al.*

**1389** MCM2 promotes symmetric inheritance of modified histones during DNA replication *N. Petryk et al.*

► PERSPECTIVE P. 1311



### DEPARTMENTS

#### 1293 EDITORIAL

Crisis in Brazil

*By Beatriz Barbuy*

#### 1410 WORKING LIFE

Breaking the silence

*By Kate Funnell*

Science Staff .....	1290
AAAS News & Notes .....	1327
New Products .....	1396
Science Careers .....	1397

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 #R125488122. 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

**Editor, Insights** Lisa D. Chong

**Editors, Research** Valda Vinson, Jake S. Yeston

## Research and Insights

**DEPUTY EDITORS** Julia Fahrenkamp-Uppenbrink(UK), Stella M. Hurlley(UK), Phillip D. Szurromi, Sacha Vignieri **DEPUTY EDITOR, EMERITUS** Barbara R. Jasny **SR. EDITORIAL FELLOW** Andrew M. Sugden(UK) **SR. EDITORS** Gemma Alderton(UK), Caroline Ash(UK), Pamela J. Hines, Paula A. Kiberstis, Marc S. Lavine(Canada), Steve Mao, Ian S. Osborne(UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Jelena Stajic, Peter Stern(UK), 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 Jack, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Amber Esplin, Chris Filiatreau, Cynthia Howe **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Shannon McMahon, Jerry Richardson, Alice Whaley(UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Ope Martins, Nida Masuilis, Dona Mathieu, Ronmel Navas, Hilary Stewart(UK), Alana Warnke, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements(UK), **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldo(UK)

## News

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

**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 **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

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

**WEB TECHNOLOGIES TECHNICAL DIRECTOR** David Levy **PORTFOLIO MANAGER** Trista Smith **PROJECT MANAGER** Dean Robbins **DEVELOPERS** Liana Birke, Ryan Jensen

**DIGITAL MEDIA DIRECTOR OF ANALYTICS** Enrique Gonzales **DIGITAL REPORTING ANALYST** Timothy Frailey **MULTIMEDIA MANAGER** Sarah Crespi **MANAGING WEB PRODUCER** Kara Estelle-Powers **DIGITAL PRODUCER** Jessica Hubbard **VIDEO PRODUCERS** Chris Burns, Meagan Cantwell **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 **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 **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 **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **SR. PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle **SR. TRAFFIC ASSOCIATE** Christine Hall **DIRECTOR OF BUSINESS DEVELOPMENT AND ACADEMIC PUBLISHING RELATIONS, ASIA** Xiaoying Chu: +86-131 6136 3212, xchu@aaas.org **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu + 81532-81-5142 asandhu@aaas.org **EAST COAST/E. CANADA** Laurie Faraday: 508-747-9395, FAX 617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FAX 415-520-6940 **MIDWEST** Jeffrey Dembski: 847-498-4520 x3005, Steven Loerch: 847-498-4520 x3006 **UK EUROPE/ASIA** Roger Goncalves: TEL/FAX +41 43 243 1358 **JAPAN** Kaoru Sasaki (Tokyo): +81 (3) 6459 4174 kasaki@aaas.org

**ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT** Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING ASSOCIATE** Aimee Aponte **MARKETING MANAGER, JOURNALS** Shawana Arnold **MARKETING ASSOCIATES** Mike Romano, Tori Velasquez **SENIOR DESIGNER** Kim Huynh **TRADE SHOW COORDINATOR** Andrew Clamp

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

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

**SUBSCRIPTION SERVICES** For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005 **INSTITUTIONAL SITE LICENSES** 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Central Support** 866-434-2227 www.aaas.org/membercentral

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

**INFORMATION FOR AUTHORS** See [www.sciencemag.org/authors/science-information-authors](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, Pontificia Universidad Católica de Chile  
Ray H. Baughman, U. of Texas at Dallas  
Carlo Beenakker, Leiden U.  
Kamran Behnia, ESPCI  
Yasmine Belkaid, NIAID, NIH  
Philip Benfey, Duke U.  
Gabriele Bergers, VIB  
Bradley Bernstein, Massachusetts General Hospital  
Peer Bork, EMBO  
Chris Bowler, Ecole Normale Supérieure  
Ian Boyd, U. of St. Andrews  
Emily Brodsky, U. of California, Santa Cruz  
Ron Brookmeyer, U. of California, Los Angeles (\$) **\$**  
Christian Büchel, UKE Hamburg  
Dennis Burton, Scripps Research  
Carter Tribley Butts, U. of California, Irvine  
Gyorgy Buzsaki, New York U. School of Medicine  
Blanche Capel, Duke U.  
Mats Carlsson, U. of Oslo  
Ib Chorkendorff, Denmark TU  
James J. Collins, MIT  
Robert Cook-Deegan, Arizona State U.  
Lisa Coussens, Oregon Health & Science U.  
Alan Cowman, Walter & Eliza Hall Inst.  
Roberta Croce, VU Amsterdam  
Jeff L. Dangel, 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 Diaz, 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, U. of Oxford  
Todd Ehlers, U. of Tübingen  
Jennifer Elisseeff, Johns Hopkins U.  
Tim Elston, U. of North Carolina at Chapel Hill  
Nader Engheta, U. of Pennsylvania  
Barry Everitt, U. of Cambridge  
Vanessa Ezenwa, U. of Georgia  
Ernst Fehr, U. of Zürich  
Michael Feuer, The George Washington U.  
Toren Finkel, U. of Pittsburgh Medical Ctr.  
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  
Marta Gonzalez, U. of California, Berkeley  
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  
Louise Harra, U. College London  
Michael Hasselmo, Boston U.  
Jian He, Clemson U.  
Martin Heimann, Max Planck Inst. Jena  
Carl-Philipp Heisenberg, IST Austria  
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 Johnsson, 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 Kochlin, Ecole 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  
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.  
Chris Memmott, U. of Bristol  
Tom Misteli, NCI, NIH  
Yasushi 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. (\$) **\$**  
Samuel Pfaff, Salk Inst. for Biological Studies  
Matthieu Piel, Institut Curie  
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 Shalae, 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, Scripps Research (\$) **\$**  
Timothy D. Wilson, U. of Virginia  
Yu Xie, Princeton U.  
Jan Zaanen, Leiden U.  
Kenneth Zaret, U. of Penn. School of Medicine  
Jonathan Zehr, U. of California, Santa Cruz  
Maria Zuber, MIT



# Crisis in Brazil

Earlier this month, Brazil's National Museum—the oldest, largest, and arguably the most important historical and scientific museum in Latin America—was consumed by fire, stoked by a neglected and degraded infrastructure. It has been a tragic reminder to Brazil and to the rest of the world of how important it is for societies to support the institutions and endeavors that preserve and promote science and culture. This devastating event should serve as a harsh wake-up call for Brazil to bolster, rather than neglect, its scientific enterprise. The general elections next month are an opportunity for Brazil to prioritize science.

The financial crisis in Brazil has been the rationale for a steady decrease in support for science. For example, this year, the Brazilian astronomy community watched Brazil become further disengaged from major international telescope resources and projects. Specifically, Brazil's opportunity to join the European Southern Observatory (ESO) was suspended in March by the Observatory's multinational consortium, thereby halting Brazil's access to the world's largest and most complete observatory located in South America. In 2010, the ESO Council had approved a plan in which Brazil pledged to pay €270 million over 10 years for full member status. Although the plan was approved by the Brazilian Congress in 2015, the government failed to ratify the agreement in the interim. Not surprisingly, Brazilian astronomers have been frustrated by the lack of commitment by Brazil to science and technology, especially given that a strong community of Brazilian astronomers has been fortified over the past 50 years through training programs at home and abroad. Now, this community is seeing its work, and Brazil's investment, strangled.

As of now, Brazil is still involved in developing cutting-edge instruments for telescopes in Chile, including the Extremely Large Telescope, an observatory that will vastly advance astrophysical knowledge of the

universe. But Brazil needs to be part of the ESO, the preeminent intergovernmental science and technology organization that plays a leading role in astronomical research cooperation. Without access to the best instruments, the new generations of astronomers in Brazil will not thrive.

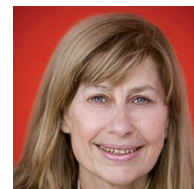
Brazil's elections next month, including for the presidency, are a chance to reestablish commitment to science and rescue the country from further economic decline. The current budget of the Ministry of Science, Technology, Innovation, and Communication (MCTIC) is only 40% of that in 2010 (corrected for inflation),

even after it merged with the Ministry of Communication in 2016. At the same time, the currency has devalued by half. This year, a letter protesting budget cuts in science was signed by 56 Brazilian scientific societies and sent by the Brazilian Academy of Sciences and Brazilian Society for the Advancement of Science to Brazil's president, Michel Temer. The problem is that research and its supporting infrastructure are largely dependent on government support. Funds for research projects necessarily come from the MCTIC and state agencies such as the São Paulo Research Foundation (FAPESP). There is also a problematic misperception in Brazil that science and

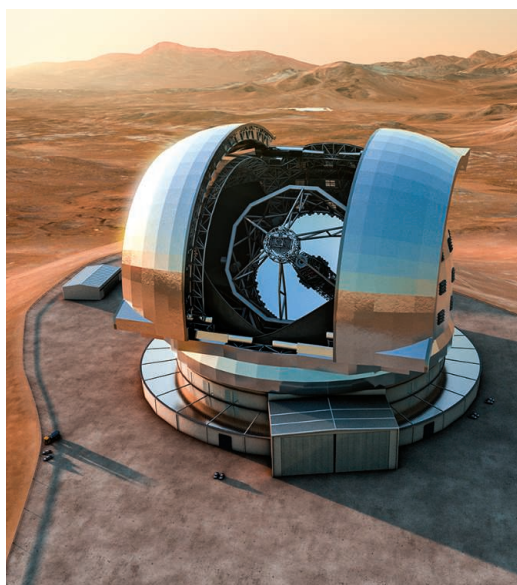
technology have little impact on the economy. By comparison, the League of Research Universities in Europe estimated that in 2016, the output of research-intensive universities bolstered the European economy by generating approximately €100 billion in gross value as well as 1.3 million jobs.

The strongest societies reach their competitive strengths by supporting the scientific enterprise. Even if the transfer-of-knowledge mechanisms are still maturing, their exercise is needed now so that economic prosperity is attainable later on, lest the country be condemned to lag behind other nations.

—Beatriz Barbuy



**Beatriz Barbuy**  
is a professor in the Astronomy Department of the University of São Paulo, Brazil, and coordinator of the pro-ESO Committee of the Brazilian Astronomical Society. [b.barbuy@iag.usp.br](mailto:b.barbuy@iag.usp.br)



The Extremely Large Telescope (part of the ESO) is under construction in Chile.

“This is the first time we’ve shown that we can, in principle, manipulate the fate of an entire species.”

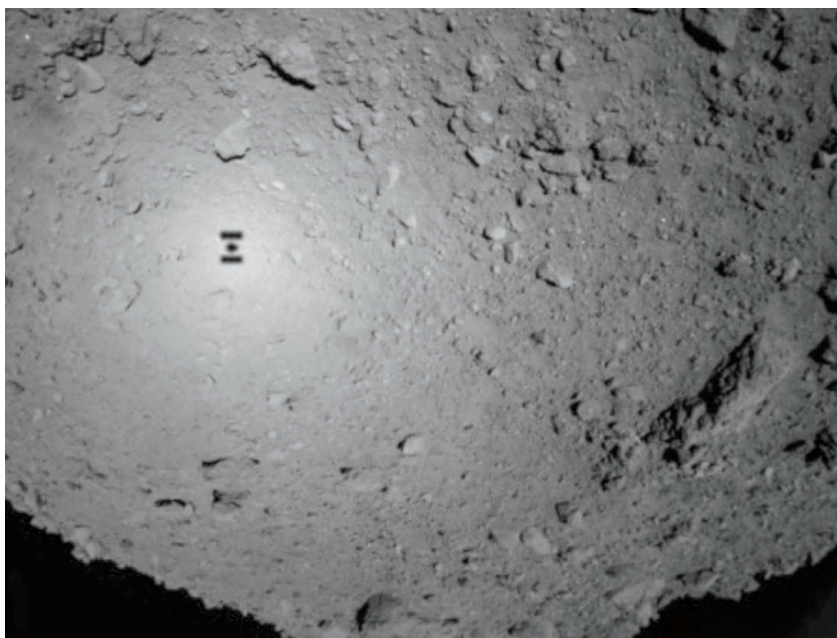
Biologist **Andrea Crisanti**, in *Wired*, about editing the genome of the malaria-carrying mosquito *Anopheles gambiae* so that lab cohorts die.

## IN BRIEF

Edited by **Jeffrey Brainard**

### SPACE SCIENCE

## Asteroid rovers hit the ground hopping



The Hayabusa2 probe casts its shadow on the asteroid Ryugu from about 135 meters away.

In a spacefaring first, two robotic rovers last week began to hop across an asteroid, snapping pictures and gathering data. The rovers, no wider than pie plates, landed on the asteroid Ryugu as part of Japan’s Hayabusa2 mission. On 21 September, the spacecraft released Rover-1A and Rover-1B from a position 55 meters from the 1-kilometer-wide asteroid. Ryugu’s weak gravity then pulled the probes to the surface. (The original Hayabusa probe, which reached the asteroid Itokawa in 2005, released a similar probe prematurely and it floated off into space.) Wheels or crawlers wouldn’t work in Ryugu’s microgravity, so the rovers have internal motors that generate torque; this force propels them up and across the asteroid’s surface, and gravity pulls them back down. Hops can last 15 minutes and cover 15 meters. The cameras on the separate rovers can work together to take stereoscopic pictures. Two additional landers will be released later. The main Hayabusa2 spacecraft will touch down on Ryugu three times between now and May, fire a fingertip-size projectile into the asteroid’s surface, and collect rock fragments before returning to Earth in 2020.

## Gender imbalance affects degrees

**HIGHER EDUCATION** | Female Ph.D. students in the sciences graduate at higher rates in cohorts that contain a higher fraction of women, according to an un-peer-reviewed study published this month by the National Bureau of Economic Research. Researchers examined graduation data for 2541 students in science, technology, engineering, and math who entered 33 doctoral programs at six universities in Ohio between 2005 and 2009. Women with no female peers completed degrees within 6 years at a rate 12 percentage points lower than men in the same cohort. But for each increase of 10% in the proportion of female students in a cohort, their graduation rate increased by 1 percentage point. Programs that enrolled fewer than four in 10 women, on average, accounted for most of the disparity in graduation rates. The findings suggest a lack of gender balance in an incoming cohort may influence its cultural environment, making a Ph.D. program less friendly to women.

## TB vaccine shows promise

**PUBLIC HEALTH** | A tuberculosis (TB) vaccine has thwarted the life-threatening disease in some people latently infected with the TB bacterium. “It’s a breakthrough,” says Barry Bloom, a TB researcher at the Harvard T.H. Chan School of Public Health in Boston. “No vaccine has ever worked in people already infected with TB.” An existing TB vaccine, Bacillus Calmette-Guérin (BCG), is widely used in uninfected infants, but trials have found its efficacy varies widely. The new vaccine, dubbed M72/AS01<sub>g</sub>, was tested in a placebo-controlled study of 3283 latently infected adults—most of whom had received BCG early in life—in three sub-Saharan African countries. It had an efficacy of 54%, as reported online on 25 September in *The New England Journal of Medicine*. Bloom, who was not involved with the study, co-wrote an accompanying editorial stressing the urgent need for a new vaccine as some 1.7 million people still die from TB each year. Curative antibiotics do exist, but many people with TB are not





## ANIMAL BEHAVIOR

### Octopuses on ecstasy get huggy

**H**umans and octopuses are separated by 500 million years of evolution, but both get high on the party drug ecstasy, according to a study reported last week. In humans, ecstasy—or 3,4-Methylenedioxymethamphetamine—causes neurons to pump out massive amounts of serotonin, which produces the warm and friendly feelings of ecstasy’s high. Humans and octopuses share a gene that mediates this neural pathway. To test ecstasy’s effects, researchers placed four California two-spot octopuses (*Octopus bimaculoides*, pictured) in a tank containing ecstasy and then, one by one, in a tank with three chambers: one empty, one containing a plastic action figure, and one with a laboratory-bred octopus under a cage. Octopuses are usually antisocial outside of their mating season, but while under the influence, all four spent the most time with the other octopus and hugged the cage in a way typical of mating season behavior, the team reported in the 20 September issue of *Current Biology*.

diagnosed and others do not complete the 6-month treatment.

### Discredited food expert resigns

**MISCONDUCT** | Brian Wansink, a Cornell University nutrition researcher who won public attention for probing the psychology behind human eating habits, resigned last week after the university determined he violated its academic misconduct policy and six more of his papers were retracted, bringing his retraction count to 13. In a statement, Cornell’s provost said its investigation revealed “misreporting of research data, problematic statistical techniques, failure to properly document and preserve research results, and inappropriate authorship.” Wansink’s headline-friendly findings suggested it’s relatively easy to manipulate people’s eating choices by tweaking features of their environments, such as whether they eat from big or small bowls. In a statement, Wansink admitted to mistaken reporting, poor documentation, and “some statistical mistakes” in his work but said he made no intentional misrepresentations.

### Observatory’s closure explained

**CRIMINAL JUSTICE** | The mysterious weeklong closure of a solar observatory in New Mexico this month was triggered by evidence that a janitor was downloading

and distributing child pornography using its internet connection, *The Washington Post* reported last week. Federal Bureau of Investigation agents had traced the pornography to an IP address at the Sunspot Solar Observatory and seized a laptop computer containing the pornographic images. The contract janitor then told management he was concerned about lax security there and claimed a killer was on the loose in the area who could harm staff members, leading observatory officials to evacuate and close the facility. The janitor has not been arrested or charged, but an investigation continues.

### Italy waters down vaccine law

**PUBLIC HEALTH** | After months of political wrangling, the Italian government has softened requirements in a controversial vaccine law implemented by its predecessor—although not nearly as much as the coalition’s two parties had promised during their election campaigns. The law, adopted last year to fight flagging vaccination rates and major measles outbreaks, requires that parents show proof that their children have received vaccines against 10 diseases before they can be enrolled in school. It was one of the strictest such laws in Europe, and the two ruling parties, which have frequently peddled antivaccine rumors, had vowed to repeal it. The new amendment, approved last week,

doesn’t rescind the obligation to vaccinate but lets parents testify themselves that kids have had the shots. Doctors must still certify the self-reports by March 2019, however. Although the changes may have little practical effect on vaccinations, “they unnecessarily confuse schools and health professionals,” says epidemiologist Pier Luigi Lopalco of the University of Pisa in Italy.

### Agency probes fetal tissue work

**DRUG DEVELOPMENT** | The U.S. Department of Health and Human Services (HHS), citing “serious regulatory, moral, and ethical considerations,” said on 24 September it is comprehensively reviewing “all research involving [human] fetal tissue” to ensure it is compliant with laws and regulations. The department also announced the cancellation of a \$15,900 contract issued in July under which Advanced Bioscience Resources, Inc. of Alameda, California, supplied fetal tissue to the Food and Drug Administration (FDA) for the creation of humanized mice for testing drug safety and effectiveness. FDA wrote to the company that it is “not sufficiently assured” the firm will abide by federal law in procuring and providing the tissue. FDA canceled the contract after 45 anti-abortion groups wrote to HHS Secretary Alex Azar on 11 September, saying they

## CONSERVATION

## Judge cancels hunt of Yellowstone grizzlies

In the latest twist in a long-running battle, a judge has restored legal protections for about 700 grizzly bears living in and around Yellowstone National Park in the western United States, halting a planned bear hunt. The 24 September ruling reversed a 2017 decision by the U.S. Fish and Wildlife Service (FWS) to remove the bears from the federal endangered species list. The agency failed to “rationally consider and apply the best available science” in reaching its decision and “illogically cobbled together” research studies to support it, wrote U.S. District Court Judge Dana Christensen. FWS had argued that the Yellowstone bear population, one of five grizzly populations in the lower 48 states, had grown large enough to withstand limited hunting. But Christensen found that the agency gave up on using state-of-the-art science when it struck a deal with three states—Montana, Wyoming, and Idaho—about how to estimate bear populations.



were “shocked and dismayed” by the sole-source FDA contract. On 17 September, 85 members of Congress wrote a similar letter to FDA Commissioner Scott Gottlieb asking him to cancel the contract.

## Migration brings citations boost

**PUBLISHING** | Physicists who moved between countries for their work received 17% more citations than their stay-at-home counterparts, according to a paper in this week's issue of the *Journal of the Royal Society Interface*. Their publications also covered a wider range of topics and had more co-authors. The study's author, Alex Petersen, a complex systems scientist at the University of California, Merced, looked at 26,000 physicists with papers published in journals of the American Physical Society between 1980 and 2009 and examined how their performance changed over the decade that included a move. To account for the possibility that scientists who are stronger performers get more chances to move,

Petersen's study design paired mobile scientists with nonmobile scientists whose scientific accomplishments were otherwise similar. The study is one of several in recent years finding that cross-border mobility among scientists helps boost their research performance.

## Court exonerates Indian scientist

**CRIMINAL JUSTICE** | India's highest court cleared Sankaralingam Nambi Narayanan, a former top Indian space scientist, of allegations he sold his country's space program secrets to Pakistan. The 14 September decision by the Supreme Court of India condemned the Kerala state police for arresting and torturing him in 1994 on what the court called concocted charges. It awarded Narayanan compensation of 5 million rupees (about \$70,000) and ordered the creation of a committee to take action against the police officials responsible. In 1996, a local court declared him and co-defendants innocent, but litigation continued.

## THREE Qs

## Blowing glass for science

When NASA's Europa Clipper spacecraft sails by Jupiter's icy moon in the late 2020s, it will carry a handful of glass tubes breathed into existence by Mike Souza. The 63-year-old professional glassblower at Princeton University started out helping his father at a glassblowing plant in Evanston, Illinois, and has been creating tricky scientific instruments for universities and companies for nearly 5 decades.

## Q: What are you doing for the Europa mission?

**A:** The glass cylinders I'm making will kind of dangle on the tail end of the spacecraft, like a string of cans on a wedding car. They look like little salt shakers. The cells have to hold helium gas for several years, and since helium is a small molecule, it would work its way through typical glass in 2 to 10 years—and the Europa mission could last much longer than that. Aluminosilicate [the glass Souza is using to make the cells] can hold helium for 20,000 years. ... It's a horrific glass to work with. You need a special touch and a lot of practice.

## Q: You also handled a specialty glass meant to help Princeton researchers track down dark matter particles. What was unique about this glass?

**A:** Normal glass contains trace elements, and if they decay, they could cause all this radioactive background that would disrupt the dark matter signal. So instead, we had to work with an ultrapure glass. It costs \$200,000 for a slab that's about 5 centimeters thick and 35 centimeters across. It had to be put together in a clean room, and it was nerve-wracking because the lead scientist was right there, watching his glassware that cost more than your house.

## Q: Is it strange, knowing the number of people in your career is dwindling?

**A:** It's really a pressure all universities are facing. Chemistry has shrunk vastly in the amount of glasswork required, but glass has expanded into new sciences, like ion traps for quantum computing research and replicas of patient blood vessels for research. I don't know what the long-term future will be, but I think glass as a material will be more and more valuable.

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





The two surviving wolves on Isle Royale, shown last winter, are closely related and have not bred.

pects of the island we haven't researched fully," says Isle Royale National Park Superintendent Phyllis Green. "How do we restart the science?"

In the 1960s the project firmly established that wolves focus on old, young, and infirm prey, rather than slaughtering indiscriminately. The study was among the first to document in a terrestrial system how a food web is controlled from the top down, by connecting wolf predation with more plant growth as shown in tree rings (*Science*, 2 December 1994, p. 1555). More recently, researchers have documented the effects of inbreeding and the short-lived genetic rescue of the pack by an immigrant wolf known as the "old gray guy." He crossed an ice bridge from the mainland in 1997, but his impact and origin were only recognized a decade later, with analyses of DNA from scat from him and his numerous descendants (*Science*, 24 May 2013, p. 919).

For population biologist John Vucetich of Michigan Technological University (MTU) in Houghton, who joined the study in 1991, that epitomizes its greatest value: documenting chance events whose impact can't be recognized until later. "In ecosystems, the most important things are inherently unpredictable," he says. "There is no chance to understand how ecosystems work unless you watch them over time."

The wolves arriving this fall will make up the first of several waves over the next 3 years, with the goal of establishing a genetically diverse, reproducing population of 20 to 30 wolves to control the booming moose. Moose numbers have been trending up for the past decade with the absence of predation, and some 1475 animals now chomp away on vegetation including the island's balsam fir, many of which

## IN DEPTH

### ECOLOGY

# Classic wolf-moose study to be restarted on Isle Royale

Wolf airlift will reboot world's longest running predator-prey study, ended by climate change, inbreeding

By Christine Mlot

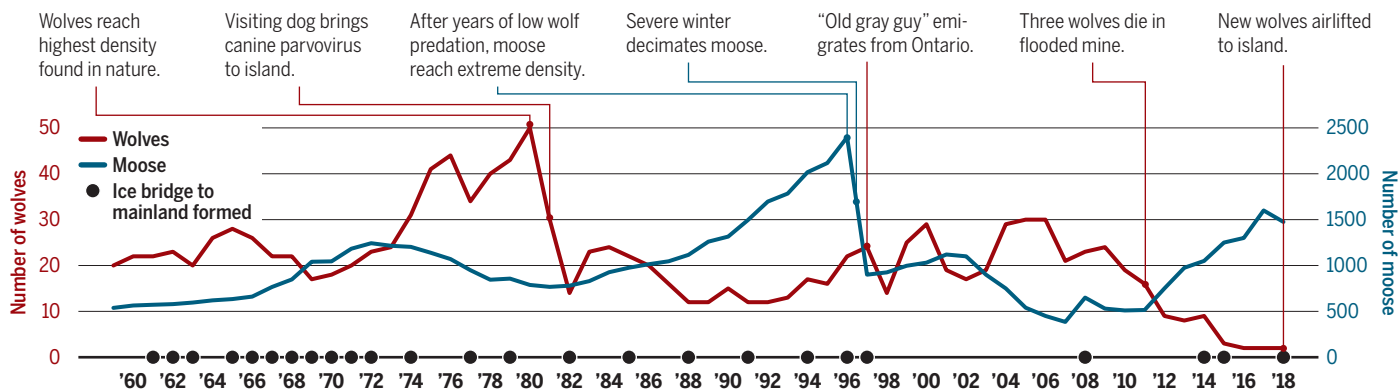
**M**ore than 60 years ago, a couple of wolves wandered across a frozen channel on Lake Superior and settled on moose-rich Isle Royale in Michigan, touching off a dance of predators and prey and a classic study in ecology. Now, a new experiment is starting. If the wind and waves on Lake Superior cooperate, by the end of this month the U.S. National Park Service (NPS) will airlift the first of six wolves from the U.S. mainland to Isle Royale by float plane. The goal is to re-establish predators now that

the original wolf population has dwindled to a vulnerable pair (*Science*, 21 April 2017, p. 226). Planners aim to move a wolf each week over the next month.

The operation is a turning point for the wilderness island and NPS, which has largely taken a hands-off approach to managing natural areas. But inbreeding as well as warmer winters, which mean less lake ice and fewer opportunities for mainland wolves to restock the population, have essentially ended the classic natural experiment. With the reboot, NPS hopes to attract researchers with new questions to explore. "We know there are many as-

## A dance of predators and their prey

The Isle Royale study shows the importance of chance events as well as climate change, as ice bridges that allow immigration by mainland wolves are now rare.



are stunted from overbrowsing. On some ponds, moose have eliminated the once dominant native watershield, a floating aquatic plant.

NPS plans to monitor the populations of both moose and wolves, which they will track with GPS collars, and the impact of the new arrivals on the ecosystem. It is budgeting about \$2 million over 20 years for the project, and it is encouraging scientists to ask other questions, for example about plants' response to predation and the genetic makeup of the new wolves.

This time around, planners hope to avoid inbreeding by ensuring geographic and genetic diversity in the imported wolves, along with an equal ratio of males to females. Two wolves will come from the Upper Peninsula in Michigan and two to four from the northeastern Minnesota reservation of the Grand Portage Band of Lake Superior Chippewa, which was likely the source of the wolves that found their way to Isle Royale in past decades. Other wolves are expected to be brought in from Canada's Ontario province in 2019.

Ontario wolves have experience killing moose, but Michigan wolves may have never seen one, and Minnesota wolves may have limited experience killing them, says MTU wildlife ecologist Rolf Peterson. The new wolves will have to learn how to bring down the 400-kilogram mammals. "However, wolves are wonderful observational learners, and hunger is a strong motivation," Peterson says. The newcomers will also find plenty of familiar beaver, which have been booming as the wolves declined.

The fate of the two surviving Isle Royale wolves—heard howling last month—remains to be seen. Peterson spotted them on Valentine's Day, bedded down with full bellies. But there was little sign of courtship. The 9-year-old male "was looking grizzled," says Peterson, who has studied the wolves since 1971. The 7-year-old female was "solicitous," and at times pawed the male, Peterson says, but "it seemed like she was [providing] elder care." She is the male's daughter and half-sibling.

NPS plans to avoid the pair's territory when it releases the mainland animals onto the 544-square-kilometer island. Because the aging, existing pair is bonded, no one expects the old and new wolves to interbreed, though it can't be ruled out. The animals may fight, but managers will be hands off once the wolves are released. "We intend to let them work it out," Green says. How they do so is just one of the surprises the experiment has in store. ■

*Wisconsin-based Christine Mlot has been reporting on Isle Royale wolves since 1993.*

## SCIENTIFIC COMMUNITY

# NSF issues sexual harassment policy as NIH promises action

## Funding agencies face challenge of sanctioning grantees found guilty without disrupting academic research

By Jeffrey Mervis and Jocelyn Kaiser

Nobody condones sexual harassment by scientists. That much is obvious. But last week it also became clear that two leading U.S. government science agencies—the National Institutes of Health (NIH) in Bethesda, Maryland, and the National Science Foundation (NSF) in Alexandria, Virginia—are still wrestling with how to sanction such behavior without disrupting the overall academic research enterprise.

Under growing pressure from Congress and the scientific community to act, NSF made the first move. On 19 September, it announced new reporting requirements for institutions designed to help the agency track sexual harassment by grantees. The change prompted NIH to explain why it has yet to take a similar step—and elicited calls for the government to take even stronger action.

Under the change, universities must tell NSF if any faculty member with an NSF grant has been found guilty of sexual or other forms of harassment and if they have placed an NSF-funded scientist on administrative leave for a matter relating to harassment. (Institutions sometimes take administrative action against faculty members after they receive an allegation but long before they have concluded an investigation.) The changes finalize a draft policy that NSF posted in March and apply to grants awarded after 20 October.

"The scientific community has not sufficiently protected all of its members," NSF Director France Córdoba said when announcing the new rules, noting the rising number of harassment cases. "This neglect must end," she said, adding that she believed the policy would help NSF "provide targeted, serious consequences for harassers ... without disturbing others' careers and lives. We think this is a big deal." Those consequences could include replacing the principal investigator on a grant or even terminating funding.

Congress has been waiting impatiently for funding agencies to clarify how they will deal with harassers. In February, the

science committee in the U.S. House of Representatives asked the Government Accountability Office (GAO), a congressional watchdog agency, to examine the policies of NSF and four other science agencies. And a few hours after NSF's announcement, its chairman, Representative Lamar Smith (R-TX), sent a 14-page letter to GAO reiterating his concern that the government lacks uniform policies. "No taxpayer dollars should be awarded to a researcher



***"The scientific community has not sufficiently protected all of its members. This neglect must end."***

France Córdoba, National Science Foundation

who engages in harassment toward a colleague or student under their charge," Smith wrote.

Smith's letter asserts that the agencies and universities need to do a better job of reminding scientists that an agency can terminate a grant if a scientist is found guilty of harassment. (Awards are given to institutions, not individuals, although scientists invariably refer to "my" grant.) But



Córdova says termination should be a last resort, pursued only if NSF decides no other faculty member is capable of taking over the project. A university would typically propose such a transfer in hopes of retaining a grant that supports students and other members of the research team.

NIH is under similar pressure. Last month, Senator Patty Murray (WA) and Representative Rosa DeLauro (CT), the senior Democrats on panels overseeing NIH, told the agency it should be “doing more to prevent and address discriminatory practices and harassment by NIH-funded researchers.” NIH told them what it is doing to address their concerns, but last week’s NSF announcement prompted NIH Director Francis Collins to publicly explain why his agency doesn’t have similar reporting policies.

NIH faces “legal constraints” that don’t apply to NSF, and that “prevent NIH from immediate implementation of an identical policy,” Collins said in a statement. He was referring to the Administrative Procedures Act, a law that requires NIH to go through a lengthy rule-making process that includes approval from its parent agency, the Department of Health and Human Services, and from the White House Office of Management and Budget. Although the law also applies to NSF, agency officials have interpreted it to mean NSF can revise its grantmaking policies by using a simpler—and shorter—“notice and comment” approach.

But Collins also promises future steps. “We’re going to ... put a little pressure on the boundaries and we’ll see how far we get.”

The research community is also watching. Earlier this year, for example, a coalition of higher education associations lamented the potential “chilling effect” of informing NSF about cases where a university has imposed administrative leave or taken other “non-punitive” steps before an investigation is complete. NSF could use that information to act “prematurely,” they fear.

Some researchers, however, would like to see the government take an even more aggressive approach. “I recognize it’s complicated for the NIH to work through, but I think the community is looking for more, and rightly so,” says Benjamin Corb, public affairs director for the American Society for Biochemistry and Molecular Biology in Rockville, Maryland. In particular, some scientists and lawmakers want the govern-

ment to prevent scientists found guilty of harassment from ever receiving another federal grant. Currently, however, there is no clear mechanism for imposing such a lifetime debarment.

Agencies can debar a scientist for a specific length of time—often 1 to 3 years—if they are found guilty of research misconduct. But harassment isn’t part of the current federal definition of misconduct, which covers fabrication, falsification, and plagiarism. A recent National Academies of Sciences, Engineering, and Medicine report that found widespread sexual harassment in academia suggested adding sexual harassment to the federal definition. And last fall the American Geophysical Union, based in Washington, D.C., took that step in revising its ethics policy for members.

NSF settled on its new reporting rules, Córdova noted, “because that’s what we could do swiftly and with high impact. But I expect that question [of expanding the misconduct definition] will come up again.”

Neither she nor Collins seems eager to broaden the federal misconduct definition, however. The current version is the product of a hard-fought compromise struck in the late 1990s that addressed misconduct without trying to define all types of unacceptable

behavior—a discussion Córdova took part in as NASA’s chief scientist under then-President Bill Clinton. Prior to that, NSF used its own definition that also covered what it called “serious deviation from accepted practices”—including harassment.

Evoking that earlier controversy, Collins thinks the government would “end up in a very muddy situation” if it reopens the debate over defining misconduct. “This is a separate issue that requires separate oversight and policy,” he says.

The likely forum for any such debate would be the National Science and Technology Council, an interagency body overseen by the White House Office of Science and Technology Policy (OSTP). The Senate is poised to confirm meteorologist Kelvin Droegemeier, former vice president for research at The University of Oklahoma in Norman, as OSTP’s leader. Given the issue’s high profile, it is a good bet that the question of how to handle government-funded scientists who have committed sexual harassment will be sitting in his inbox once he comes on board. ■

## QUANTUM COMPUTING

# Atomic arrays power quantum computers

## Neutral atoms rise as dark horse qubit candidate

By Sophia Chen

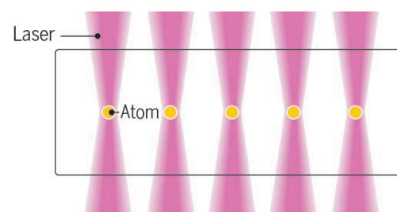
In a small basement laboratory, Harry Levine, a Harvard University graduate student in physics, can assemble a rudimentary computer in a fraction of a second. There isn’t a processor chip in sight; his computer is powered by 51 rubidium atoms that reside in a glass cell the size of a matchbox. To create his computer, he lines up the atoms in single file, using a laser split into 51 beams. More lasers—six beams per atom—slow the atoms until they are nearly motionless. Then, with yet another set of lasers, he coaxes the atoms to interact with each other, and, in principle, perform calculations.

It’s a quantum computer, which manipulates “qubits” that can encode zeroes and

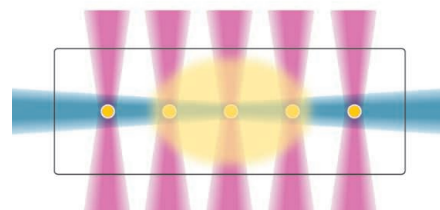
## Entangling atoms

Researchers are using lattices of neutral atoms the bits to power a quantum computer. But the shy, uncharged atoms must be excited in order to interact.

**1** Individual laser beams are used to trap arrays of atoms in vacuum chambers.



**2** Other lasers excite an atom’s outermost electron. The massively enlarged atom can interact with its neighbors, enabling the formation of entangled quantum bits.



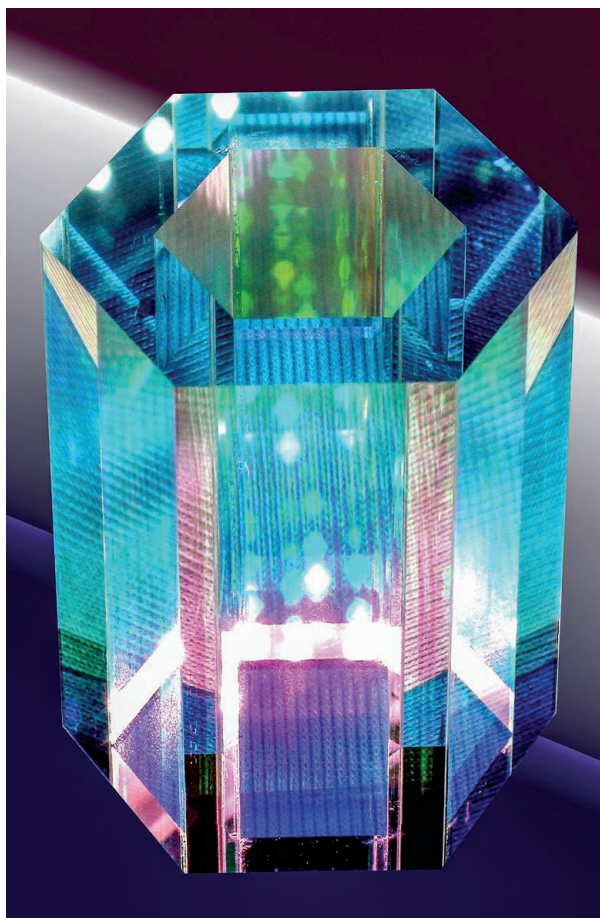
ones simultaneously in what's called a superposition state. If scaled up, it might vastly outperform conventional computers at certain tasks. But in the world of quantum computing, Levine's device is somewhat unusual. In the race to build a practical quantum device, investment has largely gone to qubits that can be built on silicon, such as tiny circuits of superconducting wire and small semiconductor structures known as quantum dots. Now, two recent studies have demonstrated the promise of the qubits Levine works with: neutral atoms. In one study, a group including Levine showed a quantum logic gate made of two neutral atoms could work with far fewer errors than ever before. And in another, researchers built 3D structures of carefully arranged atoms, showing that more qubits can be packed into a small space by taking advantage of the third dimension.

The advances, along with the arrival of venture capital funding, suggest neutral atoms could be on the upswing, says Dana Anderson, CEO of ColdQuanta, a Boulder, Colorado-based company that is developing an atom-based quantum computer. "We've done our homework," Anderson says. "This is really in the engineering arena now."

Because neutral atoms lack electric charge and interact reluctantly with other atoms, they would seem to make poor qubits. But by using specifically timed laser pulses, physicists can excite an atom's outermost electron and move it away from the nucleus, inflating the atom to billions of times its usual size. Once in this so-called Rydberg state, the atom behaves more like an ion, interacting electromagnetically with neighboring atoms and preventing them from becoming Rydberg atoms themselves.

Physicists can exploit that behavior to create entanglement—the quantum state of interdependence needed to perform a computation. If two adjacent atoms are excited into superposition, where both are partially in a Rydberg state and partially in their ground state, a measurement will collapse the atoms to one or the other state. But because only one of the atoms can be in its Rydberg state, the atoms are entangled, with the state of one depending on the state of the other.

Once entangled, neutral atoms offer some inherent advantages. Atoms need no quality control: They are by definition identical. They're much smaller than silicon-



Lasers are used to trap arrays of atoms within glass chambers made by ColdQuanta, a neutral atom quantum computing startup.

based qubits, which means, in theory, more qubits can be packed into a small space. The systems operate at room temperature, whereas superconducting qubits need to be placed inside a bulky freezer. And because neutral atoms don't interact easily, they are more immune to outside noise and can hold onto quantum information for a relatively long time. "Neutral atoms have great potential," says Mark Saffman, a physicist at the University of Wisconsin in Madison. "From a physics perspective, [they could offer] easier scalability and ultimately better performance."

The two new studies bolster these claims. By engineering better quality lasers, Levine and his colleagues, led by physicist Mikhail Lukin at Harvard, were able to accurately program a two-rubidium atom logic gate 97% of the time, they report in a paper published on 20 September in *Physical Review Letters*. That puts the method closer to the performance of superconducting qubits, which already achieve fidelity rates above 99%. In a second study, published in *Nature* on 5 September, Antoine Browaeys of the Charles Fabry Laboratory near Paris and his colleagues demonstrated an unprec-

edented level of control over a 3D array of 72 atoms. To show off their control, they even arranged the atoms into the shape of the Eiffel Tower. Another popular qubit type, ions, are comparably small. But they can't be stacked this densely because they repel each other, acknowledges Crystal Senko, a physicist at the University of Waterloo in Canada who works on ion quantum computers.

Not everyone is convinced. Compared with other qubits, neutral atoms tend not to stay put, says Varun Vaidya, a physicist at Xanadu, a quantum computing company in Toronto, Canada, that builds quantum devices with photon qubits. "The biggest issue is just holding onto the atoms," he says. If an atom falls out of place, Lukin's automated laser system can reassemble the atoms in less than a second, but Vaidya says this may still prohibit the devices from performing longer tasks. "Right now, nobody knows what's going to be the best qubit," Senko says. "The bottom line is, they all have their problems."

Still, ColdQuanta has recently received \$6.75 million in venture funding. Another startup, Atom Computing, based in Berkeley, California, has raised \$5 million. CEO Ben Bloom says the company

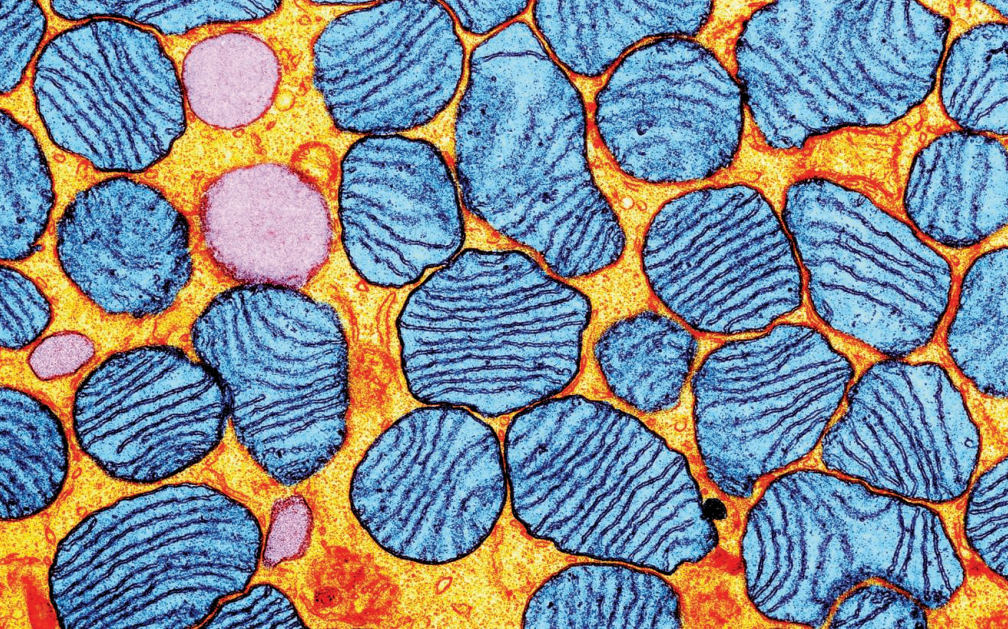
will pursue qubits made of atoms with two valence electrons instead of rubidium's one, such as calcium and strontium. Bloom believes these atoms will allow for longer-lived qubits. Lukin says he's also interested in commercializing his group's technology.

The startups, as well as Saffman's group, are aiming to build fully programmable quantum computers. For now, Lukin wants his group to focus on building quantum simulators, a more limited kind of computer that specializes in solving specific optimization problems by preparing the qubits a certain way and letting them evolve naturally. Levine says his group's device could, for example, help telecommunications engineers figure out where to put radio towers to minimize cost and maximize coverage. "We're going to try to do something useful with these devices," Levine says. "People still don't know yet what quantum systems can do."

In the next year or two, he and his colleagues think neutral atom devices could deliver an answer. ■

*Sophia Chen is a journalist in Tucson, Arizona.*





Cells can have thousands of mitochondria (blue), each with DNA that can harbor disease-causing mutations.

## GENE THERAPY

# ‘Old’ genome editors might treat mitochondrial diseases

In mice, disrupting mutant DNA appears to relieve its effects

By **Mitch Leslie**

**C**RISPR, the genome editor celebrated as a potentially revolutionary medical tool, isn’t omnipotent. Mitochondria, the organelles that supply a cell’s energy, harbor their own DNA (mtDNA), and mutations there can have devastating consequences, including deafness, seizures, and muscle weakness. Genome editing might be a remedy, but mitochondria appear to be off-limits to CRISPR.

This week, two studies in *Nature Medicine* reveal that older genome-editing tools can slash the amount of defective mtDNA in mice bred to have a mitochondrial mutation, counteracting its effects. The proof-of-principle results could open the way for the first treatments for mitochondrial diseases. “These are remarkable findings that make it possible to even consider doing this in humans,” says mitochondrial biologist Martin Picard of the Columbia University Irving Medical Center, who was not involved in the work.

Turning these results into a treatment will be tricky. The genes encoding the genome editors had to be introduced by viruses, and researchers have struggled to make similar gene therapy efforts work. But “these are the right experiments to get ready to go into people,” says molecular geneticist Stephen Ekker of the Mayo Clinic in Rochester, Minnesota, who wasn’t connected to either study. In fact, both groups are considering clinical trials.

Descendants of ancient bacteria, mitochondria sport their own small genomes and a distinct set of proteins. Each cell can contain thousands of these organelles, and mutations in mtDNA cause a range of illnesses. “Together, they are one of the most common causes of genetic disease in humans,” says molecular biologist Michal Minczuk of the University of Cambridge in the United Kingdom, who led one of the research teams.

The controversial “three-parent baby” approach can protect children from inheriting mitochondrial diseases. It involves replacing the defective mitochondria in the mother’s egg with those of a healthy donor. But researchers haven’t discovered any treatments for a person who inherits faulty mtDNA. “It’s a large unmet need,” Ekker says.

Snipping the mutant DNA could help because mitochondria destroy the severed molecules. Moreover, potential treatments might not need to eliminate all the defective mtDNA in the body’s myriad mitochondria. Patients have mtDNA copies with and without the harmful mutation, and the ratio between the two varieties must reach a certain level before symptoms occur, notes mitochondrial biologist Carlos Moraes of the University of Miami Miller School of Medicine in Florida, who led the other research team. “If you can lower this ratio below the threshold, the clinical manifestations might go away.”

CRISPR, however, was not an option. It depends on a RNA strand to guide the DNA-cutting protein to the right spot in the

genome, and most researchers doubt that mitochondria can take up these guide RNAs. So both teams turned back the clock and tested two earlier editing approaches—zinc finger nucleases (ZFNs) and transcription activatorlike effector nucleases (TALENs). Both consist of DNA-cutting proteins designed to home in on DNA without guide RNAs. Although more cumbersome and less versatile than CRISPR, they, too, can slice DNA at a specific location.

Both groups of researchers harnessed similar viruses, considered harmless, to ferry genes for the DNA-editing proteins into the mice. In this strain, some mtDNA copies have a mutation in the gene coding for a type of transfer RNA (tRNA), which helps assemble mitochondrial proteins. The animals have less of this tRNA variety than normal, although they only develop a subtle symptom, a mild heart abnormality.

In their study, Moraes, his colleague Sandra Bacman, and their team injected viruses loaded with TALENs genes into a leg muscle on each animal’s right side. As a control, they shot viruses lacking the TALENs genes into the same muscle on the left side. After 6 months, the amount of mutant mtDNA was more than 50% lower in the muscle that received TALENs, and the ratio of damaged to normal DNA was below the 50:50 threshold that typically produces symptoms.

Minczuk, along with his postdoc Payam Gammage and colleagues, designed equivalent ZFNs and injected viruses carrying them into the tail veins of the mice. Once in the bloodstream, the viruses traveled to the heart, which also harbored the defective mtDNA. Analyzing the animals’ cardiac tissue 65 days later, the team found the fraction of mutant mitochondrial DNA was about 40% lower.

Because the mild heart abnormality is hard to document, the researchers used molecular indicators to gauge success. Both groups determined that levels of the tRNA that is scarce in the mutant mice surged after gene therapy. Minczuk’s team also measured several metabolic molecules that suggested the animals’ mitochondria were working better.

Researchers agree that to apply the strategy in people, they will have to ensure that the genes for TALENs and ZFNs reach the right tissues, in the right amounts. Nonetheless, Moraes says his team is trying to organize a safety trial of their approach in people with a mtDNA mutation, starting as early as next year. Minczuk says his group also hopes to launch clinical trials, but he doesn’t have a timetable. “It’s a hopeful moment for these diseases,” Gammage adds. ■



## MATERIALS SCIENCE

# Cool paint job fights solar warmth

New material could reduce cost of cooling buildings by reflecting light and shedding heat

By Robert F. Service

Homes in Greece and other sunbaked countries are regularly painted white to reflect as much sunlight as possible. Researchers are building on that age-old strategy with a new wave of “passive radiative cooling” materials that shed sunlight and heat. Most are not easy to apply to existing roofs and walls, but a team of U.S. researchers has now created a cooling paint that can coat just about any surface, lowering its temperature by 6°C.

The advance underscores “terrific progress in this field,” says Xiaobo Yin, a materials scientist at the University of Colorado in Boulder whose team has developed a passive radiative cooling plastic film and has formed a startup company, Radi-Cool, to commercialize it. The new materials, Yin says, could drop cooling costs by up to 15% in some climates. “It’s quite a big number,” he says. And with 17% of all residential electricity use in the United States going toward air conditioning, the savings could be substantial.

White paints typically reflect only about 80% of visible light, and they still absorb ultraviolet (UV) and near-infrared (near-IR) rays, which warm buildings. To do better, the new materials start by incorporating materials or structures that reflect nearly all the sun’s incoming rays, including near-IR heat and, in some cases, UV as well. They also contain polymers or other substances that, because of their chemical makeup, radiate away additional heat as mid-IR light, at wavelengths of 8 to 13 micrometers. The atmosphere does not block these wavelengths, effectively allowing the materials to shed excess heat into space without warming the surrounding air.

In 2014, researchers led by Shanhui Fan, an electrical engineer at Stanford University in Palo Alto, California, reported in *Nature* that by alternating layers of silicon dioxide and hafnium dioxide, they could create a highly reflective surface that stayed 5°C cooler than the surrounding air. Last year, Fan and his colleagues used another material—a polymer and silver film combo—

to cool water for use in air conditioning. The team showed that it had the potential to save 21% of air conditioning costs over the summer. Fan’s team has since formed its own startup, SkyCool Systems.

Yin and his colleague Ronggui Yang jumped in last year with their material: a plastic film embedded with tiny glass beads that cooled surfaces by up to 10°C. And in Australia, applied physicists Angus Gentle

paint comes in. Researchers led by Yuan Yang and Nanfang Yu, applied physicists at Columbia University, had been experimenting with making highly reflective materials by adding air voids to plastics. Then one of their students, Jyotirmoy Mandal, stumbled on an easy method for creating a similar texture. Mandal was studying ways of curing a common polymer called polydimethylsiloxane from a liquid to a solid thin film

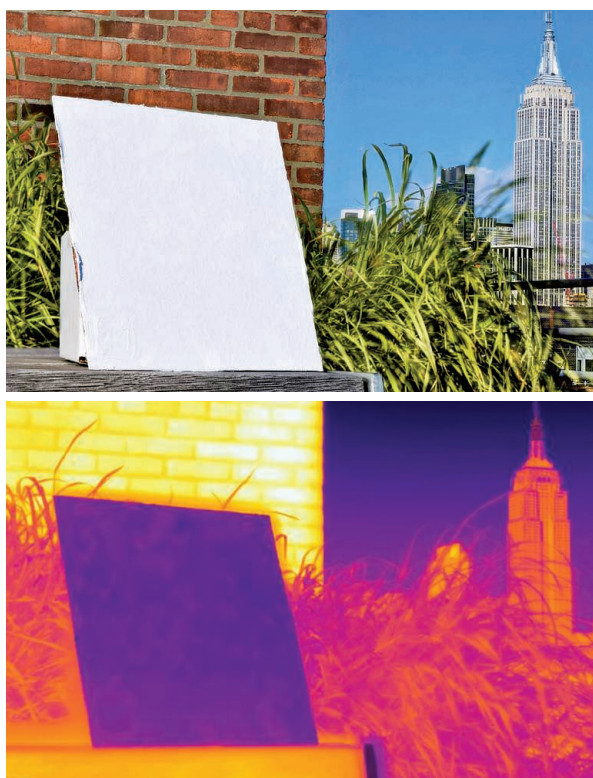
when he noticed that under certain conditions, the material turned from transparent to white as it dried. Under the microscope he saw that an array of interconnected air voids had formed in the dried film, causing it to strongly reflect light. (The same effect explains why snow is white even though ice cubes are transparent.)

The researchers then found they could re-create the effect with other polymers. Eventually, they settled on a highly durable commercial polymer called PVDF-HFP. The polymer starts as a solution in acetone, to which the researchers add a small amount of water. When painted on a surface, the acetone quickly evaporates, and the polymer separates from the water, creating a network of water droplets. Finally, the water also evaporates, leaving a sponge-like arrangement of interconnected voids that reflect up to 99.6% of light, including IR, visible, and UV. The dried film also emits heat primarily in the desirable mid-IR range.

Under the relentless midday sun of Phoenix, painted surfaces remained 6°C cooler than the surrounding air, the researchers report in a paper published online in

*Science* this week. And for good measure, they also showed that they could dye the paint, varying its appearance, although the colored paint sacrificed some cooling.

“It looks very good for being widely applicable,” Gentle says. Though durable, PVDF-HFP is roughly five times as expensive as traditional acrylics used in paints, says Ronnen Levinson, a cool roof expert at Lawrence Berkeley National Laboratory in California. But the added cost carries a benefit. Gentle says: “All these very cool coatings make it like nighttime in the middle of the day.” ■



Surfaces coated with a paint that efficiently reflects and radiates energy (top) drop in temperature, as seen in an infrared image (bottom).

and Geoff Smith of the University of Technology Sydney reported in 2015 that a cool roofing material made of a pair of polymers kept a commercial roof 3°C and 6°C cooler than ambient air in the midday sun and at night, respectively.

Applying these coatings to roofing and siding materials remains a sticking point. Highly reflective compounds can be integrated in traditional-looking shingles and clay tiles for new construction or renovation. But it’s been harder to come up with options for existing buildings.

That’s where the new passive cooling





Downloaded from <http://science.sciencemag.org/> on October 3, 2018

FEATURES

# BRIDGING THE GAP

In a bid to ease a history of mistrust,  
a summer workshop trains Indigenous scientists in genomics

By **Lizzie Wade**, in Seattle, Washington

**W**hen Ripan Malhi started graduate school in anthropology in 1996, his lab at the University of California (UC), Davis, housed what he saw as a valuable scientific resource: a freezer of Native American blood samples. Burgeoning genetic tools offered a chance to study the population history of these groups, especially the still-mysterious timing of their ancestors' arrival on the continent. Malhi began to extract and sequence DNA from the samples, which his adviser had collected over many years. As his research went on, however, Malhi realized there were few other Native American samples to compare with those on hand. So, he decided to collect more.

He kicked off his effort with a lecture at a reservation in Northern California. It was the first time he had spoken with a Native American community, despite years of studying their genetics. Expecting to gather dozens of DNA samples, "I brought a bunch of cheek swabs with me," he recalls. But at the end of his talk on DNA variation and the importance of filling in sampling gaps, the room fell uncomfortably silent. "Then one person stood up and said, 'Why should we trust you?'" Malhi remembers. "That's a formative memory. I had not learned about anthropologists going to communities, taking samples, and just leaving."

He got no samples that day.

Malhi's experience was one small manifestation of the ongoing tensions between Western scientists and Indigenous communities around the world. ("Indigenous" is an internationally inclusive term for the original inhabitants, and their descendants, of regions later colonized by other groups.) Scientists have used Indigenous samples without permission, disregarded their customs around the dead, and resisted returning samples, data, and human remains to those who claim them. Indigenous communities have often responded by severely restricting scientists' sampling of their bodies and their ancestors, even as genomics has boomed, with increasing relevance for health.

But today, more than 2 decades after his wake-up call in California, Malhi, now a molecular anthropologist at the University of Illinois (UI) in Urbana, is part of an effort to change the relationship between these communities. On a recent

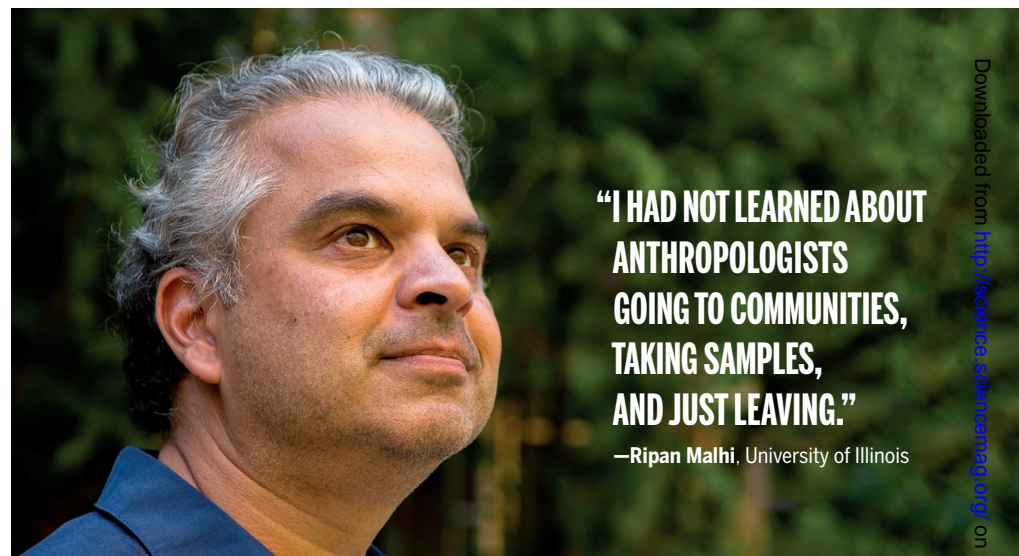
In 2010, members of the Havasupai Tribe look at DNA samples taken years before in an episode that fueled mistrust.

by the National Institutes of Health and the National Science Foundation (NSF), and held this year at the University of Washington (UW) here. About half of participants spoke in Indigenous languages spanning the globe from Alaska to New Zealand.

SING aims to train Indigenous scientists in genomics so that they can introduce that field's tools to their communities as well as bring a sorely needed Indigenous perspective to research. Since Malhi helped found it at UI in 2011, SING has trained more than 100 graduates and has expanded to New Zealand and Canada. The program has created a strong community of Indigenous scientists and non-Indigenous allies who are raising the profile of these ethical issues and

"SING is so important," says geneticist Rasmus Nielsen of UC Berkeley, who is not involved in the program. Those who have taken part say it has equipped them with increased awareness of Indigenous concerns and how to prioritize them in research. In response to new attitudes, some communities say they might now consider working with geneticists. SING is also building what may be the best kind of bridge, one that is "the obvious solution" to the problem of mistrust, Nielsen says: creating "a new generation of geneticists within Indigenous groups."

**ANY COMMUNITY DEMANDING** that researchers slow down, change their questions, destroy samples, keep data private, and per-



**"I HAD NOT LEARNED ABOUT ANTHROPOLOGISTS GOING TO COMMUNITIES, TAKING SAMPLES, AND JUST LEAVING."**

—Ripan Malhi, University of Illinois

developing ways to improve a historically fraught relationship.

SING grads and professors say the experience has profoundly affected their work. At SING, "you can exist as your authentic self, as both Indigenous and as a scientist, without having to code-switch all the time. It's like coming up for air," says Savannah Martin, a Ph.D. student in biological anthropology at Washington University in St. Louis, Missouri, and a member of the Confederated Tribes of Siletz Indians in Oregon.

SING participants are beginning to make waves in the broader scientific community. This year, SING alumni and faculty published ethical guidelines for genomic studies in *Science* (27 April, p. 384) and in *Nature Communications*. Echoing discussions at the workshops, those guidelines call for intense community engagement, especially in areas where Indigenous priorities may clash with those of Western science: questions of which research questions to tackle, when—or even whether—to publish, and how to handle samples and data.

haps not even publish their results is bound to face skepticism from Western scientists. Some Indigenous communities, such as the Navajo Nation, decline to participate in genetic research at all. And many tribes don't permit research on their ancestors' remains. Such opposition can feel like a hostile stumbling block to Western scientists, some of whom have gone to court to gain or maintain access to Indigenous samples. Not being able to study at least some early samples would "result in a world heritage disaster of unprecedented proportions," the American Association of Physical Anthropologists said in 2007 in a debate over an amendment to the Native American Graves Protection and Repatriation Act.

To understand why so many Indigenous people distrust Western scientists, consider how intertwined science has been with colonialism, says SING co-founder Kim TallBear, an anthropologist at the University of Alberta in Edmonton, Canada, and a member of the Sisseton Wahpeton Oyate in North and South Dakota. "While the U.S.



was moving westward, stealing land, and massacring Indians, you had contract grave robbers coming out onto the battlefields and immediately picking up the dead—Native people—and boiling them down to bone, and sending their bones back east,” she says. Many of those skeletons were displayed and studied in museums by researchers who used them to argue for the biological inferiority of Indigenous people. Some of those skeletons are still there.

“Science was there, always. It’s part of that power structure,” TallBear says. Just 20 years ago, researchers sued for and won the right to study the Ancient One, also called Kennewick Man, a 9000-year-old skeleton from Washington, over the objections of Indig-

tion and the Apache in the U.S. Southwest, to see how they might be related. “When I was meeting with tribes to tell them about the project, they were very enthusiastic,” Gachupin remembers. “Every tribal community that I went to gave approval on the first visit.”

But at the same time, researchers working for the Human Genome Diversity Project (HGDP), a major international effort, were collecting samples from around the world to build a public database of global genetic variation. The project publicly emphasized the importance of collecting DNA from genetically isolated Indigenous populations before they “went extinct.”

That rationale “was offensive to Indigenous

and journals require scientists to make data public, so others can check their work. But that makes scientists the custodians of data, and it’s scientists who decide what research questions to ask and how to present the results. Many Indigenous people don’t want to cede such control to researchers they don’t know and don’t trust, let alone to the entire scientific community.

Gachupin, now an epidemiologist at The University of Arizona in Tucson and a SING faculty member, represents tribes when scientists want to work with them, to make sure the tribes’ wishes are respected.

Another such pioneer is Nanibaa’ Garrison, a member of the Navajo Nation. She was in college when her tribe passed its moratorium on genetic research. (According to an article in *Nature* last year, the tribe may lift the ban.) But Garrison went on to earn a Ph.D. in the field. “I wanted to find a way to do it better. To do things right,” she says. She’s now a bioethicist at UW and the Seattle Children’s Research Institute, developing ethical approaches to research with Indigenous communities. When Malhi got in touch about SING, she signed on right away. “I wanted to see more people like me,” in genetics, she says. “And I wanted to change the story.”

**AT SING THIS YEAR**, each day’s activities began and ended with Indigenous stories, songs, and prayers. In between, participants spent 6 days extracting and analyzing their own mitochondrial DNA, getting a crash course in bioinformatics, critiquing informed consent forms, and talking about the questions DNA can and can’t answer. Students spanned the educational spectrum, from undergraduates to public health professionals.

“We’re not trying to shelter [anyone] from Western mainstream thought,” Malhi says. The bioinformatics workshop even uses the HGDP reference panel—once so controversial—because it allows students to learn about both its uses and its fraught history. But this year’s program ended with an exercise that reminded participants of the complex social backdrop of such research: a drama about a fictional project to look for genetic links to suicide in an Indigenous community.

As students and faculty adopted roles such as researcher or at-risk youth, conflicts quickly arose: At-risk teens refused to offer blood samples for research that might stigmatize them. Public health workers pressed for holistic programs. Pharma reps gave proforma lectures. Before long, the university researchers who proposed the study quietly disappeared. Overwhelmed, they decided to go back to their labs and work on something easier, they admitted at the end of the exercise. Laughs of recognition rang through the

**“IF YOU’RE GOING TO WORK WITH INDIGENOUS COMMUNITIES ON GENETICS, YOU HAVE TO BE WILLING TO MAKE LIFELONG RELATIONS.”**

—Kim TallBear,  
University of Alberta



enous groups. (The Ancient One was eventually returned to five tribes who claimed him in 2017, after DNA testing suggested a genetic link between him and living tribal members.)

Many Indigenous communities see echoes of this painful history reverberating in the 21st century. In 2003, the Havasupai Tribe in Arizona discovered that samples taken for a study on diabetes had been used for research projects they had never consented to, including on population genetics and schizophrenia. They sued Arizona State University in Tempe, which eventually returned the samples and paid \$700,000 to the tribe (*Science*, 30 April 2010, p. 558).

Missteps by Western researchers have even hampered work by Indigenous scientists. For example, in the 1990s, Francine Gachupin, a member of the Pueblo of Jemez in New Mexico, was working on a Ph.D. in anthropology at The University of New Mexico in Albuquerque. She wanted to collect genetic samples from speakers of Athabaskan languages, who range from some Alaska Native groups to the Navajo Na-

populations worldwide,” Gachupin says. “Resources for infrastructure and for the well-being of the community were not forthcoming, and yet now here were these millions and millions of dollars being invested to ‘save’ their DNA.” The message from the scientific establishment was, she says, “We don’t care about the person. We just want your DNA.” Some activists dubbed the HGDP “the Vampire Project,” believing the only beneficiaries would be Western scientists and people who could afford costly medical treatments.

In the United States, Native American support for genetic research “changed overnight,” Gachupin says. She put her research on hold because tribes became so worried about data protection. She eventually finished her work, but the tribes “were not going to give permission for anything more.”

Meanwhile, the HGDP database, which includes more than 1000 samples from 51 populations worldwide, went on to become a key genetic reference panel.

What happens after data are collected can also lead to conflict. Many granting agencies



Participants in the Summer Internship for Indigenous Peoples in Genomics study DNA extraction techniques at the University of Washington in Seattle this summer.

classroom, as participants noted just how complicated ethical research with Indigenous communities can be.

In real life, everyone at SING aims to be the researcher who won't disappear. Lessons gleaned from the workshop may help. One key, TallBear says: "If you're going to work with Indigenous communities collaboratively on genetics, you have to be willing to make lifelong relations."

Malhi, for example, has spent years building relationships with the First Nations of British Columbia in Canada, particularly the Metlakatla and Lax Kw'alaams. He has shifted away from focusing solely on his original questions about ancient migrations to questions that matter to the communities themselves, such as their relationships with their ancestors. His study of ancient DNA published in the *Proceedings of the National Academy of Sciences* in 2017, for example, showed at least 10,000 years of genetic continuity in the region, supporting Indigenous oral traditions.

Malhi's graduate student Alyssa Bader, an Alaska Native with ancestors from British Columbia, is now studying the oral microbiome of these communities' ancestors by sequencing DNA preserved in their dental plaque. That's less destructive than sampling bones or teeth, and can reveal what these ancient North Americans ate, a subject their contemporary descendants are interested in because their traditional diets have been altered by Western foods.

SING has helped forge new research relationships. Through the program Deborah Bolnick, an anthropological geneticist at the University of Connecticut in Storrs, has established a collaborative research project with Indigenous partners in the southern United States. It took 4 years of conversation before they collected a single sample, but now they have nearly 150. One project is to see whether

maternally inherited mitochondrial DNA (mtDNA) corresponds with the communities' matrilineal clans. If so, mtDNA analyses might be able to restore clan identities to community members who had that knowledge stripped from them by colonization.

Malhi and Bolnick both say the communities they work with will always have absolute control over their samples and data, and even whether and how they publish their results. That's because many Indigenous people, still facing racism, worry that certain types of studies—such as the one on genetic risk for suicide in the SING role play—may further stigmatize them. As a non-Indigenous researcher, "You have to be willing to know that history and put in the labor to get beyond that," Bolnick says. "To do this work you have to be willing to not see yourself as the authority, but rather as somebody who is going to listen to other authorities."

**THE INDIGENOUS RESEARCHERS** SING aims to foster understand that history better than almost anyone. They are likely to remain a small minority, at least in the near future: In the United States, less than 1% of doctorates are awarded to American Indian and Alaska Native students, according to NSF, a statistic that has held steady since 2006. But SING offers them the chance to collectively think through whether and how they want to use genetic tools to study their own people. "If you're working with your own community, you're less likely to back out when you hit a wall," says Anežka Hoskin, a graduate student in genetics at the University of Otago in Dunedin, New Zealand, and a member of the Māori tribes Ngāti Porou and Ngāti Kahu. "And you're going to hit walls."

Martin says doing research with Indigenous people has prompted difficult reflections. She's studying the biological effects of racism and historical trauma on tribes in

the Pacific Northwest—work she hopes will include searching for epigenetic changes linked to that history. But she was wary of what might happen if a university or a granting agency demanded access to her samples. Going to SING for the first time in 2015 helped her figure out how to present data protection as a priority in grant proposals. "SING made me feel a lot more comfortable with pushing back against Western institutions," she says.

The tribes she works with have full control over their samples and data—and will decide whether the results are published. If not, "That's it, I don't get my Ph.D.," Martin says. "I've made my peace with that ... Indigenous sovereignty is more important to me than three letters after my name."

SING faculty member Keolu Fox, a postdoc in genetics at UC San Diego and a Native Hawaiian, sees a future in which genomics supports Indigenous self-governance rather than undermines it. "Our genomes are extremely valuable," he says. For example, he's starting to study a genetic variant first identified among Polynesian populations, including Native Hawaiians, that may protect against heart disease and diabetes, especially in people with high body mass indexes. It should be Polynesian communities who profit from that research, he says.

SING faculty member Rene Begay, a geneticist at the University of Colorado Anschutz Medical Campus in Aurora and a member of the Navajo Nation, is excited about her role in building this bridge. "I want to be at the table, to advocate for my people, to advocate for research," especially studies that may improve health care, she says. "I want us to ... have the advancements and the technologies that the world outside the Navajo Nation has. But I want to do it in a way that's on our terms." ■



## PERSPECTIVES

### BIOCHEMISTRY

## The making of a plankton toxin

A combined genetic and chemical study reveals the biosynthetic steps involved in making the toxin domoic acid

By Georg Pohnert, Remington X. Poulin,  
Tim U. H. Baumeister

Phytoplankton are unicellular photosynthetic organisms that occur in vast abundance in the world's oceans. This diverse group of algae is responsible for nearly half of the total global CO<sub>2</sub> fixation and forms the base of the marine food web. However, some phytoplankton taxa produce toxins that can poison marine wildlife and humans. Mass occurrences of such toxic unicellular algae in the plankton occur periodically and are

known as harmful algal blooms (see the first photo). For example, the toxic diatom *Pseudo-nitzschia* (see the second photo) regularly emerges in the Pacific. In 2015, this microalga formed the largest harmful algal bloom ever recorded, which stretched along nearly the entire coast from Alaska to Mexico and caused massive ecologic and economic damage (1). On page 1356 of this issue, Brunson *et al.* (2) describe the biosynthesis of the toxin domoic acid that is produced by this diatom.

Domoic acid is the causative agent responsible for amnesic shellfish poisoning in

humans after eating contaminated seafood. Interest in this toxin increased after 1987, when consumption of blue mussels from Prince Edward Island led to an incident of severe poisoning and human fatalities (3). Single cells of the algae produce only picogram or smaller amounts of domoic acid, but the compound accumulates as it is transferred through the food web. Mussels feeding on the diatoms enrich the toxin without being affected. Fish are also vectors that are not poisoned themselves but transfer the toxin to higher trophic levels.

In humans, domoic acid acts as an analog of the neurotransmitter glutamate in the central nervous system, leading to symptoms that range from nausea and diarrhea, to confusion and other neurological effects, before coma and death (4). As a safety measure,

Institute for Inorganic and Analytical Chemistry,  
Friedrich Schiller University Jena, Jena, Germany  
Email: georg.pohnert@uni-jena.de





A toxic algal bloom clouds open ocean waters.

many countries have installed surveillance programs to monitor *Pseudo-nitzschia* in coastal waters and the domoic acid concentration in seafood; fisheries are shut down and beaches closed at times of contamination. Incidents of domoic acid poisoning of sea lions with more than 400 dead animals further boosted interest in the toxin (5). Seabirds are also affected, and historic samples from the 1960s suggest that the inspiration for Alfred Hitchcock's famous movie *The Birds* came from crazed animals affected by domoic acid poisoning (6).

The structures of many natural toxins, including that of domoic acid, have been known for decades, but their biosynthesis is not always understood. Knowledge of how nature assembles a complex toxin can facilitate follow-up research. For example, identification of the genes responsible for biosynthesis allows the heterologous expression and modification of a toxin and drug candidate for production in large quantities,

or the monitoring of toxin producers in the environment (7). In recent years, we have seen a number of biosyntheses that have been unraveled, including the biosynthetic pathway of the anticancer drug vinblastine and the identification of all enzymes involved in morphine biosynthesis, more than a century after its isolation in pure form in 1895 (8, 9). These breakthroughs have been made possible by the combination of advanced genetic methods, natural product chemistry, and physiological studies.

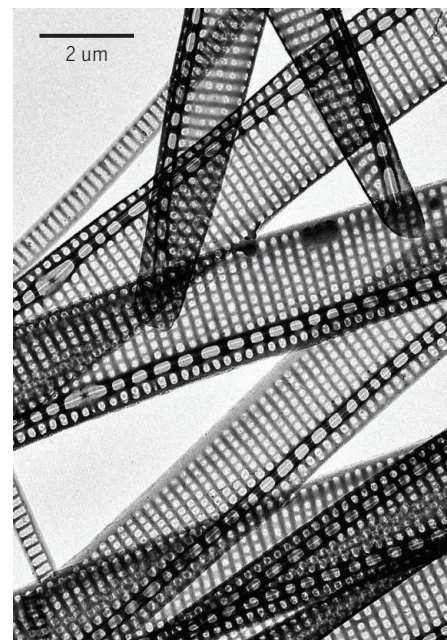
Brunson *et al.* use such a combined approach to uncover central steps in the biosynthesis of domoic acid in the harmful alga *Pseudo-nitzschia multiseries*. The endeavor was based on earlier observations that toxicity of the diatom increases under phosphate limitation and elevated CO<sub>2</sub> concentration. A comparison revealed that the transcription of ~500 genes increases under phosphate starvation. Transcription of only 43 of these genes also increased under high-CO<sub>2</sub> conditions; the authors used these transcripts to look for domoic acid biosynthetic genes.

The oxidase with the most pronounced increase in transcription was located in a genomic region where additional genes showed up-regulated expression. This clustering facilitated the breakthrough, because it turned out that all these genes were involved in domoic acid biosynthesis. The genes could be expressed in bacteria and yeast; application of tailored synthetic substrates, together with elaborate chemical analysis, led to the assignment of the biosynthesis of isodomoic acid. The authors have thereby uncovered the assembly line for producing an isomer of domoic acid, and in theory, this isomer can be transformed in one enzymatic step to the toxin, but the enzyme responsible is not yet known. Isodomoic acid is formed from two common primary metabolites, the amino acid glutamic acid and a terpenoid precursor geranyl pyrophosphate, in a few steps. The two precursors are first linked via *N*-prenylation and then further transformed by two oxygenases.

Knowledge of the biosynthetic pathway offers the opportunity to establish screening techniques for genetic sequences that indicate toxin production in plankton samples. These techniques will enable rapid bloom monitoring and environmental consulting. More importantly, the biosynthetic pathway could pave the way to answering scientific questions about this toxin. For example, the sequence opens the opportunity to address how metabolites with structural similarity to domoic acid are synthesized in red algae. Initial screening in these red algae has already revealed oxidases similar those required for domoic acid biosynthesis (2). Also, the ob-

served gene clustering, which is not common in diatoms but widespread in fungi and bacteria, will stimulate evolutionary discussion.

Answering the more fundamental question of why the diatoms produce the toxin is now also within reach. Given that there is no direct ecological connection between mammals or birds and the unicellular algae, toxicity to the top consumers might be considered as collateral damage caused by a compound that has different importance for the unicellular producer. Many studies have been undertaken to evaluate the role of domoic acid in inhibiting competing



Electron microscopy reveals the silicified cell wall of a *Pseudo-nitzschia* diatom from the Gulf of Naples.

phytoplankton (3), defending against herbivorous predators (10), or facilitating iron uptake (11). However, findings have been contradictory and do not provide a conclusive picture. Knowledge of the biosynthetic pathway can be used to elucidate domoic acid's function for the diatom by studying the induction of toxin production and by studying the physiological and ecological performance of mutant strains that lack domoic acid biosynthesis genes. ■

## REFERENCES

1. R. M. McCabe *et al.*, *Geophys. Res. Lett.* **43**, 10366 (2016).
2. J. K. Brunson *et al.*, *Science* **361**, 1356 (2018).
3. A. Lelong *et al.*, *Phycologia* **51**, 168 (2012).
4. K. A. Lefebvre, A. Robertson, *Toxicon* **56**, 218 (2010).
5. C. A. Scholin *et al.*, *Nature* **403**, 80 (2000).
6. S. Bargu *et al.*, *Nat. Geosci.* **5**, 2 (2012).
7. S. Galanie *et al.*, *Science* **349**, 1095 (2015).
8. L. Caputi *et al.*, *Science* **360**, 1235 (2018).
9. J. M. Hagel, P. J. Facchini, *Nat. Chem. Biol.* **6**, 273 (2010).
10. M. B. Olson, E. J. Lessard, *Harmful Algae* **9**, 540 (2010).
11. E. K. Prince *et al.*, *Mar. Drugs* **11**, 2398 (2013).

10.1126/science.aau9067



## DEVELOPMENT

# Hox genes and body segmentation

An ancient gene cluster controls the formation of repetitive body parts in a sea anemone

By Detlev Arendt

On page 1377 of this issue, He *et al.* (1) elucidate two long-standing problems in animal evolution: the ancient function of the homeobox (*Hox*) gene cluster, which has puzzled scientists for decades (2), and the centuries-old debate on the emergence of the segmented animal body.

*Hox* genes were first discovered in flies and mice, where they specify different body segments along the anterior-posterior (A-P) axis (3). Although their expression often overlaps in posterior body regions, they show spatially distinct anterior expression boundaries (4) (see the figure). Importantly, the A-P sequence of *Hox* gene expression in the body matches their 3' to 5' sequential occurrence within a chromosome cluster, a principle called spatial collinearity. Moreover, the more anteriorly expressed 3' *Hox* genes are often expressed earlier in development, which is called temporal collinearity. In addition, individual *Hox* proteins are typically active close to their anterior expression boundary, because the more 5', or posterior, proteins counteract the function of the more 3', or anterior, ones whenever both products coexist. This is called posterior prevalence (4).

Making sense of these rules has been challenging. For vertebrates, *Hox* collinearity and posterior prevalence are explained with the sequential activation of these genes from within a posterior growth zone (5). This growth zone iteratively produces mesodermal body segments, called somites, and pushes them toward the anterior. If, during this process, *Hox* genes are sequentially activated and stay on in the newly generated segments, then staggered expression of *Hox* genes arises along the A-P axis.

What is the relevance of spatial collinearity for invertebrates? Collinear *Hox* gene activity is described for patterning ectodermal derivatives such as the nervous system in insects (3). Few studies revealed *Hox* gene expression in mesodermal structures that resemble vertebrate somites. Moreover, *Hox* gene clusters are active in both segmented and unsegmented invertebrates such as sea urchins (6). This has prompted the view that the tight link between *Hox* genes and body segmentation observed in vertebrates, insects, or ane-

a cluster of only three *Hox* genes existed (8), with one anterior (3'), one middle, and one posterior (5') gene. The sea anemone has a fragmented version of the ancient three-gene *Hox* cluster, with additional, lineage-specific duplications (1). Previous expression comparisons revealed expression of *N. vectensis* *Hox* genes in the developing gastric pouches, with distinct *Hox* gene expression boundaries matching the epithelial boundaries between the pouches (9–11). These boundaries are collinear with the assumed sequence of genes in the cnidarian-bilaterian ancestor [referred to as trans-spatial collinearity (4)].

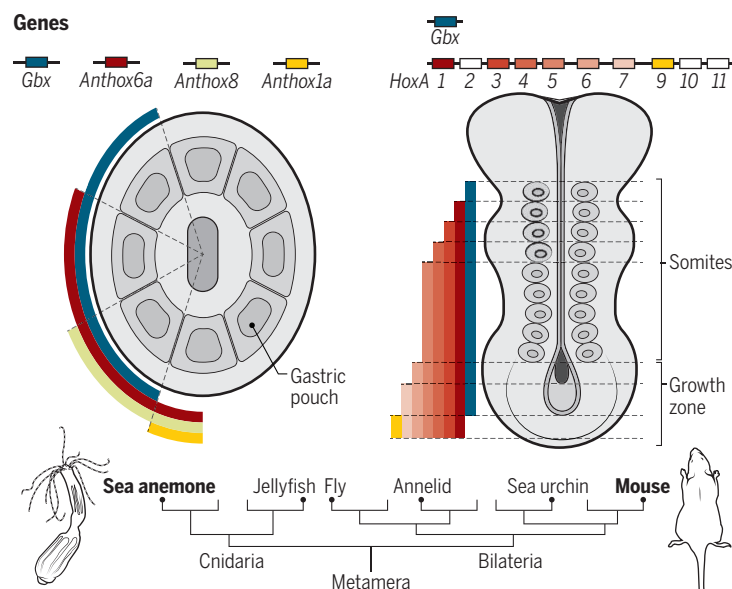
He *et al.* established gene ablation techniques in *N. vectensis* to investigate the function of the *Hox* genes. They uncovered a fundamental role of these gene products in the formation of the gastric pouches. In the absence of *Hox* activity, the epithelial folds separating these pouches are lost. Therefore, similar to vertebrates, *N. vectensis* *Hox* genes differentially and specifically control the generation of boundaries between repeated body parts, in conjunction with a later role in controlling the different fates of these parts (12). He *et al.* also show that the *N. vectensis* *Hox* genes are expressed sequentially, in concert with the stepwise appearance of folds during polyp development. Thus, *N. vectensis* *Hox*

genes show spatial and temporal collinearity. Additionally, the loss of each *Hox* gene abolishes the epithelial folding at the most anterior expression limit of that gene—where they are not coexpressed with a more posterior *Hox* gene. In the absence of the boundary formed by the folds, the two adjacent pouches fuse. These phenotypes are consistent with the posterior prevalence principle.

This is strong evidence that the link between *Hox* gene function and some kind of body segmentation is ancestral. But how does *N. vectensis* segmentation—the sequential

## Hox genes in sea anemone and mouse development

Collinear expression of gastrulation brain homeobox (*Gbx*), "anterior" *Hox* (*Anthox6a*), "middle" *Hox* (*Anthox8*), and "posterior" *Hox* (*Anthox1a*) genes defines gastric pouches in sea anemone, and that of *HoxA* genes defines somite positioning in mice.



lids has evolved independently, that is, by evolutionary convergence (7).

The best way to challenge this view is to investigate an evolutionary outgroup. Accordingly, He *et al.* investigated *Hox* gene function in the cnidarian *Nematostella vectensis*, the starlet sea anemone. Cnidarians are our most distant relatives to possess a *Hox* cluster (see the figure). Their inner surface is folded, so that their primitive gut is subdivided into chambers, called gastric pouches. These are also continuous with the lumen of the tentacles. The cnidarian lineage diverged from ours when

generation of gastric pouches by epithelial folding—relate to bilaterian segmentation? The simple mode of epithelial folding observed in cnidarians has inspired morphologists for more than 150 years, who saw similarities with mesodermal segmentation in bilaterians. Bilaterian mesodermal somites often emerge as outpocketings from the gut. The enterocoel theory proposed that both cnidarian pouches and bilaterian somites are derived from such outpocketings (7, 13). The findings of He *et al.* support this theory.

One possible caveat is that the sequentially emerging *N. vectensis* folds are not generated from a posterior growth zone. This might represent a secondary simplification of cnidarian development, given that in the tube anemones (Cerianthida), which represent an early diverging branch of Cnidaria, additional folds are pushed forwards from a terminal growth zone (9). Examining the expression of *Hox* genes and of growth zone markers in these cnidarians could be especially rewarding, as it might establish a similar link between *Hox* spatial and temporal collinearity and the generation of body segments from a growth zone as is observed in vertebrates (4, 5). Another note of caution concerns the unsolved axial relationships between cnidarians and bilaterians (14), which led to conflicting views about the nature of the cnidarian *Hox* axis (15). However, the data of He *et al.* seem to firmly settle this issue.

Twenty-five years after the revolutionary zootype concept, which considered the *Hox* genes a shared feature of all animals (2), we can now refine this idea and propose that the *Hox* gene cluster evolved with the emergence of segmental epithelial folds and pouches in the cnidarian-bilaterian ancestor. These persisted as gastric pouches in today's cnidarians and gave rise to mesodermal somites in the bilaterians. ■

#### REFERENCES

1. S. He *et al.*, *Science* **361**, 1377 (2018).
2. J. M. W. Slack, P. W. H. Holland, C. F. Graham, *Nature* **361**, 490 (1993).
3. M. Akam, *Cell* **57**, 347 (1989).
4. D. Duboule, *Development* **134**, 2549 (2007).
5. J. Deschamps, D. Duboule, *Genes Dev.* **31**, 1406 (2017).
6. R. Mooi, B. David, *Annu. Rev. Ecol. Syst.* **39**, 43 (2008).
7. D. Tautz, *Dev. Cell* **7**, 301 (2004).
8. D. E. K. Ferrier, in *Hox Genes: Studies from the 20th to the 21st Century*, J. S. Deutsch, Ed. (Landes Bioscience and Springer Science, 2010), pp. 91–100.
9. D. Arendt *et al.*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20150286 (2015).
10. G. Genikhovich *et al.*, *Cell Rep.* **10**, 1646 (2015).
11. D. Arendt, M. A. Tosches, H. Marlow, *Nat. Rev. Neurosci.* **17**, 61 (2016).
12. T. Alexander, C. Nolte, R. Krumlauf, *Annu. Rev. Cell Dev. Biol.* **25**, 431 (2009).
13. A. Sedgwick, *Q. J. Microsc. Sci.* **24**, 43 (1884).
14. C. Nielsen, T. Brunet, D. Arendt, *Nat. Ecol. Evol.* **2**, 1358 (2018).
15. T. Q. DuBuc *et al.*, *Nat. Commun.* **9**, 2007 (2018).

10.1126/science.aav0692

#### DNA REPLICATION

## No strand left behind

Histone chaperones direct how epigenetic information is inherited in dividing cells

By Kami Ahmad<sup>1</sup> and Steven Henikoff<sup>1,2</sup>

**S**emiconservative replication of DNA faithfully transmits genetic information, but the copying of epigenetic information, which distinguishes cell identity and is embedded in chromatin, is more complicated: Unwinding and separating the parental double-stranded DNA for replication displaces all chromatin proteins, including the histones that package eukaryotic DNA into nucleosomes. This necessitates that chromatin organization be reestablished after every round of DNA replication. Furthermore, bound chromatin proteins must double after DNA replication, as one parental DNA chromatid becomes two daughters. The rules governing the distribution of histones to two daughter chromatids

**“Why are chaperones required to achieve the seemingly simple outcome of random symmetric parental-histone partitioning...?”**

have long been the subject of speculation. However, on pages 1389 and 1386 of this issue, Petryk *et al.* (1) and Yu *et al.* (2), respectively, show that histone distribution is inherently asymmetric, but they identify mechanisms of chromatin regulation to achieve nearly equal distribution of parental nucleosomes to daughter chromatids. These findings have implications for how epigenetic information is propagated through cell divisions.

Chromatin assembly in the wake of the replication fork relies on two distinct processes: first, the transfer of parental histones, and second, the deposition of new histones. Experiments with bulk chromatin demonstrated the retention of parental histones on daughter strands and enabled isolation of histone chaperones, which pro-

mote the efficient deposition of new histones in the wake of the replication fork (3), but distinguishing how exactly this happens has been limited by methods to track histones on daughter chromatids. This is critical because all chromatin features (including histone modifications and nucleosome positioning relative to DNA sequence) must be copied onto both daughter chromatids to be propagated from one cell generation to the next. The two daughter chromatids differ in how they are replicated (see the figure), constrained by the requirement of DNA polymerases (Pols) to synthesize DNA in a 5'-to-3' direction: the top “leading” DNA strand is replicated rapidly and processively by DNA Pol  $\epsilon$ , whereas the bottom “lagging” DNA strand is replicated in interspersed segments of RNA primers and new DNA fragments by DNA Pols  $\alpha$  and  $\delta$ . These Okazaki fragments are further processed to displace the RNA primers and fill in the gaps to complete the double-stranded daughter chromatid. These complicated events on the lagging strand delay its completion, resulting in the generation of an extensive single-stranded DNA loop of the lagging strand behind the replicative helicase [the minichromosome maintenance (MCM) complex consists of MCM2 to MCM7].

How are these two strands packaged in the wake of the replication fork? The bulk of parental nucleosomes are distributed to daughter chromatids with tetramers of histone H3 and histone H4 [referred to as (H3-H4)<sub>2</sub>], intact, and two H2A-H2B dimers are added to complete the nucleosome (3). Early experiments with viral circularized DNA (minicircles) concluded that parental histones were equally partitioned between the two daughter strands, with the remaining gaps filled by new histone deposition (4). This conclusion was further supported by genome-wide mapping of DNA replication forks (5, 6).

The studies of Petryk *et al.* and Yu *et al.* reveal that this partitioning is carefully orchestrated by the replicative machinery with the help of specific histone chaperones that are well positioned to capture histones displaced from the parental chromatid. Yu *et al.* used genome-wide mapping to follow old and new nucleosome assembly on the leading and lagging strands in budding

<sup>1</sup>Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. <sup>2</sup>Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. Email: kahlmad@fredhutch.org; steveh@fredhutch.org



yeast, detecting a slight lagging-strand preference for deposition of parental histone (H3-H4)<sub>2</sub> tetramers. Deletion of the two histone-binding proteins, Dpb3 and Dpb4, which are subunits of DNA Pol  $\epsilon$ , increased lagging-strand preference, revealing a role for histone chaperones in regulating parental histone-assembly preference. In mouse embryonic stem cells, Petryk *et al.* mapped Okazaki fragments to determine replication fork directionality and to map parental and new histone deposition across the genome. They detected a slight preference of parental histones for the leading strand. By mutating histone-binding residues within a histone-binding domain of MCM2, a replicative helicase subunit, they observed a dramatic loss of symmetric assembly, whereby parental histones were predominantly deposited on the leading strand and new histones on the lagging strand. This implies that deposition of parental (H3-H4)<sub>2</sub> tetramers on the leading strand is the default state but that deposition is delayed by binding to a chaperone domain of the replicative helicase.

Why are chaperones required to achieve the seemingly simple outcome of random symmetric parental-histone partitioning into daughter chromatids? DNA replication in eukaryotic cells displaces ~10 nucleosomes each minute, and displaced histones must be quickly captured. The leading daughter strand is available for capturing displaced histones first, because it is produced much faster than the lagging one. If binding to replicative chaperones delays histone transfer, this would allow replication of the lagging strand to be completed and thereby promote equal partitioning.

The finding that nucleosome partitioning is carefully controlled has important implications. Gene expression requires the precise localization of chromatin features, including the localization of histone modifications to promoters and regulatory elements, and the binding of transcription factors to their cognate DNA sites. The relatively slow reassembly of chromatin in the wake of the replication fork probably provides a window of opportunity for displaced transcription factors to rebind DNA, maintaining gene expression programs (7). The equal partitioning of histones allows equal transcription factor-binding on both daughter chromatids, ensuring that daughter cells maintain the

gene expression program of their lineage.

The symmetric partitioning of parental histones also creates a problem for copying patterns of histone variants and modifications. Each daughter chromatid only receives half the parental histones; thus, parental modification patterns are diluted. Further, precise nucleosome positions are not maintained because nucleosomes are displaced and then reassembled. This explains why chromatin features in eukaryotic genomes primarily rely on transcription factors that bind precisely at DNA sequences in the genome, recruiting histone modifying activities and chromatin remodelers to reestablish genomic

H3.3-containing nucleosomes are split (by an unknown chaperone) into two H3.3-H4 dimers. This allows “semiconservative” partitioning of an H3.3-H4 dimer to each daughter chromatid (10, 11). These kinds of regional controls over partitioning can be imagined to affect chromatin accessibility and the dilution of histone modifications.

The regulated distribution of nucleosomes assures equal partitioning, but is unequal partitioning ever beneficial? Although, for most cells, gene expression programs in daughter cells should resemble those of the parent cell, stem cells are an exception, where one daughter cell retains stem cell identity and the other begins to

differentiate. Notable differences in chromatin accompany this division in the fruitfly male germline, whereby the daughter stem cell retains the bulk of the parental histones and the differentiating daughter cell genome is packaged with new histones (12). This is a spectacular example of asymmetric partitioning between cells, and preliminary data show that this is the consequence of asymmetric histone partitioning in the wake of the replication fork (13). Asymmetric partitioning in this example results from greatly delaying the completion of the lagging strand; thus, only the leading strand is available to capture displaced histones. Asymmetric partitioning in the wake of the replication fork may promote initiation of a differentiation program because the daughter cells with lagging-strand chromatids lose all pa-

rental histone modifications and are thus open for new transcription factor binding. Understanding how asymmetric nucleosome assembly behind the replication fork may be coordinated with developmental events remains an exciting future challenge. ■

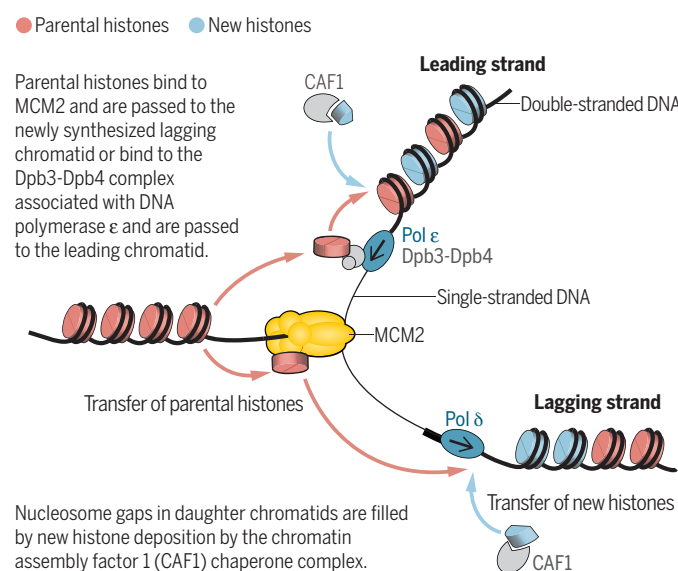
## REFERENCES

1. N. Petryk *et al.*, *Science* **361**, 1389 (2018).
2. C. Yu *et al.*, *Science* **361**, 1386 (2018).
3. A. T. Annunziato, *Genes (Basel)* **6**, 353 (2015).
4. M. E. Cusick *et al.*, *J. Mol. Biol.* **178**, 249 (1984).
5. D. J. Smith, I. Whitehouse, *Nature* **483**, 434 (2012).
6. P. Vasseur *et al.*, *Cell Rep.* **16**, 2651 (2016).
7. S. Ramachandran, S. Henikoff, *Cell* **165**, 580 (2016).
8. D. Reinberg, L. D. Vales, *Science* **361**, 33 (2018).
9. M. Foltman *et al.*, *Cell Rep.* **3**, 892 (2013).
10. M. Xu *et al.*, *Science* **328**, 94 (2010).
11. C. Huang *et al.*, *PLOS Genet.* **9**, 618 (2013).
12. V. Tran, C. Lim, J. Xie, X. Chen, *Science* **338**, 679 (2012).
13. M. Wooten *et al.*, *bioRxiv* 242768 [Preprint], 23 August 2018.

10.1126/science.aav0871

## Histone partitioning at the replication fork

Chromatin-associated proteins, including parental histones, are displaced from DNA as the MCM helicase melts double-stranded DNA into single-stranded DNA.



landscapes. In limited cases, modified nucleosomes may recruit the cognate modifying enzyme, leading to modification of new histones and restoration of modification patterns (8). It is notable that many of the replicative histone chaperones are required for heterochromatic (densely packed chromatin) silencing of gene expression (6, 9). Indeed, heterochromatic silencing of the yeast mating-type locus is defective in the absence of Dpb3 and Dpb4 (2), underlining the importance of delaying histone partitioning in regions where completion of the lagging strand is delayed.

Replicative histone chaperones also allow histone partitioning to be altered. Histone variants form specialized nucleosomes in certain regions of genomes. The H3.3 histone variant is enriched in active chromatin regions, and during replication, parental

# Above and below the Maya forest

Advanced remote sensing technology raises questions about settlement and land use

By Anabel Ford<sup>1</sup> and Sherman Horn<sup>2</sup>

Reports of lost cities in the Maya forest began with the remarkable pictorial folios that Stephens and Catherwood produced from their journeys in the 1830s (1) and that inspired the Carnegie Institution of Washington's expeditions in the early 20th century (2). This research focused on major architecture: the grandest temples, plazas, tombs, and palaces built for Maya aristocrats. The singular attention paid to rulers illuminated Maya politics and religion but left a major lacuna in knowledge of the populace. To fill this gap, later fieldwork mapped the humble housemounds to interpret settlement patterns at the detailed site scale (3–5). On page 1355 of this issue, Canuto *et al.* (6) use a remote sensing method called light detection and ranging (lidar) to examine regional-scale settlement patterns and implications for sustainability in the Maya forest. Their conclusions support extensive settlement densities, as has been expected, and suggest substantial land modifications that affect assessments of sustainability.

Twentieth-century scientific surveys in the Maya area (4–6) typically proceeded by cutting a breach—a *brecha*—through the jungle and then using paced distances and compass bearings to compile a mapped 250- to 500-m-wide transect. Striking into the unknown, surveyors relied on calculated azimuths, basic ordnance maps, and aerial photos. These tools served archaeologists well, providing a basis for discovering settlement patterns and interpreting land use, as seen at El Pilar where settlement patterns have demonstrated sustainability at the local scale (7, 8). Does lidar supersede the traditional tools of the tropical surveyor's kit? The revelation of the topography below the forest canopy by lidar, which illuminates target areas with laser light and measures reflectance to create three-dimensional rep-

resentations of the ground surface, is like a magic wand. Rather than striking into the unknown, we now can see the landscape and features that may relate to ancient settlement. The interpretations of cultural remains are only impressions, however, until the remotely sensed images are verified by teams on the ground. Interpretations of ancient land use depend on archaeological survey to locate and document prehistoric settlements. Understanding settlement patterns over large areas requires extrapolation from samples of surveyed land, which has proved challenging in the densely forested lowlands occupied by the ancient Maya. Lidar clearly visualizes the geographic essentials critical to

which are extrapolated to a population of 7 to 11 million across the 95,000 km<sup>2</sup> of the central Maya Lowlands in the Late Classic (600–900 CE). These broad extrapolations are based on the 61,480 structures identified in the lidar visualizations. Canuto *et al.* used existing maps to interpret unvisited areas. This procedure will produce reasonable results at the 1:250,000 regional scale offered by Canuto *et al.*, and even at the 1:50,000 local scale (8). However, fieldwork is needed at the detailed 1:10,000 site scale. Large features are clearly visible. Field validation should follow up on the proposed terrain modifications (such as the wetland drainages) as well as small, low features (such as isolated structures) that affect our understanding of human-environment interactions (11).

Interpretations by Canuto *et al.* raise intriguing hypotheses. The authors define a rural-to-urban continuum based on preliminary structure density and propose a threshold of sustainability beyond which food importation would be required. They identify features in lidar imagery, such as terraces and canals, that are ascribed to agricultural intensification. Investing in infrastructure to overcome environmental limitations along with putting marginal land (areas with too much or too little water) under cultivation is one



Dense forest covers the Verdant Maya Biosphere Reserve, as seen from El Pilar.

understanding settlement patterns: topography, drainage, and slope. Canuto *et al.* now bring together 18 archaeologists working in 10 areas, covering more than 2000 km<sup>2</sup>, to examine regional relationships between environment and settlement. In addressing land use sustainability in the Maya Lowlands, where contemporary land use is expanding at the expense of the forest, Canuto *et al.* suggest that the ancient Maya could have faced similar issues. We, too, have been working with lidar at El Pilar to address sustainability in the tropics (9) and to ask whether deforestation is the only outcome of ancient Maya settlement expansion, as the prevailing narrative in Maya archaeology proclaims (10).

On the basis of a concordance of lidar visualizations, existing maps, and preliminary field examinations, Canuto *et al.* estimate population densities of 80 to 120 persons/km<sup>2</sup>,

strategy for intensifying food production. But what about strategies that intensify food production without infrastructure? Agricultural intensification strategies that rely on labor, skill, and scheduling can increase food yields but are not reflected in surface features detected by lidar. Intensified food production can result from the management of infield forest gardens around structures as well as in outfields. Again, on-site geographic and archaeological validation will resolve imagery ambiguities and provide the appropriate scale to assess land use alternatives and to validate agro-engineering features.

The terrestrial features revealed by lidar technology are indeed exciting and have opened the door to the next step of studying the Maya forest: assessment of the imagery. As exciting as the imagery is, we need to understand the nuances of where small, and

<sup>1</sup>ISBER/MesoAmerican Research Center, University of California, Santa Barbara, CA 93106, USA. <sup>2</sup>Department of Anthropology, Grand Valley State University, Allendale, MI 49401, USA. Email: ford@ucsb.edu



presumably domestic, structures may or may not be present. With standard Global Positioning Systems, one can locate and validate features identified remotely in lidar imagery within 7- to 14-m accuracy. Structure locations and densities as well as the “zones of intensive cultivation” may well change with field inspection, as lidar has a tendency to image false positives (11). The archaeological survey methods developed in the Maya forest over the past 60 years are basic to the validation process (3–5). Archaeologists with “boots on the ground” will be able to determine whether proposed structures are instead treefalls, palm debris, or technical errors in the lidar visualizations. Equally challenging are linear features—presumed signatures of land modifications (particularly in wetlands)—that may be artifacts of lidar visualizations. While lidar technology has improved over the past decade, it is the innovative visualization methods that are critical (12). Assessment of lidar imagery will help the field evolve the use of this technology. These developments are where the greatest prospects for the future of archaeology lie.

Canuto *et al.* recognize that fieldwork is the way forward and prophetically conclude that lidar will become indispensable for settlement research. The authors also present an excellent example of a regional scale evaluation of settlement patterns, and by proxy, to analyze sustainability. They suggest that the patterns they elucidated are not to be solely based on geography; political and economic forces must be evoked to explain Maya land use. Their interpretation provides critical hypotheses for future testing and is a call to action for archaeologists to scrutinize the findings with survey at the detailed site scale. As usual, the devil is in the details. ■

## REFERENCES

1. J. L. Stephens, F. Catherwood, *Incidents of Travel in Yucatan* (Univ. of Oklahoma Press, 1854/1962; 2 volumes).
2. R. E. W. Adams, *Lat. Am. Res. Rev.* **4**, 3 (1969).
3. W. R. Bullard Jr., *Am. Antiq.* **25**, 355 (1960).
4. D. E. Puleston, thesis, University of Pennsylvania (1973).
5. A. Ford, *Population Growth and Social Complexity: An Examination of Settlement and Environment in the Central Maya Lowlands* (Arizona State University, 1986).
6. M. A. Canuto *et al.*, *Science* **362**, eaau0137 (2018).
7. A. Ford, R. Nigh, *The Maya Forest Garden: Eight Millennia of Sustainable Cultivation of the Tropical Woodlands* (Left Coast Press, 2015).
8. A. Ford, K. C. Clarke, G. Raines, *Ann. Assoc. Am. Geogr.* **99**, 496 (2009).
9. A. Ford, P. Morales, S. Horn, in *XXXI Simposio de Investigaciones Arqueológicas en Guatemala*, B. Arroyo, L. Méndez Salinas, L. Paiz, Eds. (Ministerio de Cultura y Deportes, Instituto de Antropología e Historia, Asociación Tikal, Guatemala, 2018).
10. B. L. Turner 2nd, J. A. Sabloff, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13908 (2012).
11. A. Ford, *Res. Rep. Beliz. Archaeol.* **11**, 270 (2014).
12. T. J. Pingel, K. C. Clarke, A. Ford, *Cartogr. Geogr. Inf. Sci.* **42** (suppl. 1), S18 (2015).

10.1126/science.aav0887

## CANCER

# How dormant cancer persists and reawakens

Insights reveal possible avenues to prevent metastasis

By Julio A. Aguirre-Ghiso

**T**reatment of stage IV cancer, when metastases are detectable by conventional imaging methods, is commonly the only strategy we have to defend against this lethal stage of disease.

However, the seeds of those lesions, the disseminated cancer cells (DCCs), can reside in a patient's organs for long periods before their outgrowth (1). DCCs are proposed to be able to persist for years because they interpret homeostatic signals from the host microenvironment and respond by entering a long-lasting dormant state, with occasional cell divisions (1, 2). In response to signals not yet fully identified, dormant DCCs reawaken and grow into larger lesions. So why not intervene earlier when DCCs are dormant? This has presented a challenge because the biology of DCC dormancy was mechanistically unclear and there was no way to determine whether DCCs were dormant. However, two decades of mechanistic research (1, 2) and recent work by Pommier *et al.* (3) and on page 1353 of this issue by Albregues *et al.* (4) have provided important insights into the microenvironmental control of DCC dormancy and reawakening that leads to metastasis, which could have therapeutic implications.

The dormancy of DCCs has been mostly explained by the ability of these cancer cells to enter quiescence, a reversible and programmed growth arrest (1). Clinical evidence and experimental models support this mechanistic view, which helps to explain why in some cancers (for example, estrogen receptor-positive breast cancer and prostate cancer), patients can be free of disease for more than a dozen years. The implications of this biology are profound; it offers a window of opportunity to treat metastasis early and to understand how host and epigenetic factors override the driver genetics that supposedly can only fuel cancer growth. This new theory

of metastasis progression (1) has led to innovative questions such as those addressed by Albregues *et al.* and Pommier *et al.*

Pommier *et al.* investigated how DCCs evade the immune system. They found that mice immunized against pancreatic cancer cells produced an immune response that killed most of the DCCs when they were subsequently engrafted. But some solitary DCCs persisted, which suggests that CD8<sup>+</sup> T cells did not recognize them. Careful analysis confirmed that DCCs are quiescent and lack expression of the epithelial cell marker E-cadherin (epithelial cadherin), but do not undergo an obvious epithelial-mesenchymal transition, which has been linked to increased

cancer cell adaptability to changing microenvironments and initiation of metastasis. Importantly, they found that dormant DCCs activate components of the unfolded protein response (UPR), a proteotoxicity survival pathway (5). The UPR was previously found to promote the survival of dormant cancer cells im-

planted in immunodeficient mice [patient-derived xenograft (PDX)], and it has been linked to how cancer cells respond both to internal (metabolic) stress and external stress (adaptation to foreign extracellular matrix composition) (6, 7). Pommier *et al.* found that the UPR caused posttranscriptional downregulation of major histocompatibility complex (MHC) class I molecules, which present antigens to CD8<sup>+</sup> T cells. Thus, the UPR also controls how dormant DCCs are interacting with the immune microenvironment. This decrease in MHC class I expression was associated with activated UPR and rendered dormant DCCs undetectable by CD8<sup>+</sup> T cells; restoration of UPR activity caused DCCs to reexpress E-cadherin and MHC class I, reversing the immune-evasive phenotype so that CD8<sup>+</sup> T cells could detect them in vivo.

Different cancer types exhibit predictable metastatic distribution, but the timing of relapse is variable and appears to be random. This has prompted the suggestion that both DCC-intrinsic and microenvironmental factors influence the rate of relapse (1, 2). Albregues *et al.* investigated what reactivates dormant DCCs. A prevalent theory is

**“...disseminated cancer cells... can reside in... organs for long periods...”**

Division of Hematology and Oncology, Department of Medicine, Department of Otolaryngology, Tisch Cancer Institute, Black Family Stem Cell Institute, and Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. Email: julio.aguirre-ghiso@mssm.edu

that inflammation may activate immune or other host cells to produce signals that break dormancy (8). Albrengues *et al.* showed that bacterial-derived lipopolysaccharide (LPS), a trigger of inflammation, or cigarette smoke (which may carry LPS as a contaminant) can activate neutrophils to release their DNA content into the lung parenchyma to form neutrophil extracellular traps (NETs) that usually capture microorganisms. NETs are also associated with neutrophil degranulation and release of proteases such as neutrophil elastase (NE) and matrix metalloprotease 9 (MMP9). The NETs work as scaffolds, and with the proteases they enhance the processing of laminin-111, a basement membrane component, which consequently binds  $\alpha\beta 1$  integrin and activates focal adhesion kinase (FAK) signaling in dormant DCCs. This FAK activation induced proliferation of DCCs in several mouse models of cancer, confirming previous findings from various laboratories (1). An  $\alpha\beta 1$  integrin-blocking antibody prevented reactivation of dormant DCCs in NET-rich lungs. These findings might explain why heavy smoking, before and after treatment of primary breast tumors, increases the risk of recurrence. However, Albrengues *et al.* also found that proliferative breast cancer DCCs could induce NETs, which suggests that signals other than smoking-induced inflammation may also reawaken dormant DCCs (patients who do not smoke also relapse). Perhaps NETs that are induced by active DCCs promote their growth and/or induce an immune-suppressive microenvironment.

These two elegant studies reveal how dormant DCCs may evade adaptive immunity and can be reawakened by innate immune cells (neutrophils) responding to nontumor inflammation (see the figure). The ability of DCCs to evade adaptive immunity has previously been shown for the survival of dormant leukemic cells in a mouse model of acute my-

eogenous leukemia (AML) (9). The dormant AML cells up-regulated the immune checkpoint proteins PD-L1 (programmed cell death protein 1 ligand 1) and the CTLA-4 (cytotoxic T lymphocyte-associated antigen 4) ligand B7.1, which prevent T cell activation. It would be interesting to determine whether PD-L1 and CTLA-4 signaling are also involved in immune evasion of dormant DCCs with UPR activation because of the availability of immune checkpoint inhibitors and the development of drugs targeting the UPR (10).

What is the source of UPR activation? Is it simply a response to enhanced translational output from oncogenic signaling (11)? Or is it induced by the microenvironment? Quiescent hair follicle stem cells (HFSCs) reside in niches, specialized microenvironments that control their quiescence, and in this state, HFSCs down-regulate MHC class I and evade CD8<sup>+</sup> T cell immune recognition (12). HFSCs also engage specific pathways to repress translation initiation (a function of the UPR) for proper stem cell function during entry into and exit from quiescence (13), which suggests that attenuation of translation is not only a response but also a morphogenetic program. Niche-driven quiescence and translational attenuation are also linked to dormancy of DCCs in various models (even immune-deficient models) (1, 2), raising the possibility that MHC class I down-regulation in DCCs may recapitulate an adult stem cell program that maintains quiescence.

The findings of Albrengues *et al.* suggest that antibodies that block the binding of  $\alpha\beta 1$  integrin to laminin-111 (rather than blocking all  $\alpha\beta 1$  integrin binding events) could prevent reactivation of DCCs. Detection of the laminin-111 epitope, exposed by proteolytic processing, by the antibody Ab28 in patients who have undergone surgery and have minimal residual disease may be used as an early marker for relapse, allowing for

prompt intervention. Furthermore, the proteolytic degradation of thrombospondin-1 (TSP-1), a protein linked to dormancy and metastasis suppression (14), also has a role in NET-induced reawakening of DCCs. Albrengues *et al.* found that laminin-111 processing could not induce DCC activation unless TSP-1 was degraded. Therefore, TSP-1 activity may need to be overcome for inflammation or other signals to reawaken dormant DCCs.

Together, the innovative studies by Pommier *et al.* and Albrengues *et al.* open exciting avenues to explore opportunities for monitoring and targeting dormant DCCs during minimal residual disease stages, to explore how DCC dormancy is affected by inflammation, and to identify possible targets in dormant DCCs that may impair the success of stage IV cancer treatments. This work also offers an opportunity to study how the mechanisms found in dormant DCCs are related to organism development and tissue homeostasis. Moreover, how oncogenic driver alterations affect DCCs and how the function of such driver mutations is influenced by dormancy mechanisms can be investigated. This could help to better define safer therapies that specifically target dormant or reactivating DCCs to prevent metastasis and relapse. ■

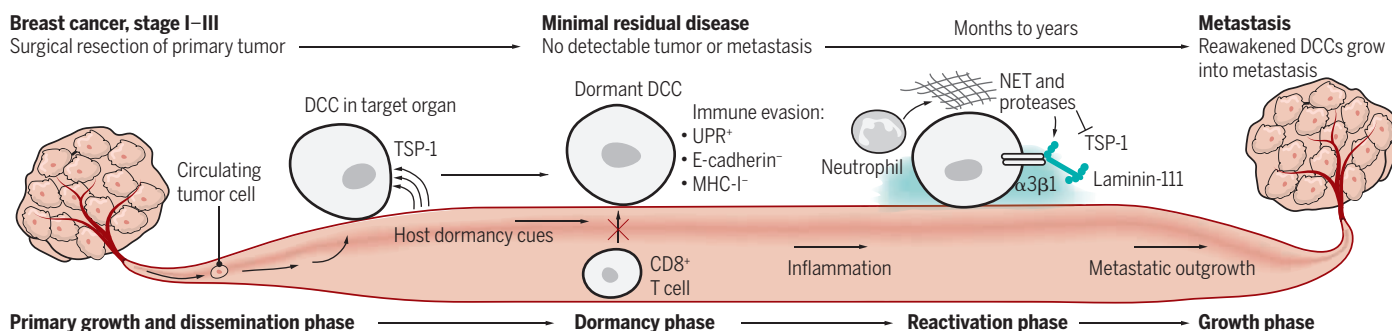
## REFERENCES

1. M. S. Sosa *et al.*, *Nat. Rev. Cancer* **14**, 611 (2014).
2. F. G. Giaccotti, *Cell* **155**, 750 (2013).
3. A. Pommier *et al.*, *Science* **360**, eaao4908 (2018).
4. J. Albrengues *et al.*, *Science* **361**, eaao4227 (2018).
5. C. Hetz, F. R. Papa, *Mol. Cell* **69**, 169 (2018).
6. D. Schewe, J. A. Aguirre-Ghiso, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10519 (2008).
7. A. C. Ranganathan *et al.*, *Cancer Res.* **66**, 1702 (2006).
8. T. El Rayes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 16000 (2015).
9. A. Saudemont, B. Quesnel, *Blood* **104**, 2124 (2004).
10. D. Pytel *et al.*, *PLoS Genet.* **12**, e1006518 (2016).
11. L. S. Hart *et al.*, *J. Clin. Invest.* **122**, 4621 (2012).
12. J. Agudo *et al.*, *Immunity* **48**, 271 (2018).
13. S. Blanco *et al.*, *Nature* **534**, 335 (2016).
14. C. M. Ghajar *et al.*, *Nat. Cell. Biol.* **15**, 807 (2013).

10.1126/science.aav0191

## Dormant DCC immune evasion and activation to form metastasis

This simplified model shows that microenvironmental cues, such as TSP-1, induce and maintain dormancy. Dormant DCCs down-regulate E-cadherin expression and up-regulate UPR activity, which results in MHC-I down-regulation and immune evasion from CD8<sup>+</sup> T cells. Eventually, unpredictable changes in the host cause inflammation and other tissue-wide changes. Inflammation can recruit neutrophils to produce NETs, and the associated proteases cause proteolytic processing of laminin-111 and TSP-1 degradation. This causes  $\alpha\beta 1$  integrin-mediated activation of dormant DCCs that proliferate to form metastasis.





## ULTRAFAST OPTICS

# Electro-optic combs rise above the noise

Electro-optic modulation of light can have a precision equivalent to one optical-field cycle

By Victor Torres-Company

**L**asers typically emit light at one frequency, but some are engineered to emit light at a set of evenly spaced optical frequencies that maintain a stable phase relation, as if a myriad of lasers were radiating in unison. Most of these laser frequency combs are based on mode-locked lasers that deliver ultrashort light pulses in a repetitive manner. This approach has found use in attosecond physics, optical clocks, molecular spectroscopy, and the generation of ultrastable microwaves (1). Before the advent of the mode-locked frequency

the bandwidth of mode-locked lasers to be expanded to one octave (that is, the optical frequency on the blue side of the spectrum is twice the frequency on the red side), as well as the stabilization of the location of the optical frequencies with unprecedented accuracy and precision (2, 3). Nonetheless, such mode-locked laser implementations can suffer from some disadvantages, such as limited tuning of the spacing between lines in the comb.

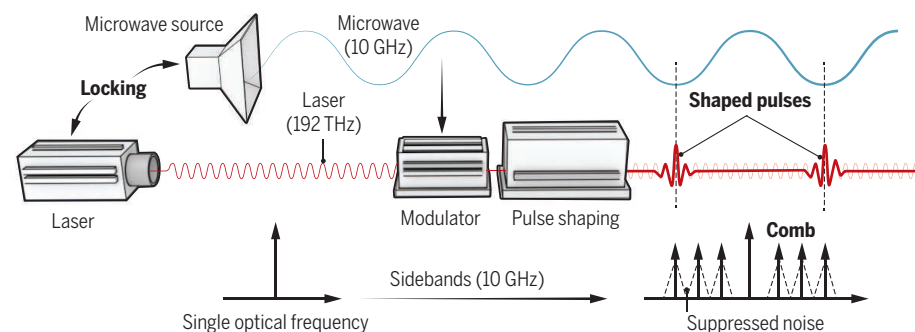
Electro-optic modulation offers potential advantages for frequency comb generation, one being its simplicity (see the figure). The input laser defines the central frequency of the comb, and a modulator driven by a mi-

Carlson *et al.* took advantage of key advances in ultrafast nanophotonics and microwave engineering to suppress this detrimental noise. They achieve a true optical frequency comb spectrum derived from an electro-optic source by forcing the microwave signal to follow the laser, using optical and microwave filtering to suppress the noise, and broadening the spectrum to form few-cycle optical pulses. Their study unravels the fundamental origin of the noise formation in electro-optic frequency combs and paves the way for technological innovations that could benefit from advances in integrated modulators (12) and resonant nonlinear optics (13). Their frequency comb extends beyond the octave while keeping the width of the frequency lines below 1 Hz. This light source delivers optical pulses shorter than 20 fs in duration at a 10-GHz repetition rate with a timing precision better than one period of the oscillating electric field.

Such a remarkable accomplishment brings the performance of electro-optic combs to a level that allows meaningful comparisons with mode-locked frequency combs. The light source by Carlson *et al.* is not meant to replace the latter, but rather to enable applications that benefit from the combination of ultrafast repetition rate and timing precision. Possibilities for future applications are only limited by our imagination, and range from calibration of astronomical spectrographs for the search of Earth-like exoplanets (14) to impulsive Raman-scattering microscopy (15). ■

## Microwave and lasers generate frequency combs

A microwave signal is added to a single-frequency laser with an electro-optic modulator. Carlson *et al.* used nanophotonic and microwave engineering to suppress noise.



comb (2, 3), the electro-optic modulation method (4) was a popular technique for measuring frequency differences between lasers (5). On page 1358 of this issue, Carlson *et al.* (6) report that electro-optic modulation of light can be used to create an ultrafast laser source for the generation of frequency combs. The “comeback” of this method is part of a larger set of efforts exploring alternative techniques for creating frequency combs to enable new applications and facilitate their operation outside the laboratory (7).

Mode-locked lasers consist of a cavity that contains an optical amplifier and an optical element that favors the formation of pulses (8). Dramatic progress in ultrafast optics and specialty optical fibers in the 1990s allowed

crowave signal sets the distance between consecutive frequencies. Thus, a comb of optical frequencies can be produced by piggybacking optoelectronic components formerly developed for telecommunications applications. Consecutive lines in an electro-optic comb can be easily set apart by 10 GHz or more (9), an aspect that is notably challenging to achieve with mode-locked lasers (10). A critical issue with electro-optic combs has been finding ways of expanding their bandwidth while keeping the comb-like structure. The spectral purity of the optical sidebands degrades the further away they are from the central line because of the accumulation of microwave noise. The degradation for the lines near the center is barely noticeable, but when the spectrum spans an octave, the comb structure is almost washed out by noise. This problem cannot be solved even by using the purest microwave signals available (11).

## REFERENCES AND NOTES

1. N. R. Newbury, *Nat. Photonics* **5**, 186 (2011).
2. D. J. Jones *et al.*, *Science* **288**, 635 (2000).
3. R. Holzwarth *et al.*, *Phys. Rev. Lett.* **85**, 2264 (2000).
4. T. Kobayashi *et al.*, *IEEE J. Quantum Electron.* **24**, 382 (1988).
5. M. Kourogi *et al.*, *IEEE J. Quantum Electron.* **29**, 2693 (1993).
6. D. R. Carlson *et al.*, *Science* **361**, 1358 (2018).
7. T. J. Kippenberg *et al.*, *Science* **332**, 555 (2011).
8. A. M. Weiner, *Ultrafast Opt.* (Wiley, 2009).
9. A. J. Metcalf *et al.*, *IEEE J. Sel. Top. Quantum Electron.* **19**, 3500306 (2013).
10. A. Bartels *et al.*, *Science* **326**, 681 (2009).
11. K. Beha *et al.*, *Optica* **4**, 406 (2017).
12. M. Zhang *et al.*, paper presented at the Conference on Lasers and Electro-Optics (CLEO), San Jose, CA, 2018 10.1364/CLEO\_QELS.2018.FW3E.4.
13. E. Obrzud *et al.*, *Nat. Photon.* **11**, 600 (2017).
14. X. Yi *et al.*, *Nat. Commun.* **7**, 10436 (2016).
15. T. Ideguchi *et al.*, *Nature* **502**, 355 (2013).

## ACKNOWLEDGMENTS

Funding support from the Swedish Research Council and the European Research Council (GA 771410) is gratefully acknowledged.

10.1126/science.aau7507

Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-41296 Gothenburg, Sweden.  
Email: torresv@chalmers.se

## CANCER RESEARCH

# Cancer prevention: Molecular and epidemiologic consensus

Research in many fields emphasizes the value of prevention

By **Mingyang Song**<sup>1,2</sup>, **Bert Vogelstein**<sup>3</sup>,  
**Edward L. Giovannucci**<sup>1,4</sup>, **Walter C.**  
**Willett**<sup>1,4</sup>, **Cristian Tomasetti**<sup>5,6</sup>

**P**revention of any disease can occur at two levels: (i) avoiding or reducing risk factors coupled with increases in protective factors (primary prevention, which is preferable when it can be practiced) and (ii) detection and intervention early in the course of disease evolution (secondary prevention). But despite substantial epidemiologic data showing that a large proportion of cancers and cancer deaths are preventable, decreases in cancer mortality rates in developed countries have lagged far behind decreases in mortality rates from heart disease (1), another major disease amenable to prevention (for example, 18 versus 68% decrease, respectively, between 1969 and 2013 in the United States) (2). We believe that one main factor explaining the relatively modest reduction in mortality is the limited support for cancer prevention research, which receives only 2 to 9% of global cancer research funding (3). As a United Nations (UN) High-Level Meeting begins this week to review efforts to combat noncommunicable diseases, a key question is how to prioritize resources to realize the potential of cancer prevention.

## LATE VERSUS EARLY CANCERS

The great majority of cancer research is focused on curing late cancers that have already spread throughout the body by the time they are detected. The reasons for this heavily skewed focus are manifold. First, societies, in general, are reactive rather than proactive. Second, the final stages of treatment research (and regulatory approval) can be simpler to perform than prevention research (requiring just hundreds of patients

versus tens of thousands of patients). Third, it is much more dramatic to effectively treat a patient with advanced disease than to prevent disease. Thus many patients who are effectively treated donate large sums to cancer centers; there are few thanks given for preventing cancers. Fourth, there are few financial incentives for industry to support primary prevention measures based on avoidance of risk factors. And finally, the financial incentives to develop new therapeutics are far more lucrative than those for new diagnostic tests for early detection and prevention.

Recent research has illuminated why it is so difficult to cure advanced cancers. Even the best new targeted therapies can generally only induce transient responses because hundreds to thousands of cells that are resistant to such therapies already exist within any advanced cancer (4). These preexisting resistant cells will eventually emerge, causing relapse. On the other hand, recent research has solidified the view that many cancers are entirely preventable through changes in environment or lifestyle. A cancer that is prevented is “cured,” not simply driven into a transient remission. Moreover, primary prevention eliminates the considerable morbidity associated with surgery and adjuvant therapy.

However, not all cancers are preventable by changes in environment or lifestyle. Recent research (5, 6) has shown that mutations due to random mistakes during normal DNA replication (R) play a major role in cancer etiology, along with environment and lifestyle (collectively denoted E) and heredity (H). R can also explain extreme variation in

cancer incidence across different tissues. R, calculated from the lifetime number of cell divisions in a tissue, is correlated with the lifetime risk of cancer in that tissue, indicating a role for R, independent of E, worldwide.

This EHR model (see the figure) highlights the connection between epidemiologic and molecular perspectives and informs cancer research and prevention strategies in two ways. First, explicit quantification of different types of mutations in cancer reinforces the importance of prevention: Cancers can still be preventable as long as one of the mutations driving toward cancer is caused by E. Second, the model highlights the heterogeneity of cancer arising in different tissues. For cancers in which most of their driver mutations are caused by E—such as lung cancers, melanomas, and cervical cancers—about 85 to 100% of incident cases could be eliminated through smoking cessation, avoidance of ultraviolet radiation exposures, and vaccination against human papillomavirus, respectively. For other cancer types that have a large proportion of R mutations—such as those of the pancreas, breast, and prostate—less

than half of incident cases can be attributed to known environmental risk factors (7, 8). Variations in cancer rates among countries and studies of migrants indicate the existence of additional environmental factors that contribute to more of these cancers than are currently known. However, mutations that occur naturally regardless of the external environment undoubtedly play a

role (6). Fortunately, many of these cancers will still be amenable to secondary prevention, with morbidity and mortality from the disease minimized.

## CONNECTIONS AND MECHANISMS

More research is needed to strengthen the connection between epidemiology and molecular biology. It will be important to identify mechanisms through which diet, exercise, and other lifestyle factors that are unambiguously associated with cancer lead to the disease (9). Can such factors be associated with genetic or epigenetic signatures that are often found in cancers, tying sequencing and epidemiology together? At present, most molecular signatures identified in genome-wide sequencing studies of cancer cannot be attributed to any environmental factor.

A specific example might put these challenges in perspective. Obesity is appreciated to be a major risk factor for cancer. Although the mechanisms through which obesity increases cancer risk remain to be fully under-

<sup>1</sup>Departments of Epidemiology and Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA. <sup>2</sup>Clinical and Translational Epidemiology Unit and Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. <sup>3</sup>Ludwig Center and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, Baltimore, MD, USA. <sup>4</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, MA, USA. <sup>5</sup>Division of Biostatistics and Bioinformatics, Department of Oncology, Sidney Kimmel Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA. <sup>6</sup>Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. Email: bertvog@gmail.com; wwillet@hsph.harvard.edu



stood, several explanations can be offered (10). First, obesity may increase the number of mutations by increasing the concentration of endogenous mutagenic substances, such as reactive oxygen species. Second, obesity could increase the number of mutations in a tissue by increasing cell divisions in that tissue through hormones, growth factors, metabolites, or other chemicals produced by adipocytes and other tissues (for example, estrogens, insulin, adipokines, and prostaglandins). Because cell division is intrinsically mutagenic, each time a cell divides, three to five new mutations occur. Moreover, increased proliferation itself, in a tumor that already harbors mutations, could determine whether a tumor becomes clinically evident

generally not considered in the preventability estimate, partly because of uncertainty in the risk-benefit profile. For example, there is convincing evidence that regular use of aspirin can prevent colorectal cancer (12) and that tamoxifen and raloxifene lower the risk of estrogen receptor–positive breast cancer (13).

### CHANGING RESEARCH PRIORITIES

The World Health Organization Global Action Plan for the Prevention and Control of Noncommunicable Diseases calls for a 25% reduction in cancer mortality rates by 2025, and the UN Sustainable Development Goals program calls for a 33% reduction by 2030. Achieving these goals requires not only more research but also strong government com-

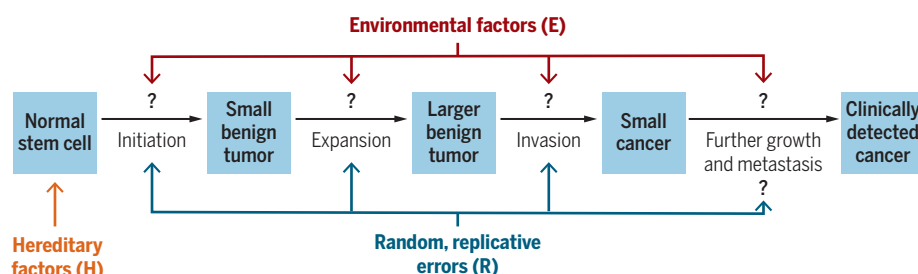
cancers. Interventions based on improvements in minimally invasive surgical procedures, accomplished through endoscopic means, are equally important initiatives.

Of course, primary and secondary prevention strategies should be balanced and prioritized on the basis of country-specific cancer demographics, risk factor profile, and resource availability. As in developed countries, primary prevention is preferable in developing countries, but secondary prevention with simple, cost-effective tests—such as Pap smears followed by surgical excisions for prevention of cervical cancer—is feasible in the developing world. Moreover, in developing countries, the costs of expensive new drugs that treat advanced disease are not feasible (15). Optimally, prevention efforts will also confer co-benefits for other chronic diseases with shared risk factors (for example, type 2 diabetes, depression, and dementia).

We believe that both epidemiologic and molecular research support a congruent public policy recommendation for consideration by the UN: A much greater emphasis on cancer prevention research is needed to avoid, detect, and treat early cancers and premalignant lesions. With such further research, we envision that cancer death rates could be reduced by 70% around the world, even without the development of any new therapies (7). Such a reduction, similar to that for heart disease over the past six decades, will only come about if research priorities are changed. ■

## Connecting epidemiologic and genetic perspectives on cancer

Solid tumors of adults evolve through a stepwise accumulation of three or more genetic alterations, but these are not the only determinants of cancer incidence. Hereditary factors can predispose individuals to cancer, generally via an initiating mutation to all stem cells. Random, replicative errors every time a cell divides can contribute to cancers at the indicated phases of tumorigenesis. Environmental factors can play a role during all phases.



and thus enumerated in epidemiologic studies. Third, obesity may promote cancer by influencing gene expression through epigenetic mechanisms that can be reversed by weight loss. Finally, adipose tissue is a source of mesenchymal stem cells, which may increase the pool of cells at risk through their recruitment to support tumor growth and progression. These potential pathways warrant further molecular and epidemiologic research. By incorporating the role played by R, the EHR model highlights the potential for new chemopreventive strategies, including those that reduce epigenetic changes or reduce errors made during cellular DNA replication (6).

Regardless of the precise mechanisms, substantial epidemiologic data show that at least 35% of all cancer deaths in the world could be avoided by modifying known risk factors (8), that is, through primary prevention. Preventability estimates will hopefully increase with further research, because additional preventable causes are likely to be discovered. For example, early-life exposures may play a critical role in cancer initiation, but epidemiologic studies have mostly focused on midlife risk factors (11). Moreover, the potential of chemopreventive strategies is

mitments to sustainable investments in primary cancer prevention strategies through legislative, regulatory, financial, and educational approaches. As highlighted by the U.S. National Cancer Institute's Cancer Moonshot Blue Ribbon Panel, there is an urgent need for more behavioral and policy research on overcoming barriers to adoption of effective evidence-based interventions.

Concomitantly, advances in genetics and other scientific areas promise new vistas in secondary prevention. Many patients have early, potentially curable cancers, or advanced precursor lesions, that can be detected through mutations or other tumor-specific markers present in their circulation or other bodily fluids and removed early to prevent further progression (14). Large-scale, prospective studies to determine the ability of such biomarkers to save lives will be required to gain confidence in such tests. But one can imagine a day when healthy individuals are routinely tested for these biomarkers to detect early cancers, along with lipid concentrations to detect early cardiac disease, at periodic visits to their physicians. Similar advances in physics (for example, in imaging) could provide orthogonal tests to detect early

### REFERENCES AND NOTES

1. GBD Mortality and Causes of Death Collaborators, *Lancet* **388**, 1459 (2016).
2. J. Ma, E. M. Ward, R. L. Siegel, A. Jemal, *JAMA* **314**, 1731 (2015).
3. S. Eckhouse, G. Lewison, R. Sullivan, *Mol. Oncol.* **2**, 20 (2008).
4. L. A. Diaz Jr. et al., *Nature* **486**, 537 (2012).
5. C. Tomasetti, B. Vogelstein, *Science* **347**, 78 (2015).
6. C. Tomasetti, L. Li, B. Vogelstein, *Science* **355**, 1330 (2017).
7. F. Islami et al., *CA Cancer J. Clin.* **68**, 31 (2017).
8. G. Danaei et al., *Lancet* **366**, 1784 (2005).
9. E. A. Golemis et al., *Genes Dev.* **32**, 868 (2018).
10. N. Murphy, M. Jenab, M. J. Gunter, *Nat. Rev. Gastroenterol. Hepatol.* **10**, 1038/s41575-018-0038-1 (2018).
11. M. A. Clarke, C. E. Joshi, *Epidemiol. Rev.* **39**, 11 (2017).
12. K. Bibbins-Domingo, U.S. Preventative Services Task Force, *Ann. Intern. Med.* **164**, 836 (2016).
13. J. Cuzick et al., *Lancet* **381**, 1827 (2013).
14. G. Siravegna, S. Marsoni, S. Siena, A. Bardelli, *Nat. Rev. Clin. Oncol.* **14**, 531 (2017).
15. H. Kantarjian, D. Steensma, J. Rius Sanjuan, A. Elshaug, D. Light, *J. Oncol. Pract.* **10**, e208 (2014).

### ACKNOWLEDGMENTS

**Funding:** This work is supported by the National Institutes of Health (K99 CA215314 and R00 CA215314 to M.S., UM1 CA167552 and UM1 CA176726 to W.C.W., and P30 CA006973 to C.T.), the American Cancer Society (MRSG-17-220-01 – NEC to M.S.), the Lustgarten Foundation for Cancer Research (B.V.), the Virginia and D. K. Ludwig Fund for Cancer Research (B.V.), the Marcus Foundation (B.V. and C.T.), the John Templeton Foundation (C.T.), and the Maryland Cigarette Restitution Fund (C.T.). **Competing interests:** PapGene, Inc., has licensed technologies from Johns Hopkins University, and B.V. and C.T. receive equity or royalties from these licenses. The terms of these arrangements are being managed by Johns Hopkins University in accordance with its conflict of interest policies. B.V. is a founder of Personal Genome Diagnostics, Inc., and PapGene, Inc., and is on the Scientific Advisory Board of Sysmex-Inostics.

10.1126/science.aau3830

## MEDICINE

# Stories of the sanguine

A meandering history of blood tackles transfusions, taboos, and trauma

By Susan E. Lederer

**B**lood is a vital fluid. Essential to life, it is often taken for granted, especially when it stays where it belongs, safely contained in our bodies. Rose George, freelance journalist and author of books that render legible the overlooked (1) and the distasteful (2), here turns her attention to blood in its myriad biological, cultural, historical, and international dimensions.

In nine chapters, corresponding to the roughly nine pints of blood each adult person is circulating at any given time through a complex network of arteries and veins, she explores blood donation and transfusion; leeches; the organization of transfusion services in England; HIV and other bloodborne diseases; plasma as fluid and as a major industry; menstruation, menstruation technologies, and menstrual taboos; trauma surgery and the physiology of hemorrhage; and the future of blood, including the ongoing search for artificial blood.

No armchair explorer, George makes the journey to the heart of blood a personal one. The book begins as she watches her own blood—female, type A positive—moving from her arm into the plastic bag in a British donor center. Here, she introduces

the workings of the circulatory system, the discovery of the blood types, blood products, and the merits of paying for blood versus voluntary blood donation.

In a chapter entitled “That most singular and valuable reptile,” George explores the history of medicinal leeches and the rediscovery of their value in modern plastic and reconstructive surgery. These crea-

tures—which she clarifies are not, as George Horn indicated in his 1798 treatise, reptiles—can remove excess blood without damage to the microvasculature. She visits Biopharm, a large leech breeding facility in Wales that ships leeches to surgeons around the world, where she permits Carl Peters-Bond to apply one to her hand. “I have blanked out what it felt like,” she writes, noting that a photograph taken during the experience shows her “face screwed up in a classic expression of disgust.”

She marvels at how well modern patients tolerate the application of leeches in all their living, slimy glory.

Perhaps not surprisingly given her previous book on human biotrash, George devotes two chapters to menstrual blood and its management over time and space. Traveling to rural Nepal and India, George meets women whose menstrual periods render them exiled to the outdoors, where they are vulnerable to the elements, snakes, and sexual assault.

Outraged by the treatment of these women—the girls who cannot attend

In India, patients who require a transfusion must provide blood from friends or family, writes George.

school and the conspiracy of silence about the biology of the menstrual process—George reminds readers that menstrual stigma is not confined to Asia. She considers at length the apparently enduring belief in the United States that bears and other creatures of the wild are attracted by menstrual odor. She quotes, for example, from a 1981 pamphlet from the National Park System (“Grizzly! Grizzly! Grizzly! Grizzly!”), which warns visitors to stay clean and avoid perfumes and that “women should stay out of bear country during their menstrual period.”

Later, George calculates the amount of menstrual blood she has discharged over 35 years and bemoans the lack of biomedical data on menstruation and women’s reproductive health. Here she notes that until 1992, the U.S. National Institutes of Health had no programs for vaginal research.

In the 21st century, surgery for traumatic injuries has exploded. Where once amputation was the only hope, surgeons have been able to transform injury into livable disability and productivity. George chronicles the prodigious blood requirements for such surgeries and the outcomes as she relates an experience in which she witnessed incredible, if ultimately futile, efforts to save a young female bicyclist after a traffic calamity in London.

As George notes, one of the holy grails of 20th-century medicine was the search for a blood substitute—one made ready at a moment’s notice and not burdened by the extraordinary immunological complexity of blood itself. Someday, perhaps, the necessity of blood-type congruency for transfusion may seem as quaint as the practice of blood-letting, but we surely will never cease to wonder at the exquisite differences in blood itself.

Each chapter of *Nine Pints* reflects George’s experience, personal investment, and broad attention to the historical, political, social, biological, and moral aspects of blood. If the organizational thread is not always easy to follow, the book nevertheless overflows with telling examples—some fantastic, some uncanny, all informative about the sanguinary fluid. ■

## REFERENCES

1. R. George, *Ninety Percent of Everything: Inside Shipping, the Invisible Industry That Puts Clothes on Your Back, Gas in Your Car, and Food on Your Plate* (Metropolitan Books, 2013).
2. R. George, *The Big Necessity: The Unmentionable World of Human Waste and Why It Matters* (Metropolitan Books, 2009).

10.1126/science.aau8695



**Nine Pints**  
A Journey Through the  
Money, Medicine, and  
Mysteries of Blood

Rose George  
Metropolitan Books,  
2018. 368 pp.

The reviewer is at the Department of Medical History and Bioethics, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706, USA. Email: susan.lederer@wisc.edu



## HUMAN ORIGINS

# Me and you and everyone we know

From the Big Bang to our early ancestors, a philosopher probes human origins and identity

By Michael A. Goldman

**W**ith campaigns of disinformation rampant in America, arming citizens with scientific knowledge is crucial. In the slim volume *You: A Natural History*, Wright State University philosopher William B. Irvine lays out some of the fundamental science behind human history, human ancestry, life's history and origin, and the origin of our Universe in simple and engaging prose. He set out, he says, to present a "humanistic" science, not just to tell readers how "they and their world came to exist, but to take this science personally."

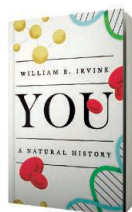
Although the human saga dominates the book, its scope isn't limited to just human biology or even biology in general. Irvine sweeps from the Big Bang to the genesis of elements and from the coalescence of planets to the origin of life. He is especially enamored of trees—family, descent, and evolutionary—an excitement he apparently shares with many others, as evidenced by our enthusiasm for direct-to-consumer genetic ancestry testing and the proliferation of television programs dedicated to uncovering family histories. His discussions, however, lack descriptions of the experimental basis for the underlying systematic and evolutionary concepts.

In just over a single page, Irvine dispenses with the idea of the ownership of genes and possibly but not explicitly their patentability—a topic that has been widely and voluminously debated. He concludes that although an individual's atoms might fairly be considered as belonging to him or her, the ownership of genes is more complex, as we each own not a gene but "a copy" of a gene shared by many others. He compares it with owning Robert Frost's poem "The Road Not Taken." Like the sequence of a gene, copies of the poem can be made without damage to the original. I'm not at all convinced that the question can be settled so simply. Irvine goes on to argue that not only do we not own our genes, our

"selfish genes"—the term coined by Richard Dawkins to describe genes that ensure their own propagation even if they do not benefit the organism—might actually be taking advantage of us.

Throughout the book, Irvine addresses issues of religion and law, and I was especially pleased to see that these discussions, although not themselves scientific, are grounded in thorough research. In his discussion of the fate of the individual after death, Irvine draws a distinction between the body and the mind. He considers the Christian view of resurrection, in which the mind lives on in a "resurrected" body, and the Eastern view of reincarnation. "People who enjoy life don't want it to end," he writes. But Irvine thinks our desires will always outpace what we have and prefers the Buddhist concept of reaching nirvana, in which the self no longer matters.

Turning to more practical considerations, Irvine mentions the potential for cryopreservation and the idea of uploading one's mind to a computer. Here, he misses a magnificent chance to highlight the 200th anniversary of the publication of Mary Shelley's *Frankenstein*. "[O]ne thing is certain," he writes, "before you experience a life after death, if there is one, you will experience a life before death." Urging readers to focus on the latter, he adds, "[Y]ou would be wise to spend your time and energy living this life as well as you can."



You

*A Natural History*  
William B. Irvine  
Oxford University  
Press, 2018. 248 pp.

Irvine mainly presents science as facts and stories rather than as a process and a history of discovery. He does, however, recognize that our understanding of how things work and how things came to be changes with time. For instance, he notes, whereas we once thought that modern humans and Neanderthals were separate species that could not, by definition, interbreed, we later learned from genomics data that Neanderthal genes remain in humans today and that we should more likely be considered a subspecies.

*You: A Natural History* came to my attention because it purported to be a natural history of humans. But it rose to the top of my list because it was written by a philosopher. I was hoping for an extensive discussion of the philosophy of science. On this front, the book did not meet my expectations. In his introduction, for example, Irvine writes that he will "address some of the questions that arise regarding the meaning of life." The bulk of this discussion, disappointingly, is limited to only a few pages near the end of the book. It is not terribly engaging or profound. For a deeper discussion of human origins, history, and ancestry, I would recommend Yuval Noah Harari's *Sapiens* or David Reich's *Who We Are and How We Got Here*.

Nevertheless, Irvine's conversational tone and brief chapters make this book accessible to a broad audience. And if I were to list the topics I'd like to see covered in a general science education, I would include many of the ones that appear in *You: A Natural History*. ■

10.1126/science.aau7306



Fitness groups that focus on support over competition may backfire.

## PODCAST

**How Behavior Spreads**  
**The Science of Complex Contagions**  
Damon Centola  
Princeton University Press, 2018.  
308 pp.

Conventional wisdom holds that behaviors spread more effectively through weak social ties rather than strong ones. But according to Damon Centola, conventional wisdom is wrong. This week on the *Science* podcast, Centola describes how societal behaviors are really propagated, revealing how networks can be exploited to effect social change.

10.1126/science.aav1974

The reviewer is at the Department of Biology, San Francisco State University, San Francisco, CA 94132, USA. Email: goldman@sfsu.edu



## LETTERS

Edited by **Jennifer Sills**

### Editorial expression of concern

On 10 September 2010, *Science* published the Report “Human SIRT6 promotes DNA end resection through CtIP deacetylation” by A. Kaidi, B. T. Weinert, C. Choudhary, and S. P. Jackson (1). On 19 August 2018, the Research Governance and Integrity Officer of the corresponding author’s institution, Cambridge University, together with the corresponding author, notified *Science* that the institution was launching a preliminary investigation under the Misconduct in Research policy. They have since notified us that the university has concluded that there is a prima facie case that requires formal investigation. *Science* is publishing this Editorial Expression of Concern to alert our readers while we await the outcome of the investigation.

**Jeremy Berg**  
Editor-in-Chief

#### REFERENCE

1. A. Kaidi, *Science* **329**, 1348 (2010).

10.1126/science.aav4528

### Lack of science support fails Brazil

On 2 September, the world watched in horror as Brazil’s National Museum, housing a vast collection of more than 20 million biodiversity and cultural artifacts, was engulfed in flames (“In a ‘foretold

tragedy,’ fire consumes Brazil museum,” H. Escobar, *In Depth*, 7 September, p. 960). The museum’s extensive natural history collections, painstakingly accumulated over more than two centuries, documented the change in species identity and distributions over time, recorded the earliest South American inhabitants’ culture and native languages, and archived the origin and historical progress of a nation. The magnitude of this loss is staggering—not just for Brazil but for the world. Scientific advancement is based on building blocks from the past, and without those components, scientists are left without points of reference. Museum collections are the foundation on which we recognize cultural and scientific novelty as we strive to understand and better the human condition, to advance our grasp of how nature’s pieces came into being and fit together, and even to predict the ecological and evolutionary future of the planet’s biodiversity (1).

Funding for the museum decreased substantially during the past 5 years (2), and calls for renewed investment in renovations, security, and protection have been ignored for decades (3). In this sense, the National Museum is an apt metaphor for the current state of science in Brazil: Leaders at all levels have failed to provide even the most basic and crucial infrastructure for preserving genuinely priceless collections and cultural resources. This fire at the National Museum follows the loss of 80,000 specimens in the fire that destroyed the collections at Instituto Butantan in 2010 (4) and the complete loss of the Museum of the Portuguese Language in São Paulo in 2015 (5), both also attributed to poor investment in infrastructure. Recent years have seen

Increased science funding might have prevented the devastating fire at Brazil’s National Museum.

large declines in the budget for basic scientific research and student training (6). The loss of unique and irreplaceable collections, owing to lackluster federal investment in science, adds salt to the growing wound.

There is some hope among the ashes. Many of the biological collections, including vertebrates, most of the marine invertebrates, and plants, as well as rare books, were spared because they were in different buildings. Fortunately, no human lives were lost. Curators and museum staff are working around the clock to reorganize and house displaced colleagues and maintain teaching and mentoring in their ongoing graduate programs. The Brazilian government has pledged funds and infrastructure to support rebuilding (7, 8), providing an opportunity for Brazil to correct mistakes of the past.

This tragic loss resonates beyond Brazil. Museum collections are timeless national treasures that represent our histories, cultures, and scientific achievements. Every institution and government should reflect and take heed at this sad moment. We must invest in and safeguard our museums and collections for the benefit of science and society worldwide.

**Kelly R. Zamudio,<sup>1\*</sup> Alexander Kellner,<sup>2</sup> Cristiana Serejo,<sup>2</sup> Marcelo Ribeiro de Britto,<sup>2</sup> Clovis B. Castro,<sup>2</sup> Paulo A. Backup,<sup>2</sup> Débora O. Pires,<sup>2</sup> Marcia Couri,<sup>2</sup> Adriano Brilhante Kury,<sup>2</sup> Irene Azevedo Cardoso,<sup>2</sup> Marcela L. Monné,<sup>2</sup> José Pombal Jr.,<sup>2</sup> Cátia Mello Patiu,<sup>2</sup> Vinicius Padula,<sup>2</sup> Alexandre Dias Pimenta,<sup>2</sup> Carlos Renato**



Rezende Ventura,<sup>2</sup> Eduardo Hajdu,<sup>2</sup>  
Joana Zanol,<sup>2</sup> Emilio M. Bruna,<sup>3</sup> John  
Fitzpatrick,<sup>4</sup> Luiz A. Rocha<sup>5</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology and  
Museum of Vertebrates, Cornell University, Ithaca,  
NY 14853, USA. <sup>2</sup>Museu Nacional, Universidade  
Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

<sup>3</sup>Center for Latin American Studies, and Department  
of Wildlife Ecology and Conservation, University of  
Florida, Gainesville, FL 32611, USA. <sup>4</sup>Laboratory of  
Ornithology, Cornell University, Ithaca, NY 14850,  
USA. <sup>5</sup>California Academy of Sciences, San Francisco,  
CA 94118, USA.

\*Corresponding author.

Email: kelly.zamudio@cornell.edu

## REFERENCES

1. L. A. Rocha et al., *Science* **344**, 814 (2014).
2. M. Brandão, "UFRJ e governo federal divergem sobre verba enviada ao Museu Nacional" Agência Brasil (2018); <http://agenciabrasil.ebc.com.br/politica/noticia/2018-09/ufrj-e-governo-federal-divergem-sobre-verba-enviada-ao-museu-nacional>.
3. M. A. Canônico, "Bicentenário Museu Nacional, o mais antigo do país, tem problemas de manutenção," Folha de São Paulo (2018); <https://www1.folha.uol.com.br/ilustrada/2018/05/bicentenario-museu-nacional-o-mais-antigo-do-pais-tem-problemas-de-manutencao.shtml> [in Portuguese].
4. J. Kemech, *Lancet* **375**, 2061 (2010).
5. V. Sreeharsha, "Brazil: Fire destroys São Paulo museum," *The New York Times* (2015); [www.nytimes.com/2015/12/22/world/americas/brazil-fire-destroys-sao-paulo-museum.html](http://www.nytimes.com/2015/12/22/world/americas/brazil-fire-destroys-sao-paulo-museum.html).
6. A. Petherick, *Nature* **548**, 249 (2017).
7. M. Brandão, "MPs criam agência para administrar museus e fundos patrimoniais," Agência Brasil (2018); <http://agenciabrasil.ebc.com.br/politica/noticia/2018-09/mps-criam-agencia-para-administrar-museus-e-fundos-patrimoniais> [in Portuguese].
8. M. Tokarnia, "Ministro da Educação diz que oferecerá apoio ao Museu Nacional," Agência Brasil (2018); <http://agenciabrasil.ebc.com.br/geral/noticia/2018-09/ministro-da-educacao-diz-que-oferecera-apoio-ao-museu-nacional> [in Portuguese].

10.1126/science.aav3296

## Relating addiction and psychiatric disorders

In their Research Article "Analysis of shared heritability in common disorders of the brain" (22 June, p. 1313), The Brainstorm Consortium shows that psychiatric disorders share common risk variants, whereas neurological disorders appear more distinct from one another and from the psychiatric disorders. The analyses include 10 psychiatric disorders. Future studies on common risk variants should include addictive disorders as well.

Addictive disorders are among the most common, debilitating, and stigmatized disorders (1). Just like other psychiatric disorders (1), they are defined in international classification systems for psychiatric disorders like DSM-5 (2) and largely based on behavioral symptoms. The high comorbidity between addictive disorders and other psychiatric disorders (3) suggests pathophysiological overlap. Combinations of these disorders negatively affect prognosis of both disorders, stressing the relevance for exploring shared mechanisms (4–6).

The magnitude of genetic correlations between addictive disorders (including substance use) and other psychiatric disorders is comparable with the genetic correlations between psychiatric disorders reported by The Brainstorm Consortium. For example, the genetic correlation of .52 ( $P = 2.18 \times 10^{-20}$ ) between major depressive disorder (MDD) and attention deficit hyperactivity disorder (ADHD), described in the Research Article (table S7A), is similar in magnitude to the correlation of .57 ( $P = 3.1 \times 10^{-4}$ ) between MDD and alcohol dependence (7). Similarly, significant genetic correlations of .53, .38, and .37 for ADHD with alcohol dependence (7), smoking quantity, and smoking initiation (table S7B), respectively, are comparable to correlations of .52, .26 and .22 between ADHD and MDD, bipolar disorder, and schizophrenia, respectively (table S7A). Genetic correlations for MDD with smoking initiation ( $r_g = .33$ ,  $P = 3.1 \times 10^{-11}$ ) (table S7B) and cannabis use ( $r_g = .21$ ,  $P = 3.0 \times 10^{-4}$ ) (8) are also significant. The Brainstorm Consortium did include smoking as a risk factor for other conditions. However, smoking is an addictive disorder on its own, with high prevalence, morbidity, and mortality (9, 10). It therefore makes sense to look at correlations between smoking as a risk factor and tobacco use disorder as an addition, as well as correlations between tobacco use disorder and other psychiatric disorders.

These correlations suggest that addictive disorders share a considerable portion of their common variant genetic risk with other psychiatric disorders. This association may result from the same genes influencing multiple phenotypes (horizontal pleiotropy) or from causal relationships (vertical pleiotropy) (11). To better understand these associations, substance use and addictive disorders should be included in future studies of the biological pathways underlying psychiatric disorders.

Jacqueline Vink<sup>1</sup> and Arnt Schellekens<sup>2,3</sup>

<sup>1</sup>Behavioural Science Institute, Radboud University, 6525 HR Nijmegen, Gelderland, Netherlands.

<sup>2</sup>Department of Psychiatry, Donders Institute for Brain Cognition and Behavior, Centre for Medical Neuroscience, Radboud University Medical Center, 6525 GA Nijmegen, Netherlands. <sup>3</sup>Nijmegen Institute for Scientist Practitioners in Addiction, 6525 HR Nijmegen, Netherlands.

\*Corresponding author.

Email: arnt.schellekens@radboudumc.nl

## REFERENCES

1. P. Corrigan, G. Schomerus, D. Smeelson, *Br. J. Psychiat.* **210**, 180 (2018).
2. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Publishing, Arlington, VA, ed. 5, 2013).
3. R. E. Roberts, C. R. Roberts, Y. Xing, *Drug Dependence* **88**, S4 (2007).
4. A. F. A. Schellekens, C. A. J. de Jong, J. K. Buitelaar, R. J. Verkes, *Eur. Psychiat.* **30**, 128 (2015).
5. C. L. Crunelle et al., *Eur. Addiction Res.* **24**, 43 (2018).
6. B. Hitsman et al., *Addiction* **108**, 294 (2013).
7. R. K. Walters et al., *bioRxiv* 10.1101/257311 (2018).



Differing classifications for Australia's dingo affect how the species is managed.

8. J. A. Pasman *et al.*, *Nat. Neurosci.* **21**, 1161 (2018).
9. R. West, *Psychol. Health* **32**, 1018 (2017).
10. T. Effertz, K. Mann, *Eur. Neuropsychopharmacol.* **23**, 742 (2013).
11. A. B. Paaby, M. V. Rockman, *Trends Genet. TIG* **29**, 66 (2013).

10.1126/science.aav3928

## Species definitions shape policy

The names we assign to organisms, and why, have important ramifications for our understanding of Earth's diversity and, more practically, how it is managed. For example, wolves, coyotes, domestic dogs, and other canids are often considered distinct (1), but their members can, and frequently do, interbreed (2). Differing concepts of species—which might take into account morphology, ecology, behavior, genetics, or evolutionary history (3)—could describe canids as very few or many species, depending on which concepts are used and how strictly they are applied. Which definition scientists adopt can have political and ecological consequences.

The dingo (*Canis dingo*) has traditionally been considered native in Australia, given evidence of its presence before the year 1400 (4) and indications that it has lived in Australia for at least 5000 years (5). This designation meant that Western Australia had to have a management strategy in place for the dingo, along with other native fauna. However, a recent paper (6) argues that dingoes are in fact *C. familiaris* because they don't satisfy zoological nomenclature protocols nor sufficiently differ genetically or morphologically from other canids, including domestic dogs. The Western Australian government cited

this work in justifying its recent decision to declare the dingo a non-native species under the state's Biodiversity Conservation Act (BCA) (7). The new order removes the government requirement to manage the species. As a result, dingoes can now be killed anywhere in the state without a BCA license. A potential increase in lethal control of dingoes could have dire consequences for Australia's ecosystems. The dingo is Australia's largest terrestrial top predator [adults typically weigh 15 to 20 kg (8)], it fulfils a crucial ecological role, and it has strong cultural significance for Australia's Indigenous people (8).

Taxonomy serves a critical purpose for cataloguing and conserving biodiversity, but different interpretations and applications of species concepts can affect management decisions. Policy-makers may use the interpretations that justify their preferred values, such as prioritizing livestock more than biodiversity protection. It is therefore imperative that scientists carefully engage in the policy decision-making process. Scientists must work with policy-makers to convey the multiple dimensions and values that can affect species delineation and make clear the potential consequences of applying such classifications.

**Euan G. Ritchie,<sup>1\*</sup> Bradley P. Smith,<sup>2</sup> Lily M. van Eeden,<sup>3</sup> Dale G. Nimmo<sup>4</sup>**

<sup>1</sup>School of Life and Environmental Sciences, Centre for Integrative Ecology, Deakin University, Burwood, VIC 3125, Australia. <sup>2</sup>School of Health, Medical, and Applied Sciences, Appleton Institute, Central Queensland University, Wayville, SA 5034, Australia. <sup>3</sup>Desert Ecology Research Group, School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia. <sup>4</sup>School of Environmental Science, Institute for Land, Water, and Society, Charles Sturt University, Albury, NSW 2640, Australia.

\*Corresponding author.  
Email: e.ritchie@deakin.edu.au

## REFERENCES

1. J. Clutton-Brock *et al.*, *Bull. Br. Mus. Nat. Hist. Zool.* **29**, 117 (1976).
2. Z. Fan *et al.*, *Gen. Res.* **26**, 163 (2016).
3. F. E. Zachos, *Mammal Rev.* 10.1111/mam.12121 (2018).
4. Department of the Environment and Energy, Australian Government, "Environment Protection and Biodiversity Conservation Act 1999 (EPBC Act)" (1999); [www.environment.gov.au/epbc](http://www.environment.gov.au/epbc).
5. K. M. Cairns, A. N. Wilton, *Genetica* **144**, 553 (2016).
6. S. M. Jackson *et al.*, *Zootaxa* **4317**, 201 (2017).
7. M. Bamford, "Dingoes to remain classified as non-native wild dogs under reform to Western Australian law," ABC News (2018); [www.abc.net.au/news/2018-08-28/dingoes-will-no-longer-be-native-animals-in-western-australia/10172448](http://www.abc.net.au/news/2018-08-28/dingoes-will-no-longer-be-native-animals-in-western-australia/10172448).
8. M. Letnic *et al.*, *Biol. Rev.* **87**, 390 (2012).

10.1126/science.aav3437

## TECHNICAL COMMENT ABSTRACTS

**Comment on "DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification"**

**Chip Stewart, Ignaty Leshchiner, Julian Hess, Gad Getz**

Chen *et al.* (Reports, 17 February 2017, p. 752) highlight an important problem of sequencing artifacts caused by DNA damage at the time of sample processing. However, their manuscript contains several errors that led the authors to incorrect conclusions. Moreover, the same sequencing artifacts were previously described and mitigated in The Cancer Genome Atlas and other published sequencing projects.

Full text: [dx.doi.org/10.1126/science.aas9824](https://doi.org/10.1126/science.aas9824)

**Response to Comment on "DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification"**

**Lixin Chen, Pingfang Liu, Thomas C. Evans Jr., Laurence M. Ettwiller**

Following the Comment of Stewart *et al.*, we repeated our analysis on sequencing runs from The Cancer Genome Atlas (TCGA) using their suggested parameters. We found signs of oxidative damage in all sequence contexts and irrespective of the sequencing date, reaffirming that DNA damage affects mutation-calling pipelines in their ability to accurately identify somatic variations.

Full text: [dx.doi.org/10.1126/science.aat0958](https://doi.org/10.1126/science.aat0958)

**Comment on "The plateau of human mortality: Demography of longevity pioneers"**

**H. Beltrán-Sánchez, S. N. Austad, C. E. Finch**

Barbi *et al.* (Reports, 29 June 2018, p. 1459) reported that human mortality rate reached a "plateau" after the age of 105, suggesting there may be no limit to human longevity. We show, using their data, that potential life spans cannot increase much beyond the current 122 years unless future biomedical advances alter the intrinsic rate of human aging.

Full text: [dx.doi.org/10.1126/science.aav1200](https://doi.org/10.1126/science.aav1200)



## TECHNICAL COMMENT

## MUTATION DETECTION

# Comment on “DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification”

Chip Stewart<sup>1\*</sup>, Ignaty Leshchiner<sup>1\*</sup>, Julian Hess<sup>1</sup>, Gad Getz<sup>1,2,3,†</sup>

Chen *et al.* (Reports, 17 February 2017, p. 752) highlight an important problem of sequencing artifacts caused by DNA damage at the time of sample processing. However, their manuscript contains several errors that led the authors to incorrect conclusions. Moreover, the same sequencing artifacts were previously described and mitigated in The Cancer Genome Atlas and other published sequencing projects.

The sequencing artifacts discussed in Chen *et al.* (1) have been described in publications from The Cancer Genome Atlas (TCGA) and other cancer genome projects (2–4). Accordingly, effective mitigation strategies have long been implemented, including software

pipelines and improved library preparation methods, as described in Costello *et al.* (2). Thus, findings described in TCGA publications (5–9) were not affected by oxidative damage, as suggested by Chen *et al.* The authors do raise general awareness of sequencing artifacts, which include ma-

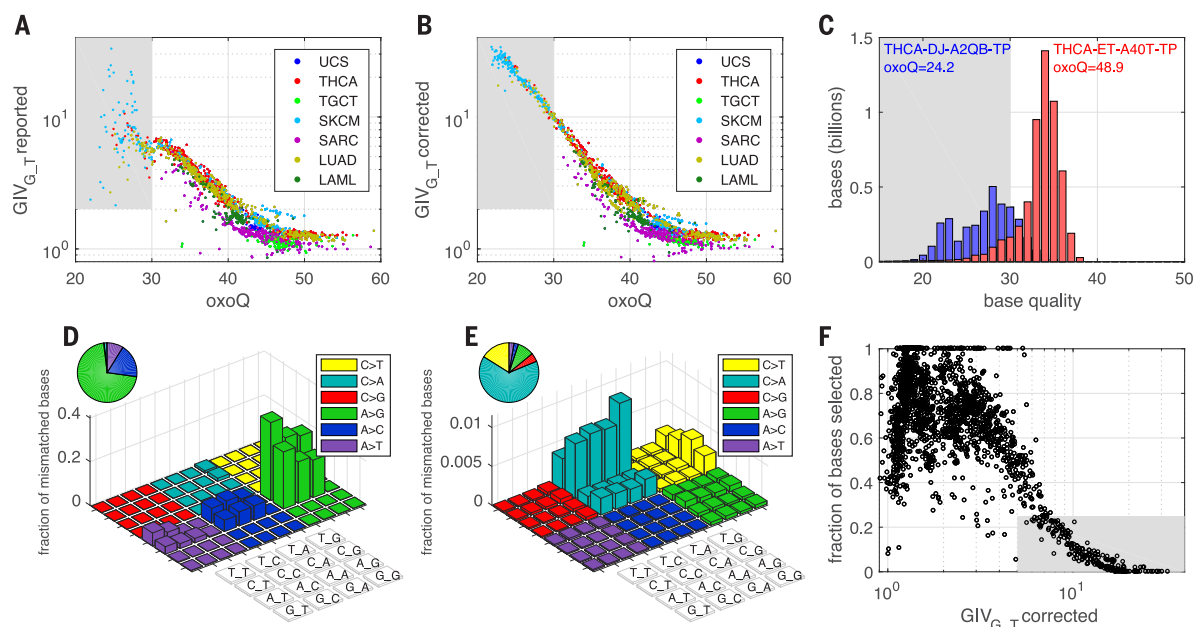
chine errors (2), DNA oxidation (2–4), and DNA cross-linking (in clinically used formalin-fixed tissues) (10) among others; however, their paper contains several errors that led to incorrect conclusions. The errors affect the following:

i) Estimation of oxidative damage levels: Although their reported oxidative damage (8-oxo-G) metric,  $GIV_{G-T}$ , is essentially equivalent to an earlier reported metric,  $oxoQ$  (2–4, 11), their software implementation is limited by using reads aligned only to the forward strand and biased by filtering out low-quality bases.  $oxoQ$  was designed to be interpreted as a (Phred-like) base quality score. As such,  $oxoQ$  can be compared to typical levels of sequencing errors, which for most Illumina sequencing protocols is roughly at the level of  $Q \sim 30$  (i.e., an error rate of 1 per 1000 bases). We initially noticed a poor agreement between the  $oxoQ$  and  $GIV_{G-T}$  metrics for damaged samples ( $oxoQ < 30$ ; Fig. 1A), whereas scores were consistent for samples with low levels of DNA damage. Examining the methods of Chen *et al.*, we found that the authors' code with default parameters applied a high base quality threshold ( $Q > 30$ ) and removed sites where total coverage exceeds

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02124, USA. <sup>2</sup>Massachusetts General Hospital Cancer Center and Department of Pathology, Boston, MA 02114, USA. <sup>3</sup>Harvard Medical School, Boston, MA 02115, USA.

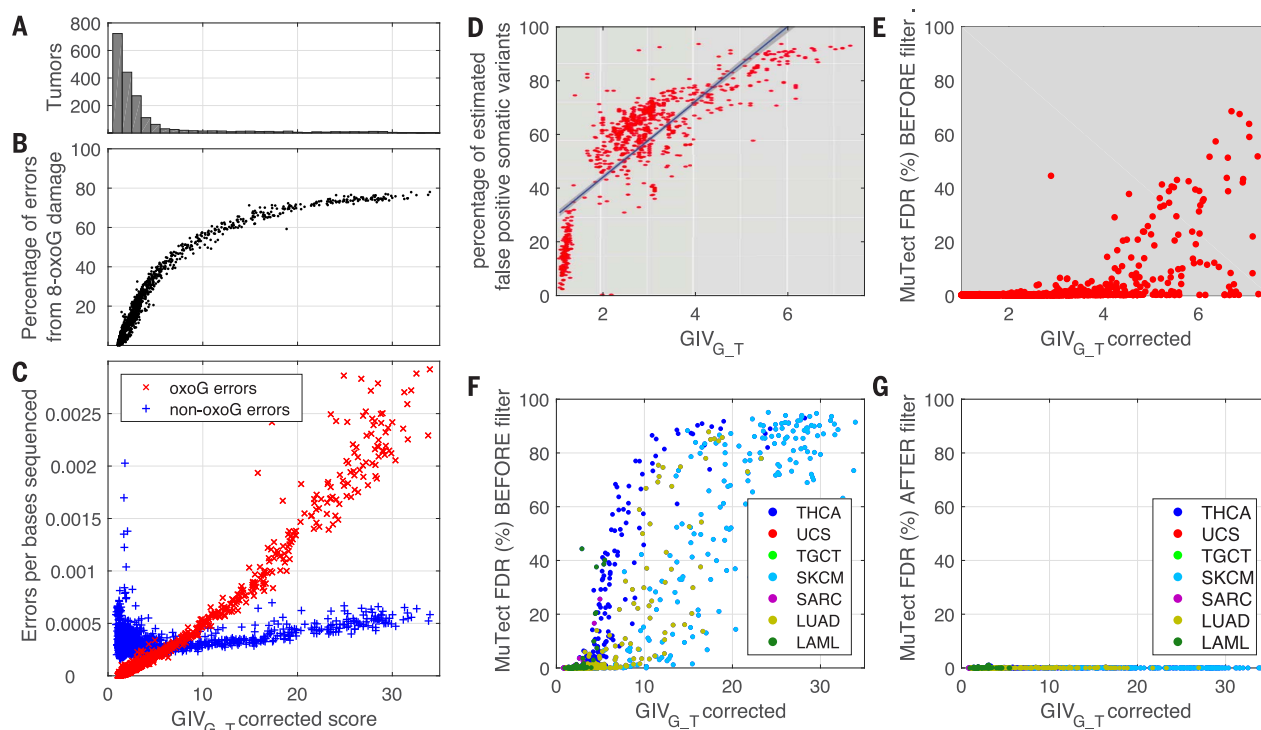
\*These authors contributed equally to this work.

†Corresponding author. Email: gadgetz@broadinstitute.org



**Fig. 1. Comparison of  $GIV_{G-T}$  to  $oxoQ$  score.** (A) Picard  $oxoQ$  score versus  $GIV_{G-T}$  for ~1900 TCGA tumor exomes. Lack of agreement between  $oxoQ$  and  $GIV_{G-T}$  scores is apparent in samples with high DNA damage ( $oxoQ < 30$ ) highlighted in the gray box.  $GIV_{G-T}$  is calculated with Chen *et al.*'s code (estimate\_damage.pl) with default parameters;  $oxoQ$  was calculated with Picard CollectSequencingArtifactMetrics (11) output. (B)  $oxoQ$  versus  $GIV_{G-T}$  corrected (using base quality  $Q > 20$  and coverage depth  $\geq 20\times$ , comparable to Picard defaults) showing excellent agreement. Color code in (A) and (B) indicates data from seven different TCGA tumor types. (C) Histogram of individual base qualities in a typical TCGA 8-oxo-G-damaged sample ( $oxoQ = 24.2$ ) compared to a low-damage sample ( $oxoQ = 48.9$ ). Most of the bases in

the damaged sample have  $Q < 30$  and hence are ignored by default  $GIV_{G-T}$ . This explains both the lack of agreement between the scores seen in (A) and the authors missing the sequence context associated with 8-oxo-G damage [(D) versus (E)]; that is, the  $Q > 30$  filtering removed bases with the type of damage that is being quantified. (D and E) “Lego” plot showing the distribution of errors in different 5' and 3' sequence contexts (plotted in reverse complement by convention), with Chen *et al.*'s (D) and our corrected (E) filtering criteria, showing distortion from the characteristic 8-oxo-G context pattern arising from data selection criteria. (F) The fraction of bases that remain after applying Chen *et al.*'s filters, showing that most bases are removed in highly damaged samples.



**Fig. 2. Comparison of 8-oxo-G-related error rates and other error modes.** (A to C) Samples with corrected  $GIV_{G,T}$  damage levels below 5 (equivalent to oxoQ  $\sim 35$ ) have fewer 8-oxo-G-related base errors (red) relative to other sequencing error modes (blue). (A) Count of tumors with a specific corrected  $GIV_{G,T}$  score (the majority of TCGA samples have very low damage levels). (B) Percentage of additional base mismatches caused by 8-oxo-G in relation to  $GIV_{G,T}$  score. (C) 8-oxo-G-related error rates

(red) and other error modes (blue). Other modes dominate until corrected  $GIV_{G,T} \sim 5$ . (D) Copy of figure 4E from Chen *et al.* (1). (E) Our estimated FDR versus corrected  $GIV_{G,T}$  for  $\sim 1900$  TCGA tumors from MuTect (14) before 8-oxo-G filtering ( $GIV_{G,T} < 7.5$ ). (F and G) Estimated FDR versus corrected  $GIV_{G,T}$  for  $\sim 1900$  TCGA tumors from MuTect before (F) and after (G) 8-oxo-G filtering [as described in (2)]. Results demonstrate no inflation in number of mutation calls after applying the 8-oxo-G software filter.

100 $\times$ , which discarded the bulk of the data for the most damaged samples (Fig. 1, C to F). Adjusting the filtering criteria to include bases with  $Q > 20$  and to allow sites with depth above 100 $\times$  restored concordance between  $GIV_{G,T}$  and oxoQ scores (Fig. 1B).

ii) Claim of no context specificity of 8-oxo-G damage: The authors' conclusion that they "did not observe nucleotide context specificity" of 8-oxo-G damage (1) is a consequence of their filtering out most of the supporting data, as described above. Sequence context specificity is a key feature of 8-oxo-G damage (12), can affect mutational signature analysis, and can be visualized using "Lego" plots (Fig. 1, D and E). These plots show the error rate of each type of base substitution in its three-base sequence context (2). The damaged bases reported by Chen *et al.*'s script (Fig. 1D) have a severely distorted error profile inconsistent with the previously reported (12) 8-oxo-G damage pattern (Fig. 1E, C > A errors in the Lego plot with a clear peak at the sequence context CCG > CAG), whereas the corrected  $GIV_{G,T}$  script [and the previous implementation in (2)] results in a distribution of errors (Fig. 1E) consistent with 8-oxo-G damage. Thus, 8-oxo-G damage does in fact have a sequence context, which the authors apparently missed as a result of their filtering strategy.

iii) Interpretation of 8-oxo-G damage levels: The authors state that "73% of the TCGA se-

quencing runs showed extensive damage, with a  $GIV_{G,T} > 2$ ." A  $GIV_{G,T}$  metric of 2 indicates that the  $G > T$  error rate in the 8-oxo-G mode is twice the error rate of the non-8-oxo-G mode (background rate for  $G > T$  errors). Even if the 8-oxo-G artifacts are twice the context-specific background level (i.e.,  $GIV_{G,T} = 2$ ), this corresponds to only a 5 to 10% increase in the overall base-level error rate (summed over all sequence contexts; Fig. 2, A to C), which is less than the intersample variability of error rates at a fixed oxoQ. A 5% increase in the base-level error rate results in a minor, if any, increase in false-positive mutation calls (Fig. 2, E and F), because calling algorithms are designed to handle typical levels of sequencing error. Only at  $GIV_{G,T} \geq 5$  (equivalent to oxoQ  $\leq 35$ ) do the additional errors from 8-oxo-G become comparable to the sum of all other errors and have an adverse impact on variant calling. The vast majority of samples in TCGA exhibit only minor 8-oxo-G damage that has minimal impact on mutation calling. Consequently, the claim that 73% of TCGA sequencing runs have extensive damage is misleading.

iv) Estimation of the false positive rate (FPR) of mutation calling: The authors define an FPR metric, which suggests that mutation calls for many samples in public databases contain >50% falsely detected somatic variants due to sequencing artifacts (Fig. 2D). However, their metric does

not reflect the actual somatic mutations used for analyses in (3–10) and other TCGA publications [e.g., a recent reanalysis of TCGA data (13)], but instead represents candidate variants supported by as few as two nonreference reads [not considered as somatic variants by most mutation callers (14)]. Moreover, the mathematical definition of their FPR metric is neither a FPR nor a false discovery rate (FDR), which is highlighted by the fact that it can range between  $-1$  and  $1$  (and not between  $0$  and  $1$ ). Therefore, Chen *et al.* incorrectly concluded that the FPR exceeds  $0.5$  in samples that have a low level of DNA damage with nearly no damage-induced false-positive mutation calls (Fig. 2, D to G).

v) Finally, Chen *et al.* overlook publications describing 8-oxo-G damage (2–10) that TCGA and other projects have mitigated with laboratory protocols and software filtering strategies when sequencing library regeneration was impractical (Fig. 2, F and G). Moreover, they claim that "recent submissions to TCGA (November to December 2015) displayed similar G-to-T imbalances" but mistakenly interpreted the time that the data repository was last updated as the time of sequencing data generation (repository updates occurred periodically for unrelated issues). The actual dates of generating the sequencing data should be obtained from the sequencing metadata. In fact, most TCGA samples sequenced after



October 2012 had low 8-oxo-G damage due to improved library preparation [as described in (2)].

In conclusion, we disagree with the assessment of Chen *et al.* regarding the quality of published cancer genome projects. Raw sequencing data inherently contain errors; therefore, to avoid misinterpreting the data, it is important that researchers use established procedures and carefully curated datasets in downstream analyses.

# REFERENCES AND NOTES

1. L. Chen, P. Liu, T. C. Evans Jr., L. M. Ettwiller, *Science* **355**, 752–756 (2017).
2. M. Costello *et al.*, *Nucleic Acids Res.* **41**, e67 (2013).
3. T. J. Pugh *et al.*, *Nat. Genet.* **45**, 279–284 (2013).
4. B. D. Crompton *et al.*, *Cancer Discov.* **4**, 1326–1341 (2014).
5. Cancer Genome Atlas Research Network, *Cell* **159**, 676–690 (2014).
6. Cancer Genome Atlas Research Network, *N. Engl. J. Med.* **372**, 2481–2498 (2015).
7. Cancer Genome Atlas Research Network, *Cell* **161**, 1681–1696 (2015).
8. Cancer Genome Atlas Research Network, *Nature* **541**, 169–175 (2017).
9. Cancer Genome Atlas Research Network, *Cancer Cell* **32**, 185–203.e13 (2017).
10. M. Giannakis *et al.*, *Cell Rep.* **17**, 1206 (2016).
11. Picard; <https://broadinstitute.github.io/picard>.
12. Y. Margolin, V. Shafirovich, N. E. Geacintov, M. S. DeMott, P. C. Dedon, *J. Biol. Chem.* **283**, 35569–35578 (2008).
13. K. Ellrott *et al.*, *Cell Syst.* **6**, 271–281.e7 (2018).
14. K. Cibulskis *et al.*, *Nat. Biotechnol.* **31**, 213–219 (2013).

# ACKNOWLEDGMENTS

We are grateful for useful comments from E. S. Lander, M. Costello, N. J. Lennon, and L. Lichtenstein as well as editing help from M. Miller. **Funding:** Partially supported by NIH grant 1U24CA210999. **Author contributions:** C.S., I.L., J.H., and G.G. conceived the analysis, carried out the analysis, and wrote the text. **Competing interests:** The authors have no competing interests. **Data availability:** Raw data are available via TCGA controlled access, <https://tcga-data.nci.nih.gov/docs/publications/tcga/accesstiers.html>; code and tables are available at <https://github.com/broadinstitute/damage>.

11 January 2018; accepted 29 August 2018  
10.1126/science.aas9824

## TECHNICAL RESPONSE

## MUTATION DETECTION

# Response to Comment on “DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification”

Lixin Chen, Pingfang Liu, Thomas C. Evans Jr.\*<sup>†</sup>, Laurence M. Ettwiller\*

Following the Comment of Stewart *et al.*, we repeated our analysis on sequencing runs from The Cancer Genome Atlas (TCGA) using their suggested parameters. We found signs of oxidative damage in all sequence contexts and irrespective of the sequencing date, reaffirming that DNA damage affects mutation-calling pipelines in their ability to accurately identify somatic variations.

Previously, we devised a metric termed the global imbalance value (GIV) to evaluate how mutagenic damage affects sequencing accuracy (1). We showed that mutagenic damage is pervasive in public sequencing datasets and confounds the identification of so-

matic variants with low to moderate (1 to 5%) allelic frequency. Following our publication, the principle of global imbalance was incorporated by the International Cancer Genome Consortium (ICGC) as one of five measures used to construct a quality rating for each cancer genome (2).

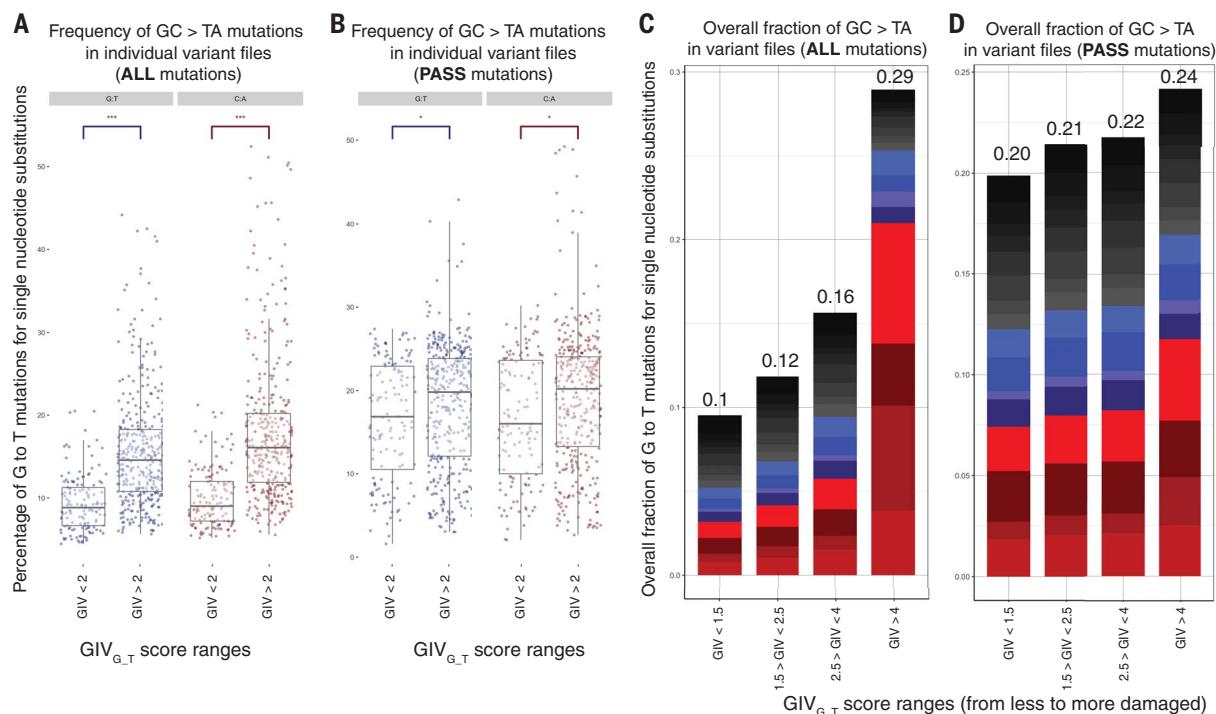
Stewart *et al.* (3) reaffirm the presence of sequencing errors derived from damage in raw sequencing reads (1, 4–6), but they question the relevance of these errors in downstream analysis. Many of their points can be distilled down to whether (i) sequencing datasets with a GIV score greater than 2 can affect the output of mutation callers, and (ii) measures have been taken since 2012 to mitigate these errors. Stewart *et al.* also claim (iii) that the default cutoff chosen in our paper led to incorrect conclusions related to the context specificity of oxidative damage-induced errors. We address these points in order.

i) DNA damage affects mutation-calling pipelines: Responding to whether damage affects downstream analysis, we examined whether damage-induced errors can be detected in mutation calls generated using standard approaches. For this, we downloaded VCF files generated in 2016 using MuTect (7) from the TCGA data portal (<https://portal.gdc.cancer.gov/>) and examined whether the fraction of G:C > T:A point mutations can be correlated with the estimated level of damage in the corresponding sequencing runs.

We observed an overall increase in the fraction of G:C > T:A mutations from a median of

New England Biolabs Inc., Ipswich, MA 01938, USA.

\*Corresponding author. Email: [evanst@neb.com](mailto:evanst@neb.com) (T.C.E.); [ettwiller@neb.com](mailto:ettwiller@neb.com) (L.M.E.)



**Fig. 1. TCGA mutation calls are affected by DNA damage.** (A) Percentage of G > T and C > A point mutations relative to the total number of point mutations in variant files downloaded from the TCGA data portal. Two groups are shown: group 1, for which the sequencing of the cancer sample leads to a GIV score of <2, and group 2, for which the sequencing data have a GIV score of >2 (damaged samples). Damaged samples show a significantly elevated fraction of G-to-T mutations ( $P < 2.2 \times 10^{-16}$ , Wilcoxon test). (B) Same as (A) using only the mutations labeled as PASS. Even for PASS mutations, samples that are

damaged show an elevated fraction of G-to-T mutations [ $P = 0.03$  (G\_T),  $P = 0.012$  (C\_A), Wilcoxon test]. \* $P \leq 0.05$ , \*\*\* $P \leq 0.001$ .

(C) Overall fraction of G-to-T point mutations relative to the total number of point mutations in the ALL-MuTect variant files downloaded from TCGA for cancer samples with  $GIV_{G-T} < 1.5$ ,  $1.5 < GIV_{G-T} < 2.5$ ,  $2.5 < GIV_{G-T} < 4$ , and  $GIV_{G-T} > 4$ . Mutations are color-coded according to sequence context. (D) Same as (C) except using only the mutations labeled as PASS. GIV scores were calculated using the cutoff (Q score > 20) as in Stewart *et al.*



8.9% in cancer samples with no or low levels of damage ( $GIV_{G,T} < 2$ ) to a median of 15.3% in samples that were damaged ( $GIV_{G,T} > 2$ ) (Fig. 1A). Restricting the analysis to only the MuTect PASS mutations (7) that represent the most confident set of mutations, we found an overall increase in the fraction of G:C > T:A mutations from a median of 16.8% to a median of 20.0% in damaged samples (Fig. 1B). This increase in the fraction of G:C > T:A mutations for damaged samples is observable even for moderately damaged samples ( $1.5 < GIV_{G,T} < 2.5$ ) (Fig. 1, C and D). Collectively, this demonstrates that standard mutation callers are affected by damage-induced errors, even in analysis pipelines from 2016.

ii) Damage can be found post-2012: Stewart *et al.* assert that the effects of oxidative damage on sequencing data were known and mitigated after 2012 (3). Although we cite the study accordingly, we note that they did not consider broad mutagenic damage in key public databases, nor is

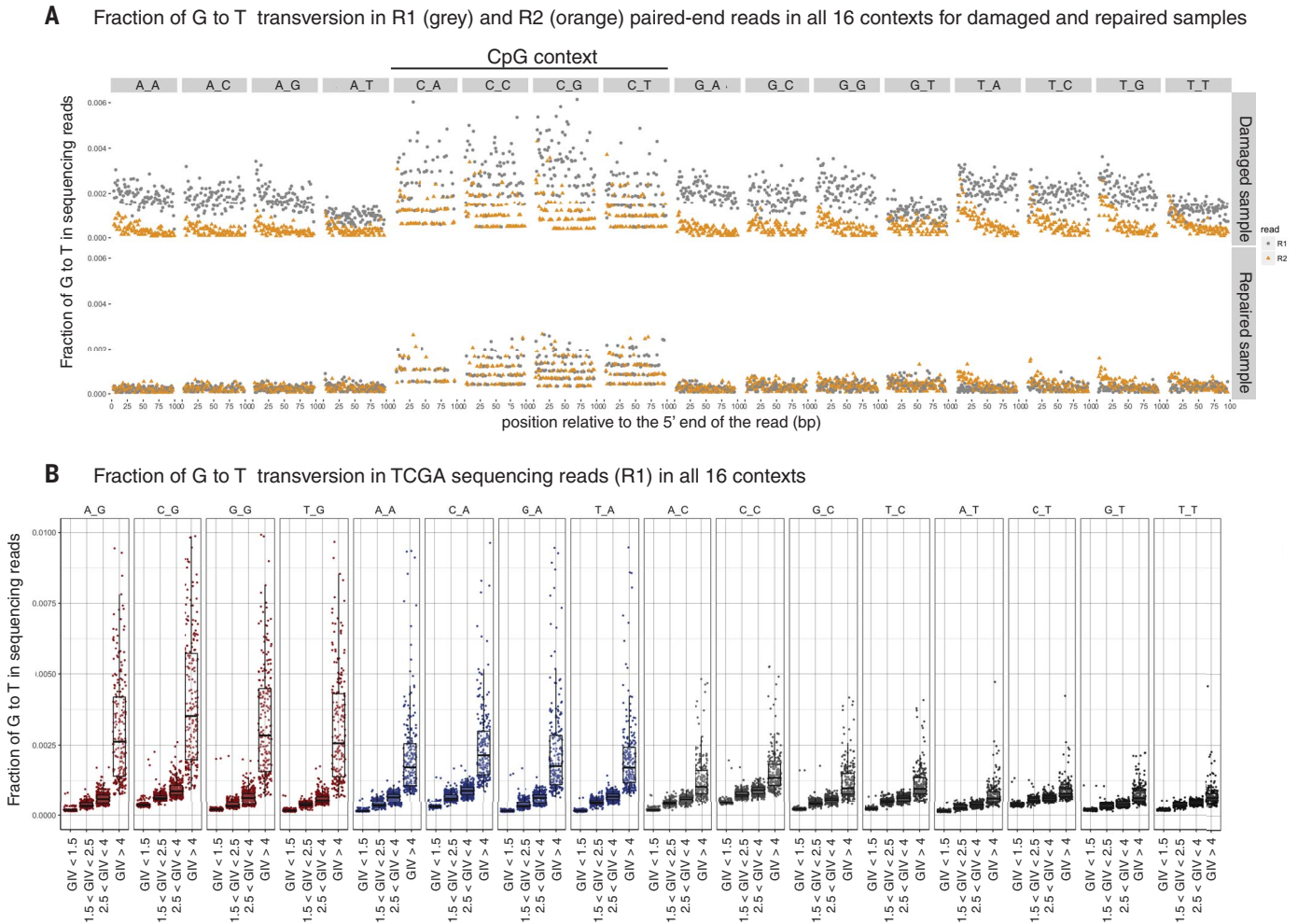
their claim of full mitigation supported by the mean GIV scores from samples of TCGA datasets from 2010 to 2015 (Table 1). Our findings indicate that artifacts consistent with DNA damage are present in these databases irrespective of the sequencing date.

iii) G-to-T transversion consistent with 8-oxo-deoxyguanosine (8-oxo-dG) damage can be found in all sequence contexts: Stewart *et al.* found that the CCG > CAG context specificity is a feature of oxidative damage of dG and is essential for mutation signature analysis (4). In contrast, we reported that a G-to-T imbalance indicative of damage is present regardless of the preceding or succeeding nucleotide. Stewart *et al.* claim that this divergence was due to a default cutoff (base quality threshold  $Q > 30$ ) (8, 9), which would eliminate most of the data in highly damaged samples, and they conclude that the oxidative damage of dG lacks context specificity. We therefore used the cutoff suggested by Stewart *et al.*, reanalyzed our

previously reported in-house sequencing runs, and again found evidence of damage-induced errors in all sequence contexts (Fig. 2A).

Extending our study to public sequencing datasets using a more quantitative measurement of damage, we observed an increase in the overall fraction of G-to-T transversions in damaged samples irrespective of sequence context (Fig. 2B). Furthermore, some samples had a greater increase in the fraction of G-to-T transversions when a purine was located 3' to the dG. Thus, the CCG > CAG context observed by Stewart *et al.* represents only a subset of errors from oxidative damage.

We value the cross-examination of our scientific work, as it has strengthened some of our earlier statements and refined others. Damage-induced sequencing errors are pervasive; we estimated that for 73% of TCGA sequencing runs analyzed in our paper, >50% of G-to-T raw read variants correspond to damage ( $GIV_{G,T} > 2$ ). Stewart *et al.* assert that we have misled readers to believe that



**Fig. 2. G > T transversions in all 16 three-base sequence contexts.** (A) Reanalysis of the sequencing dataset using the GIV scores calculated using the suggested cutoff ( $Q$  score  $> 20$ ) by Stewart *et al.* Read variant frequencies of G to T in R1 (gray) and R2 (orange) as a function of the position were plotted relative to the 5' end of the read. The upper panel shows the damaged sample [(1), figure S1, Water]; the lower panel corresponds to the same sample repaired with damage repair enzymes [(1), figure S1,

Water + repair]. A line denotes a different error pattern at CpG contexts due to the limited amount of data at these sites. (B) Quantitative measurement of the fraction of G to T transversions in sequencing reads from a subset of TCGA datasets in all 16 three-base sequence contexts (N\_N denotes NGN > NTN context). Red and blue denote 3' purine context (NGG > NTG and NGA > NTA contexts, respectively). The datasets were grouped into four categories, from no damage ( $GIV_{G,T} < 1.5$ ) to severely damaged samples ( $GIV_{G,T} > 4$ ).

**Table 1. GIV<sub>G,T</sub> scores for sequencing runs performed between 2010 and 2015.** We reanalyzed our data using the date printed on the BAM file header and calculated the mean, median, and first and third quantiles of GIV scores for each sequencing year. Although we identify an improvement in 2013, sequencing performed in 2014 and 2015 had an average GIV score comparable to the average GIV score obtained in and prior to 2012. GIV scores were calculated with the suggested cutoff (Q score > 20).

Year	First quantile	Median	Mean	Third quantile
2010	2.307	2.414	2.371	2.545
2011	2.606	3.026	4.586	5.199
2012	2.026	2.527	4.069	3.748
2013	1.219	1.288	1.678	2.136
2014	4.14	4.361	4.313	4.569
2015	2.516	2.876	2.749	3.09

this level of damage is extensive, based on the fact that damage accounts for only a fraction of the overall error rate. However, contrary to intrinsic errors, damage-induced errors such as 8-oxo-dG consistently produce the same error type (in this case, G to T) even at high Q scores. This distinction between the intrinsic error of the sequencing instrument and the error rate at high Q scores is particularly important because mutation callers rely on Q scores to filter out false-positive mutations. Thus, treating the impact of damage as if it were part of the overall sequencing error rate underestimates the effect of damage.

As part of our analysis, we required a metric to correctly estimate the percentage of false-positive

variants relative to the total number of variants. This metric, which we termed the rate of false positives, is described in the supplementary materials of Chen *et al.* and does not require knowing the false and true negatives. Its name, however, seems to have caused confusion, and in hindsight a different name could have been chosen.

Finally, the purpose of the GIV score is to provide an estimate of the fraction of raw read variants that are derived from damage. As exemplified by the ICGC, such measures can be used to flag samples of poor quality for improved downstream analysis. Our study does not call in to question previous studies that used data from TCGA, as stated by Stewart *et al.* Instead, we agree with Stewart *et al.* that DNA damage can

affect sequencing accuracy and downstream analysis. Undoubtedly, TCGA and similar public databases are valuable resources, and we hope that this dialogue increases the awareness of possible sources of error and mitigation strategies.

REFERENCES AND NOTES

1. L. Chen, P. Liu, T. C. Evans Jr., L. M. Ettwiller, *Science* **355**, 752–756 (2017).  
2. J. P. Whalley *et al.*, *BioRxiv* 140921 [Preprint]. 19 September 2017.  
3. C. Stewart, I. Leshchiner, J. Hess, G. Getz, *Science* **361**, eaas9824 (2018).  
4. M. Costello *et al.*, *Nucleic Acids Res.* **41**, e67 (2013).  
5. M. W. Schmitt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14508–14513 (2012).  
6. A. M. Newman *et al.*, *Nat. Biotechnol.* **34**, 547–555 (2016).  
7. K. Cibulskis *et al.*, *Nat. Biotechnol.* **31**, 213–219 (2013).  
8. Illumina, “Quality Scores for Next-Generation Sequencing”; [www.illumina.com/documents/products/technotes/technote\\_Q-Scores.pdf](http://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf).  
9. Current Illumina sequencing platforms have a reported minimum of 87% of bases with a Q score above 30, indicating that a cutoff of Q = 30 is not aggressive and does not significantly bias the dataset.

ACKNOWLEDGMENTS

We thank H. Runz for useful comments. **Funding:** Supported by New England Biolabs Inc. **Author contributions:** L.M.E. performed the data analysis; L.C., P.L., and T.C.E. contributed ideas and participated in writing the manuscript. **Competing interests:** The authors are listed as inventors on U.S. provisional serial number 62/376,165, submitted by New England Biolabs Inc., which covers improved sequence accuracy determination of a nucleic acid sample. **Data availability:** All data are referenced or available in the manuscript or the supplementary materials.

6 April 2018; accepted 29 August 2018  
10.1126/science.aat0958



Cite as: H. Beltrán-Sánchez *et al.*, *Science* 10.1126/science.aav1200 (2018).

# Comment on “The plateau of human mortality: Demography of longevity pioneers”

H. Beltrán-Sánchez<sup>1</sup>, S. N. Austad<sup>2</sup>, C. E. Finch<sup>3\*</sup>

<sup>1</sup>Fielding School of Public Health, University of California, Los Angeles, CA, USA. <sup>2</sup>Department of Biology, University of Alabama, Birmingham, AL, USA. <sup>3</sup>Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA, USA.

\*Corresponding author. Email: cefinch@usc.edu

Barbi *et al.* (Reports, 29 June 2018, p. 1459) reported that human mortality rate reached a “plateau” after the age of 105, suggesting there may be no limit to human longevity. We show, using their data, that potential lifespans cannot increase much beyond the current 122 years unless future biomedical advances alter the intrinsic rate of human aging.

Barbi and colleagues provide new high-quality data that show a mortality rate plateau for Italy after age 105 from birth and death records (1). Calculating from annual mortality rates of 0.475, the life expectancy at age 105 is 1.5 years. Lower plateaus are indicated for the growing survival to extreme age [figure 1A of (1)], and the authors conclude that a limit to longevity, “if any, has not been reached.”

However, our calculations suggest that the current record of 122 years cannot be exceeded by much without major advances in combatting mortality at both younger and advanced ages. Two very different models yield similar estimates of maximum survival age,  $T_{\max}$ . The simplest “radioactive decay” model assumes exponential decline of survival at a fixed mortality rate and calculates the age of the last survivor (2). Starting at the plateau level of mortality for age 105 in Barbi *et al.*, the upper part of Table 1 calculates the  $T_{\max}$  by gender for different numbers of initial survivors,  $N$ . Thus, the  $T_{\max}$  for a constant rate of annual mortality at age 105 for Italy would be at least 2 years below the record 122-year  $T_{\max}$  of Jeanne Calment, even for a cohort size of 10,000 at age 105. Notably, the impact of increasing cohort  $N$  on  $T_{\max}$  is attenuated logarithmically ( $\ln N$ ) (2). To attain a 50:50 chance of someone reaching 150 years of age given this plateau would take more than 4 trillion 105-year-olds, about 450 times the current world population of 7.5 billion (3). Further calculations based on the Gompertz model of Barbi *et al.* (lower part of Table 1) or on the basic Gompertz equation (4) also show limits to  $T_{\max}$  that differ by gender for different numbers of initial survivors. These calculations challenge conclusions that mortality plateaus allow unlimited longevity potential.

Since Calment’s death in 1997 at the age of 122, the numbers of people reaching 110 years have increased by a factor of 2 (3). Yet no subsequent lifespan has exceeded 120 years. The longest-lived Italian was Emma Morano, who

died at 117.3 years in 2017. Our calculations of  $T_{\max}$  from Gompertz exponential mortality rates closely agree with reported records, with greater potential  $T_{\max}$  for women by 7 years (4). Although Barbi *et al.* state that the Gompertz model does not hold after age 80, >95% of adult mortality in the Italian cohorts is described by the Gompertz model (5), which would concentrate the mortality plateau to a critically narrow domain of <5% of the Italian mortality regime. Moreover, the Gompertz slope has become steeper for both genders within the 20th century with the decline of baseline mortality (4). We share the caveats of Barbi *et al.* that mortality rates after age 105 have uncertainties in many populations.

In sum, regardless of whether one chooses an extrapolation assuming a mortality plateau after age 105 as calculated by Barbi *et al.* or by a continued Gompertzian increase in mortality after that age (4), the likelihood of anyone surviving much beyond the longevity record of Calment becomes remote unless unexpectedly large biomedical advances prove to ameliorate the basic biology of human aging.

## REFERENCES AND NOTES

1. E. Barbi, F. Lagona, M. Marsili, J. W. Vaupel, K. W. Wachter, The plateau of human mortality: Demography of longevity pioneers. *Science* **360**, 1459–1461 (2018). doi:10.1126/science.aat3119 Medline
2. C. E. Finch, M. C. Pike, Maximum life span predictions from the Gompertz mortality model. *J. Gerontol. A* **51A**, B183–B194 (1996). doi:10.1093/gerona/51A.3.B183 Medline
3. United Nations, World Fertility Patterns (2015); [www.un.org/en/development/desa/population/publications/pdf/fertility/world-fertility-patterns-2015.pdf](http://www.un.org/en/development/desa/population/publications/pdf/fertility/world-fertility-patterns-2015.pdf).
4. H. Beltrán-Sánchez, E. Crimmins, C. Finch, Early cohort mortality predicts the rate of aging in the cohort: A historical analysis. *J. Dev. Orig. Health Dis.* **3**, 380–386 (2012). doi:10.1017/S2040174412000281 Medline
5. N. S. Gavrilova, L. A. Gavrilov, V. N. Krut'ko, in *2017 Living to 100*, T. F. Harris, Ed. (Society of Actuaries, 2017), pp. 1–24.

# ACKNOWLEDGMENTS

Supported by grants R01 AG057434 and P30 AG050886 (S.N.A.), HD041022 (H.B.-S.), and AG051521, AG05142, and AG055367 (C.E.F.).

15 August 2018; accepted 11 September 2018  
Published online 28 September 2018  
10.1126/science.aav1200

**Table 1. Calculations of remaining lifespan at age 105 for Italian data and projections of population sizes required to reach selected age maxima.** Each model begins with the 463 males and 3373 females in the Barbi *et al.* 1904 birth cohort;  $P$  represents initial population size. In the “radioactive decay” model,  $N/P = \exp(-mt)$ , where  $m$  is the annual mortality rate; for age of the last survivor,  $N$  is set at 1, yielding  $T_{\max} = -(1/m) \ln(1/P)$ . In the Gompertz model as calculated in Barbi *et al.*,  $m(x) = a \times \exp(bx) \times \exp(\beta_1 C + \beta_2 M) \rightarrow S(x) = \exp\{a/b \times [1 - \exp(bx)] \times \exp(\beta_1 C + \beta_2 M)\}$ , where  $a$  and  $b$  are the Gompertz parameters;  $C$  is cohort birth year minus 1904;  $M = 1$  for males and 0 for females;  $\beta_1$  and  $\beta_2$  are coefficient estimates for cohort and sex, respectively; and  $S(x)$  is the survival function at age  $x$ . Then, the expression  $1/P = \exp\{a/b \times [1 - \exp(bx)] \times \exp(\beta_1 C + \beta_2 M)\}$  yields  $T_{\max} = \ln\{1 + [b \times \ln(P)]/[a \times \exp(\beta_1 C + \beta_2 M)]\}/b$ .

Male			Female		
$P$	$T_{\max}$	$T_{\max}$ age	$P$	$T_{\max}$	$T_{\max}$ age
<i>Radioactive decay model</i>					
463	9.43	114.43	3,373	12.92	117.92
4,630	12.97	117.97	33,730	16.58	121.58
46,300	16.51	121.51	337,300	20.24	125.24
463,000	20.05	125.05	3,373,000	23.90	128.90
4,630,000,000,000	44.82	149.82	3,373,000,000,000	45.86	150.86
<i>Gompertz model as calculated in Barbi et al. (1)</i>					
463	8.90	113.9	3,373	11.94	116.9
4,630	12.00	117.0	33,730	15.01	120.0
46,300	15.00	120.0	337,300	17.97	123.0
463,000	17.81	122.8	3,373,000	20.81	125.8
46,300,000,000,000,000	43.77	148.8	33,730,000,000,000,000	44.64	149.6





# AAAS EPI Center launch brings evidence to policy-makers

Communication will be key in sharing reliable scientific information, says center director

By **Becky Ham**

The Center for Scientific Evidence in Public Issues at the American Association for the Advancement of Science began work on 24 September to share scientific and technical evidence with policy-makers working at all levels of government in the United States.

The EPI Center, led by director Michael Fernandez, is a new initiative by the association to make nonpartisan information available to decision-makers as they act on issues from clean water to the opioid crisis. Instead of lobbying for a particular law or offering a years-long exhaustive study of an issue, the center hopes instead to create timely, well-communicated evidence narratives—what scientists know about a topic, how they know it, what the evidence means, and how it relates to other public policy issues.

“We want to have an impact on policy and policy-making, not by advocating for certain policies but ensuring that when decisions are being made, the evidence is being appropriately considered and evaluated,” said Fernandez.

The center’s launch was driven in part by the desire of AAAS members to have another outlet for shaping public policy and sharing their work in ways that bear on larger societal challenges, he said. “Our members are an amazing resource of expertise and commitment, and they will be the center’s best ambassadors.”

“At a time when decision-makers too often ignore, misunderstand, or misuse relevant evidence, we need new ways to commu-

nicate policy-relevant scientific evidence to decision-makers and influencers in all areas of government and society,” said Rush Holt, chief executive officer at AAAS. “The AAAS Center for Scientific Evidence in Public Issues will connect scientific expertise to the decisions and policies that affect our lives.”

Holt, a former congressman from New Jersey, has long supported the idea of such a center as one way to fill in the gap left by the 1995 elimination of the congressional Office of Technology Assessment, said Kei Koizumi, a senior science policy adviser at AAAS who guided the EPI Center’s development.

“We know it is not automatic for decision-makers to be able to access information,” said Koizumi, who served in the White House Office of Science and Technology Policy from 2009 to 2017. “But with AAAS as the world’s largest general scientific society, we have the opportunity with our community to give policy-makers the access they need.”

Fernandez’s new role at AAAS is something of a homecoming, he said, since he first came to Washington, D.C., as a 1991–1992 Congressional Science & Engineering Fellow sponsored by the American Society of Plant Physiologists working in the U.S. Senate Committee on Agriculture, Nutrition, & Forestry. Since then, he has worked in the federal government at the U.S. Environmental Protection Agency and the Department of Agriculture, on the nonprofit Pew Initiative on Food and Biotechnology project, and in the private sector at Mars, Inc.

With the guidance of an advisory board that is still being assembled, Fernandez and the EPI Center staff will analyze issues that are already at the heart of public policy

discussions or are likely to become the focus of policy in the near term. “One of the first things I’m going to do is go out and ask people about the issues they are facing as governor, or as mayoral associations, or as congressional committees,” said Fernandez. “We need to be in listening mode.”

The science policy ecosystem is crowded and diverse, he said, noting that one of his goals as director will be to ensure that the EPI Center’s work gets noticed. “The idea of narrating the evidence is a different approach than anyone else I know is taking,” he said. “Our goal is not to commission a bunch of white papers and release them to an audience where they sit on someone’s shelf. This will be more nimble, more dynamic, and more interactive.”

The narratives will be crafted with specific audiences and specific delivery tools in mind, Fernandez said. For instance, an evidence narrative about self-driving cars could be aimed at state regulators as well as congressional staff and might be best delivered to those audiences through local workshops.

Fernandez’s daughter worked this summer as an intern in the office of the mayor of Boston, reminding him of the need to “be outside the Beltway as much as we can,” he said. “If you want to have a big policy-making impact, you can’t ignore the federal government, but frankly there is a bigger need for support and help in understanding what the evidence is and the narrative around evidence

that regulators and policy-makers are grappling with."

During his leadership of the Pew initiative, Fernandez and his colleagues partnered with the U.S. Food and Drug Administration and the U.S. Department of Agriculture in small workshops to discuss biotechnology in animal agriculture. The center may look for similar opportunities to share its findings, he said. "I find that kind of participatory and interactive event to be a good way of helping to inform a policy-maker audience. Reports can be enormously influential, but I think that you can't underestimate the value of face-to-face interactions."

The center may also use interactive media and multimedia to communicate its narratives, he said. "From my work in the private sector, we saw how social media and other kinds of communication tools have really changed the landscape in the policy-making arena, and I think that we will absolutely need to understand how to leverage those tools to reach a broader audience."

One of the biggest challenges "is to be as unbiased and neutral as possible" and to avoid any notion that the center is advocating for scientists as "just another special interest group," said Fernandez.

Positioning the center as a source of trustworthy information, he

said, is important in light of "the erosion of trust in science and scientific evidence in policy-making broadly, which is particularly critical at a moment in time when so many of the issues that we're addressing have a strong scientific and evidence-based component to them."

While there are some decision-makers who willfully disregard or undermine scientific evidence, Fernandez said, "I think there's lots of reasons evidence is underappreciated, underevaluated, underutilized in policymaking that are not just people ignoring it." Some policymakers may not fully consider the evidence because it's not being presented or provided in a way that is compelling or relevant to them and their districts and their constituents, he said.

"There is a real hunger for this kind of thing among the community," he said. "As long as we can establish ourselves in a way that will be taken seriously and trusted, we will have an opportunity to do something that will make a difference."

The center is funded by the Gordon and Betty Moore, Rockefeller, Alfred P. Sloan, David and Lucile Packard, and Rita Allen Foundations; the Hellman family; the Carnegie Corporation of New York; and the Chan Zuckerberg Initiative.

## Golden Goose prize recognizes serendipitous discoveries

Roots of the immune system, cell communication, and implicit bias are honored in the 2018 awards

By **Becky Ham**

A chance encounter with a gland located in the rear end of a goose was the start of something big for Bruce Glick, a lifelong bird watcher who turned his interests toward poultry science after serving in World War II. Glick was curious about the purpose of the gland, known since the early 17th century as the bursa of Fabricius, named after Italian anatomist Hieronymus Fabricius.

"Good question," said George Jaap, his Ph.D. adviser at Ohio State University, when Glick asked about the bursa. "You find the answer."

The answer came when Glick and fellow graduate student Timothy Chang discovered by accident that geese missing their bursa glands were unable to produce antibodies. As the experiments piled up over the years, it became clear that Glick had uncovered the fundamental division of labor between B and T cells in the immune system. The finding underlies much of what scientists know about diseases from leukemia to AIDS and has transformed the produc-

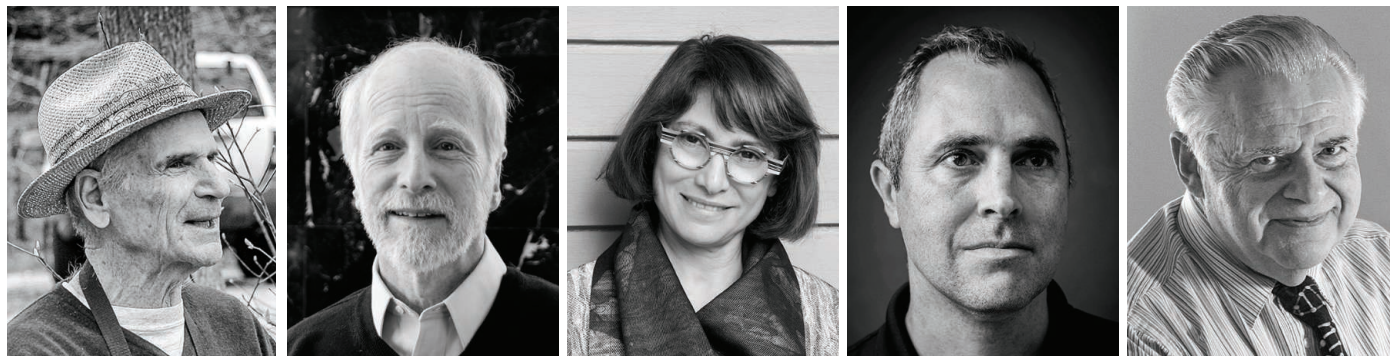
tion of vaccines and monoclonal antibody drugs to treat arthritis, multiple sclerosis, and several cancers.

Fittingly for a man with a lifelong love of birds, the late researcher was honored with one of three 2018 Golden Goose Awards, presented each year by science and higher education organizations led by the American Association for the Advancement of Science and a bipartisan group of congressional supporters. The award, established in 2012, recognizes scientists whose federally funded work may have been considered odd or obscure at first but has resulted in significant benefits to society.

AAAS has a long tradition of supporting federal funding of basic research. Long before the Golden Goose award, President Dwight D. Eisenhower underscored the necessity at a 1959 New York City symposium held by AAAS and the National Academy of Sciences.

"Government's role in research and its responsibility for advancing science must be large and there must be persistent partnership between government effort and private effort," said Eisenhower at the event. "Our science and technology are the cornerstone of American security, American welfare, and our program for a just peace. For the government to neglect this would be folly."

Glick, who was on the faculty at Mississippi State University and Clemson University, is "one of these amazing agricultural scientists who toil away and no one recognizes what they've done," said Sonny Ramaswamy, president of the Northwest Commission on Colleges and Universities, who nominated Glick for the award. "But his work, in multiple disciplines, has really been the kind of seminal



Golden Goose recipients Bruce Glick, Anthony Greenwald, Mahzarin Banaji, Brian Nosek, and Stanley Cohen.



discoveries and inventions that are the lore of textbooks."

Birds—chickens this time—played another key role in recipient Stanley Cohen's discovery of cytokines, cell secretions that affect the behavior of neighboring cells and are often called the "vocabulary" of cell communication. Cohen, a professor emeritus at Rutgers New Jersey Medical School, was studying the immune cells called lymphocytes to learn how they help to protect the body against infectious diseases like tuberculosis. Serendipity struck again, in the form of a batch of chicken eggs that Cohen had infected with a virus as part of a failed attempt to show that the virus could suppress immune responses. Before tossing them out, Cohen decided to examine them further, finding some of the small proteins that he would later name cytokines.

Scientists since have identified more than 100 cytokines across every organ system with diverse roles in growth, development, immune defense, and cancer. Cytokines are also at the base of a thriving group of therapeutics for cancer, inflammatory, and autoimmune diseases. As of 2011, more than 120 companies were developing more than 270 therapies that involve cytokines, according to the Cambridge Healthtech Institute Pharma Reports.

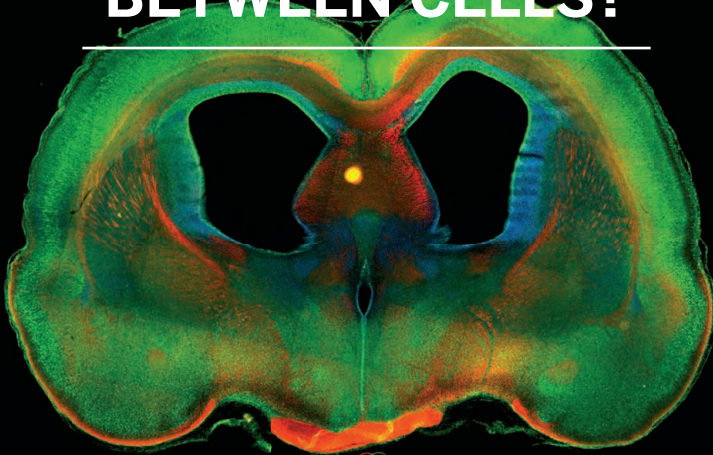
Cohen's first papers on cytokines were rejected by journals, but he persevered with funding from the U.S. National Institutes of Health. "The way to approach failure is to look carefully at what went wrong to see if there are any clues that might lead to something interesting that was unexpected," he said. "Ultimately, though, what keeps any scientist going is the thrill of discovery. There are lots of unanswered questions out there."

Harvard professor Mahzarin Banaji and her colleagues Brian Nosek at the University of Virginia and Anthony Greenwald at the University of Washington were recognized for their role in developing the concept of implicit bias, the idea that people unconsciously carry cultural beliefs and attitudes that influence whether we perceive different groups of people as "good" or "bad." Banaji and her colleagues have collected data on implicit bias through the Implicit Association Test website (<https://implicit.harvard.edu>), the results of which have led to widespread changes in hiring practices, law enforcement and physician training, and even politics. During the 2016 U.S. election, the concept of implicit bias with regard to racism and policing featured prominently in the presidential and vice-presidential debates.

Bananji said the results have been disturbing to some in the scientific and broader community since the beginning. "We knew one thing from the history of our field: that good people often are capable of surprisingly bad behavior," she said. "To us, our research was providing a new and more dramatic view of ourselves as fallible, and there was no option but to pursue it, no matter what anybody else thought."

Like previous Golden Goose recipients, she stressed that federal funding for her studies was a vital part of expanding an early idea into a concept with widespread social impact. She mourned the recent reduction of basic research funding of social cognition studies through the National Institute of Mental Health. "We will never know about the many great discoveries that could have been made with small amounts of funding to people very much like the ones who are being honored today," she said.

## CALL FOR PAPERS! DOES YOUR LAB ANALYZE THE MECHANISMS THAT MEDIATE COMMUNICATION BETWEEN CELLS?

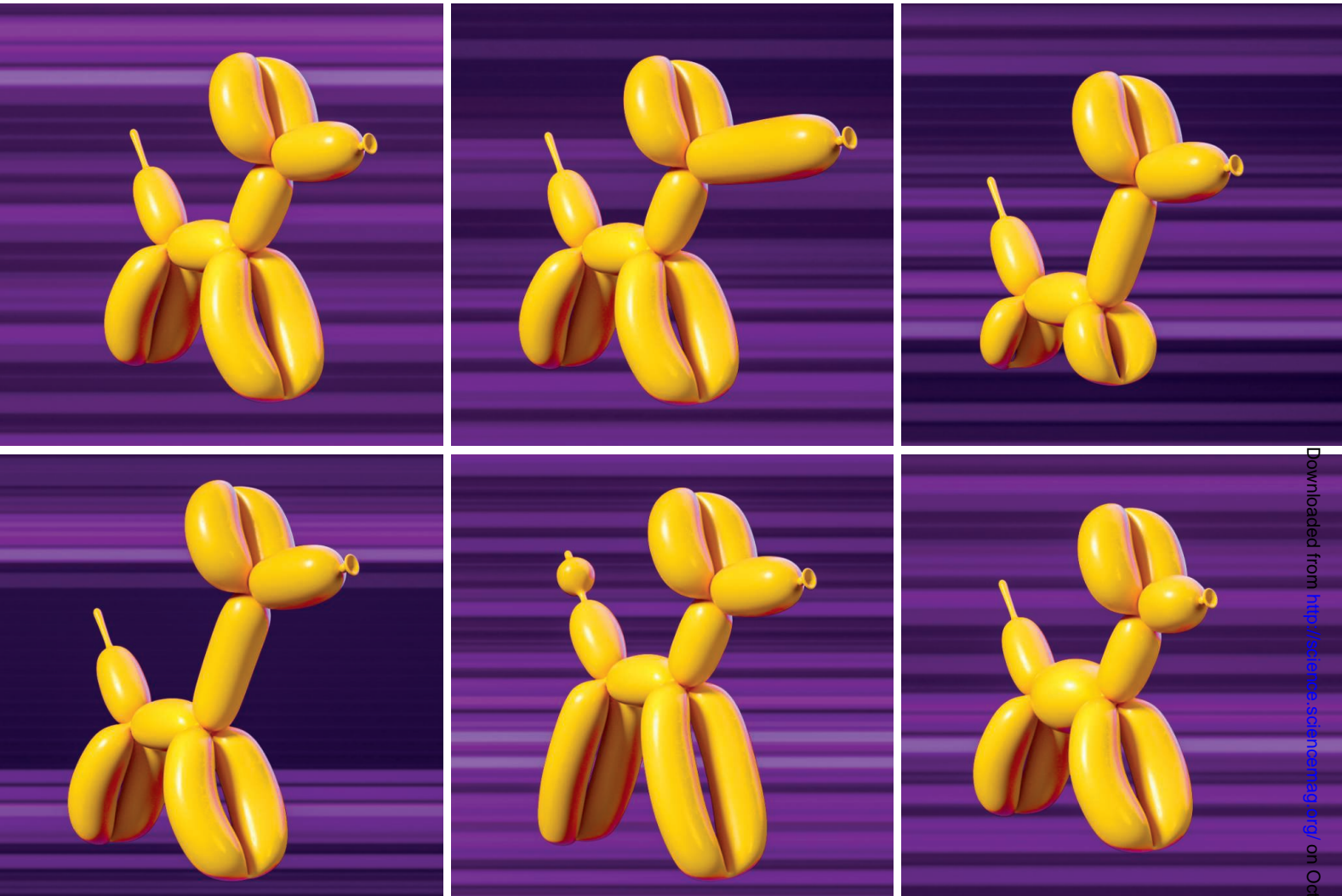


Kong-Yan W Kong-Yan Wu *et al.* (Zhen-Ge Luo), "Semaphorin 3A activates the guanosine triphosphatase Rab5 to promote growth cone collapse and organize callosal axon projections", *Sci. Signal.* 7, ra81 (2014). Rat Brain Slice. Image: Kong-Yan Wu and Zhen-Ge Luo, Chinese Academy of Sciences.

Stay on top of the latest advances in brain development and neurological disorders with *Science Signaling*, the leading online journal of cross-disciplinary cell signaling research. The journal's high-impact articles showcase basic research related to cellular and organismal regulation relevant to development, physiology, and disease as well as applied signaling research important for drug discovery and synthetic biology.

Learn more and submit your research today:  
**ScienceSignaling.org**

**ScienceSignaling** | AAAS  
CELL SIGNALING IN PHYSIOLOGY AND DISEASE

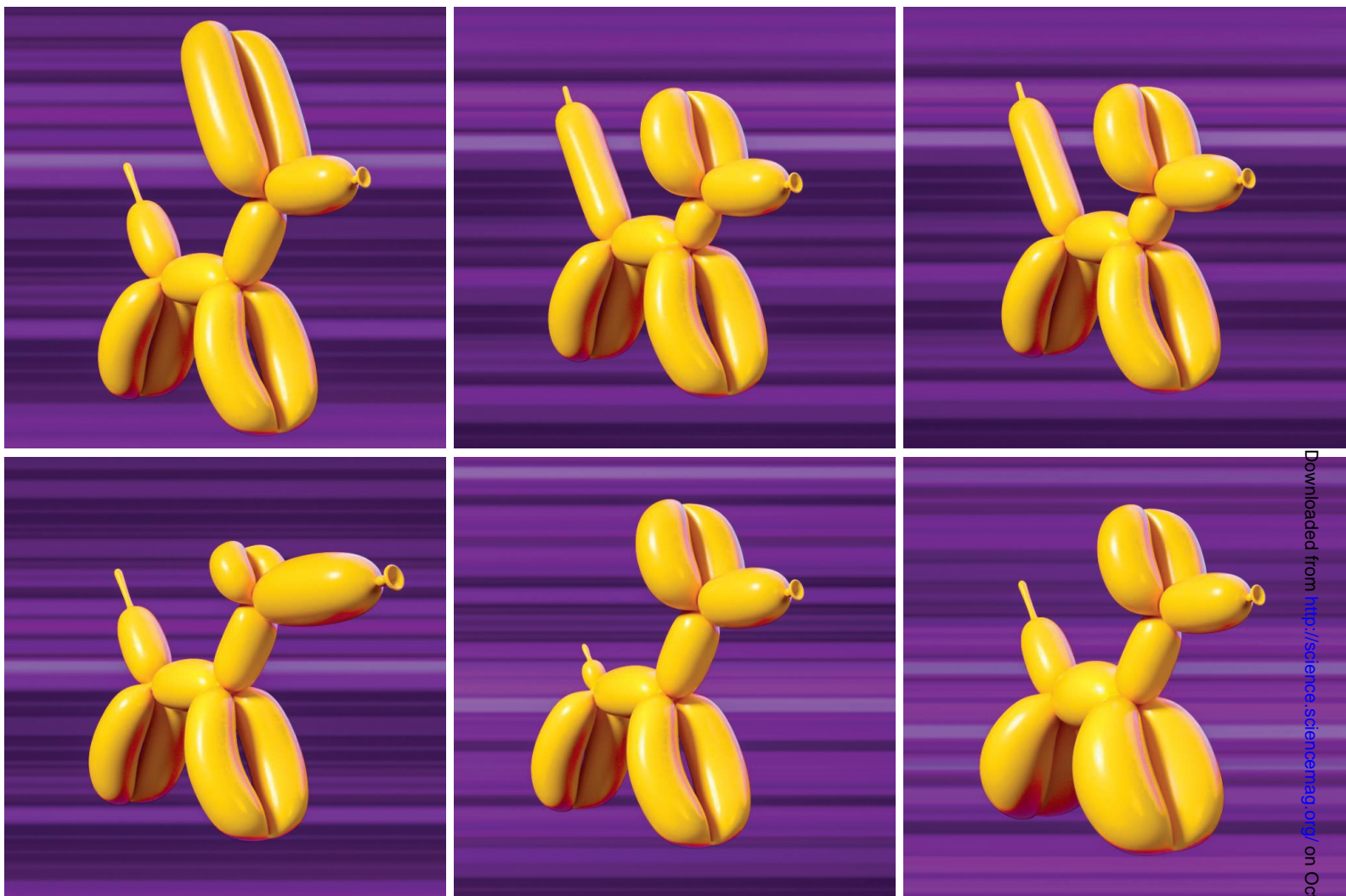


Developmental changes, represented by balloon dogs of various shapes and sizes, result from gene regulatory mechanisms altering gene expression patterns, denoted by heat maps.

# FORCES BEHIND FORM

*By* **Beverly A. Purnell, Steve Mao, and Laura M. Zahn**





Presenting a balloon dog to a 4-year-old or a sonogram to an expectant mother elicits pure amazement at the developing body. For both, forces transform dividing structures into a recognizable form. Yet living creatures require more than simple twists and turns. Elegant and precise genetic programs guide the forces that allow seemingly identical starting cells to develop into highly specialized entities: a beating cardiomyocyte, an insulin-secreting pancreatic beta cell, a bone, or a light-sensing retina. But what factors and mechanisms are at play? Researchers are using new technology to tease out the details. It is clear that molecular associations are key, whether they are changes in the topological arrangement of chromosomal regions that bring one part of the genome closer to another or chemical modification of nucleotides or proteins that recruit or repel protein binding to influence chromatin state. Histone chaperones and remodelers also contribute to the versatility of the genetic and epigenetic landscape to direct specific cell fate when cells are subjected to varied environmental conditions, metabolic changes, and even disease. This special issue highlights recent advances in our understanding of the role of gene regulation in development. By revealing the forces directing an organism's form, we gain a better understanding of the normal developmental process, with potential for clinical intervention when things go awry.

## INSIDE

### REVIEWS

Chromatin plasticity:  
A versatile landscape that  
underlies cell fate and  
identity *p. 1332*

Dynamic DNA methylation:  
In the right place at the right  
time *p. 1336*

Developmental enhancers  
and chromosome topology  
*p. 1341*

RNA modifications modulate  
gene expression during  
development *p. 1346*

### RELATED ITEMS

- PERSPECTIVE *P. 1310*
- RESEARCH ARTICLE *P. 1354*
- REPORT *P. 1377*
- REPORT *P. 1380*
- RESEARCH ARTICLE BY  
T. GERBER ET AL.  
[10.1126/science.aag0681](https://doi.org/10.1126/science.aag0681)

## REVIEW

# Chromatin plasticity: A versatile landscape that underlies cell fate and identity

Tejas Yadav, Jean-Pierre Quivy, Geneviève Almouzni\*

During development and throughout life, a variety of specialized cells must be generated to ensure the proper function of each tissue and organ. Chromatin plays a key role in determining cellular state, whether totipotent, pluripotent, multipotent, or differentiated. We highlight chromatin dynamics involved in the generation of pluripotent stem cells as well as their influence on cell fate decision and reprogramming. We focus on the capacity of histone variants, chaperones, modifications, and heterochromatin factors to influence cell identity and its plasticity. Recent technological advances have provided tools to elucidate the underlying chromatin dynamics for a better understanding of normal development and pathological conditions, with avenues for potential therapeutic application.

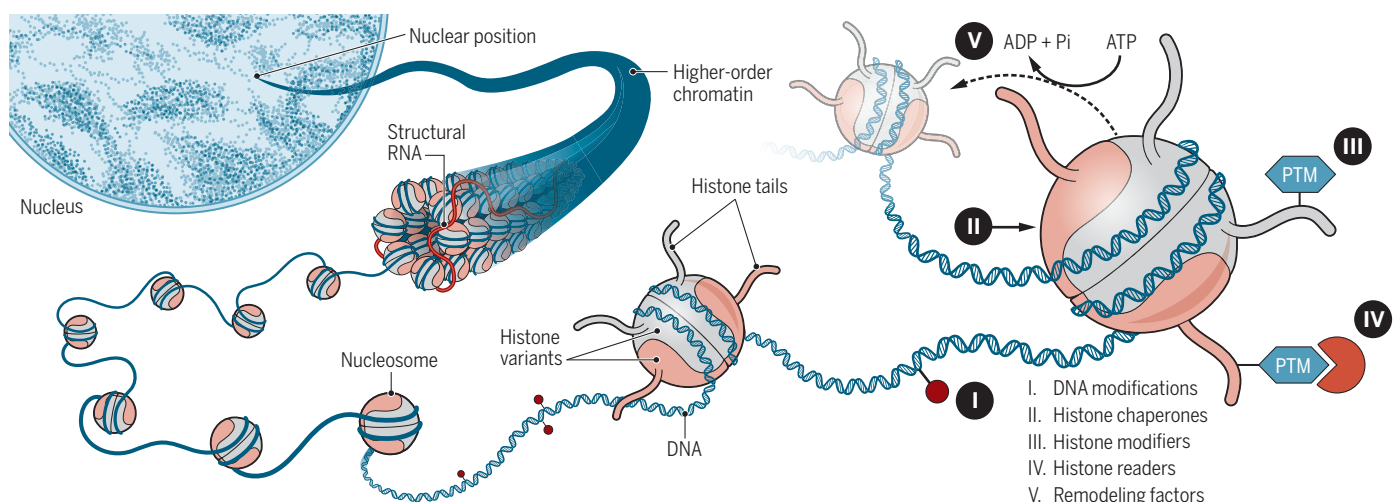
The genome of eukaryotic cells is organized into chromatin, a nuclear complex comprising DNA, RNA, and associated proteins (1, 2). Chromatin organization displays hierarchical levels ranging from the basic repeated unit, the nucleosome, to higher-level structures (Fig. 1). The nucleosome is composed of a core particle with ~147 base pairs of double-stranded DNA wrapped around histone proteins with linker DNA joining core nucleosomal units. The chromatin filament further coils and compacts DNA to reach higher-order states with interacting chromatin loops and topologically associating domains (TADs) (3). Histones come as distinct variants that undergo posttranslational modification (PTM) to provide modularity within

core particles (1). Histone chaperones, chromatin remodelers, and histone- and DNA-modifying enzymes, along with PTM readers, transcription factors, and RNA, generate specialized genomic domains for a versatile chromatin landscape. Centromeres, telomeres, and regulatory elements display unique nucleosome composition and structure. Modulation at each level enables chromatin-based information to vary in order to respond to different signals for numerous gene regulatory functions (4) (Fig. 1). This defines chromatin plasticity as a means to generate a diversity of properties for each cell type during development and also when cells face different environmental factors, genotoxic insults, metabolic changes, senescence, disease, and even death (5, 6).

Regulation of cell fate decisions and cell identity can exploit chromatin, for example, by restricting access to a particular transcription factor or by providing distinct marks that specific proteins can recognize—proteins often called “reader”—and can interpret in response to signaling (7, 8). Thus, chromatin organization is intimately linked to varied states experienced by any single cell in its lifetime, and many chromatin changes occur during embryonic development in mammals (6, 9), in embryonic stem cells (ESCs) transitioning to a differentiated state in vitro (10), during reprogramming of differentiated cells to form multi- or pluripotent cell types (11), and in various diseases (12). With the advent of single-cell approaches (13), we are beginning to capture transient or specialized states in individual cells. DNA methylation revealed major changes [see (4, 9, 14, 15) for reviews]. Here, we highlight recent reports that identify chromatin factors in controlling cell fate and identity with a focus on histone variants, PTMs, chaperones, and heterochromatin factors in the context of organism development, in vitro cell differentiation, cellular reprogramming, and disease.

## Chromatin shapes cell fate and identity in normal development

During development, highly differentiated cells, the gametes, fuse to form a totipotent zygote. This single cell then undergoes rounds of division and differentiation to give rise to every cell type in the adult organism, including gametes. Lineage-specific gene expression profiles are initiated by cell type-specific transcription



**Fig. 1. Chromatin plasticity.** The hierarchy of chromatin organization in an interphase nucleus is shown along with factors acting at each level. Arrays of nucleosomes fold into higher-order chromatin structures, and noncoding RNA participates in local organization. Inset (right): The basic repeating unit of chromatin, the nucleosome, comprises double-stranded DNA wrapped around an octamer consisting of a (H3-H4)<sub>2</sub> tetramer flanked by two H2A-H2B dimers. Variations of this basic module by dynamic combinations include the choice of histone variants, modifications of DNA bases, and reversible posttranslational modifications (PTM) of histone tails, enabling chromatin plasticity.

Institut Curie, 75248 Paris Cedex 05, France.

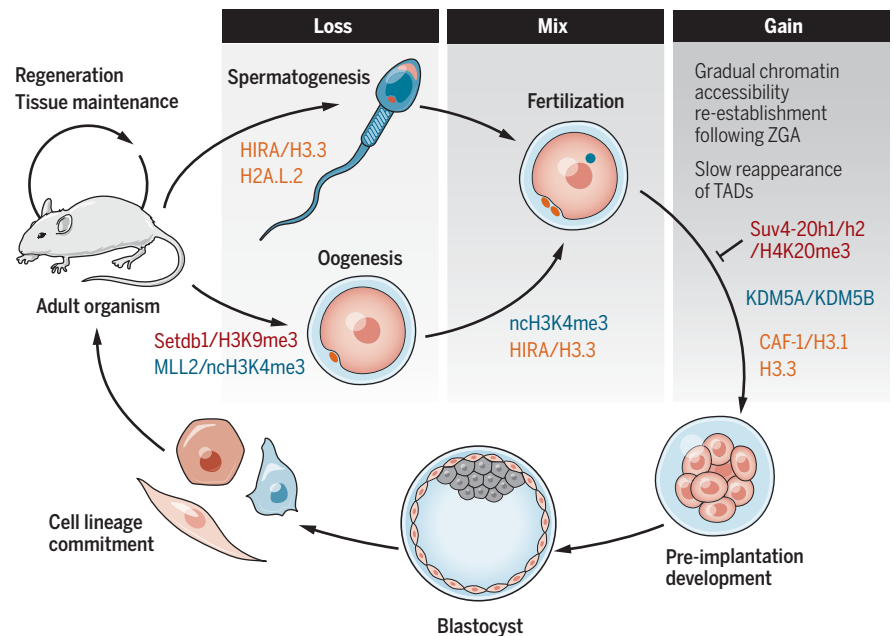
\*Corresponding author. Email: genevieve.almouzni@curie.fr



factors and controlled by signal transduction pathways. Acting in concert, chromatin factors contribute to the chromatin landscape during stages of cellular differentiation and overall organism development. This is illustrated in particular with histone chaperones and their histone variants that perform critical roles in development (6). Here, we compile recent reports placing them within the development cycle of an organism. We highlight the most drastic changes during gametogenesis and early zygotic development for reprogramming (Fig. 2), with a focus on how heterochromatin can restrict cell fate.

## Gametogenesis

In mammals, genome compaction in spermatogenesis involves major histone reshuffling. A spermatid-specific histone H2A variant, H2A.L.2, is required for the acquisition of nucleoprotamines and genome compaction in mature spermatozoa (16) (Fig. 2). During compaction, it is necessary to mark chromatin at specific genes required for embryonic development. In zebrafish, specialized “placeholder” nucleosomes constitute a bookmarking mechanism to prevent DNA methylation-mediated repression at housekeeping and early embryonic transcription factor genes. They contain a special H2A variant, H2A.Z(FV), that co-occurs with the methylated histone H3 Lys<sup>4</sup> (H3K4me1) mark in transcriptionally silent gamete and embryonic stages of zebrafish development at key genomic regions in order to ensure a supply of housekeeping gene products and transcription factors (17). Meanwhile, during oogenesis, certain histone PTMs differentiate active from inactive genomic regions. The trimethylated histone H3 Lys<sup>4</sup> (H3K4me3) mark generally correlates with active promoters but also occurs at a lesser degree at nonactive regions. Overcoming the limited amounts of mammalian genetic material available from eggs and embryos, two recent studies used low-input chromatin immunoprecipitation sequencing (ChIP-seq) to probe H3K4me3 levels during oogenesis (18, 19). Whereas only active promoters in nondividing oocytes showed H3K4me3, this PTM accumulated further as oogenesis progressed in a transcription-independent manner involving the MLL2 activity (18) (ncH3K4me3; Fig. 2). This noncanonical transcription-independent H3K4me3 marked intergenic regions, putative enhancers, and trimethylated histone H3 Lys<sup>27</sup> (H3K27me3)-silent promoters. Moreover, the noncanonical H3K4me3 distributed over broad regions correlating with partially methylated DNA domains on maternal genomes in mature oocytes. The presence of these noncanonical H3K4me3 domains on the maternal allele in zygotes and two-cell embryos suggests inheritance from oocytes (19), a feature distinct from somatic cells and ESCs. Thus, peculiar interplay between histone variants and key histone PTM shows remarkable dynamism as a marking system to prepare for subsequent development.



**Fig. 2. Histone variants and modifications involved during the development cycle.** Histone variants and their chaperones, histone modifiers and posttranslational modifications (PTMs), and heterochromatin factors are used during various stages of mouse development. The dynamic changes in the developmental cycle follow distinct steps: first an erasure or loss of parental marks during gametogenesis, followed by the mix of marks after fertilization to reach totipotency, and finally a progressive gain of specific chromatin marks and higher organization in line with the diversity of cell types forming the adult organism.

## Zygotic development and cell trajectories

After fertilization, development starts based on maternally inherited proteins and RNAs from the egg cytoplasm. A crucial reprogramming of parental epigenomes occurs to reach zygotic totipotency. This transition involves the histone H3 variant H3.3 and its associated chaperones, HIRA and DAXX, important both for meiotic segregation and fertility in mammals (6) (Fig. 2). After maternal-to-zygotic transition (MZT) and clearance of maternal products, zygotic genome activation (ZGA) is a time of extensive chromatin changes (14). It takes place around the late two-cell stage in mice and four- to eight-cell stages in human embryos. Recently, DNase I hypersensitivity (20) and scCOOL-seq (single-cell chromatin overall omics-scale landscape sequencing) experiments (21) showed a gradual reestablishment of chromatin accessibility in human embryos (Fig. 2). High-throughput ATAC-seq (assay for transposase-accessible chromatin) in human preimplantation embryos captured accessible chromatin regions existing prior to ZGA, especially at CpG-rich regions, several of them becoming inaccessible after ZGA in a transcription-dependent manner (22). Before gastrulation, spatial heterogeneity of regulatory genomic regions correlates with future cell fate. In flies, single-cell ATAC-seq studies at different developmental stages showed the dynamics of chromatin accessibility during lineage commitment (23), demonstrating the temporal placement of cells along differentiation trajectories. During mid-

and late embryonic stages, tissue-specific signatures of chromatin accessibility emerge, although individual cells still display features reminiscent of their original germ layer. The histone chaperone CAF-1 (specifically, the large subunit p150), which incorporates replicative histone variant H3.1 necessary for cell division, is crucial for preimplantation development (1, 2, 6) (Fig. 2).

ZGA uses several barriers against precocious activation of lineage-specific genes in early embryogenesis. These include active and repressive histone PTMs. During ZGA in mice, reprogramming of the noncanonical histone mark H3K4me3 is observed. H3K4me3 inherited in preimplantation embryos gets removed by the lysine demethylases KDM5A and KDM5B (Fig. 2) to constrain this mark to transcription start sites (TSS) in the late two-cell embryos at the onset of zygotic transcription (24). Thus, in contrast to oocytes with broad domains, H3K4me3 shows sharp, more confined peaks in late-stage embryos. Hence, a PTM associated with actively transcribed and poised gene promoters is highly plastic as embryos develop. Interestingly, the H3K27me3 mark associated with gene repression imposed by Polycomb complexes shows distinct dynamics. In *Drosophila* embryos, the maternal contribution of H3K27me3 counteracts premature untimely accumulation of the active H3K27ac mark at regulatory regions (25). Indeed, loss of the maternally inherited H3K27me3 mark leads to embryonic lethality that cannot be circumvented by reestablishment

of the PTM at a later zygotic stage (25). Similarly, a recent study in mice found that oocyte-acquired H3K27me3 patterns are transmitted to the zygote and involved in a novel form of DNA methylation-independent genomic imprinting of allelic loci in early embryos (26). Comparison of H3K4me3 and H3K27me3 dynamics by small-scale ChIP-seq revealed that whereas the active mark is rapidly reestablished, H3K27me3 is slower in preimplantation embryos in mouse, *Xenopus*, and zebrafish (27).

### Heterochromatin and cell fate restriction

Constitutive heterochromatin at pericentromeric regions is commonly associated with trimethylated histone H3 Lys<sup>9</sup> (H3K9me3) in cells (2, 28), a mark that acts as a key chromatin barrier to cell fate changes. The genome-wide distribution in mouse gametes and early embryos (29) shows a dynamic distribution of H3K9me3 at promoters and long terminal repeats (LTRs). After fertilization, both parental genomes show massive H3K9me3 reprogramming and reestablishment, although the discrepancy in parental H3K9me3 signals lasts until the blastocyst stage and is not fully recovered. The large subunit of histone chaperone CAF-1 (mouse *Chaf1a*) is responsible for establishing H3K9me3 on LTRs and their eventual silencing (29, 30). The observation of lineage-specific H3K9me3 raised interest in exploring roles for heterochromatin in regulating cell fate commitment. Loss of Setdb1 (H3K9 methyltransferase) in growing oocytes leads to meiotic defects and down-regulation of retrotransposon elements (31), and Setdb1 maternally deficient embryos arrest at preimplantation as a result of cell cycle progression and chromosome segregation defects (Fig. 2). The heterochromatin mark trimethylated histone 4 Lys<sup>20</sup> (H4K20me3) is undetectable in mouse preimplantation embryos, and ectopic establishment of this mark by expression of Suv4-20h1/h2 hinders development (Fig. 2), likely by altering S-phase progression in this developmental context (32). How these modifications interconnect with H3K27me3 will be important to consider.

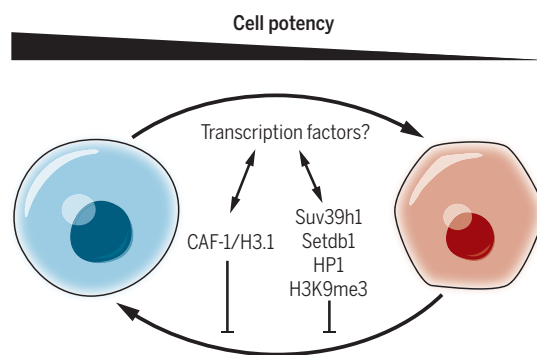
As seen with H3K9me3 marks, after fertilization both parental genomes show remarkable loss of 3D nuclear structure in a low-input Hi-C study (33). The higher-order states, as reflected with interacting chromatin loops and defined topologically associating domains (TADs), are thus lost. TAD boundaries and differential genomic compartments arise only gradually in the zygote, and a full restoration of the overall higher-order architecture involves a long maturation process (3) (Fig. 2).

Together, these recent discoveries highlight critical chromatin dynamics during development and reproduction. Gametogenesis and developing embryos in model organisms teach us how the chromatin machinery contributes

to fine-tuning of cell fate commitment. These mechanisms may also be reused during adult life in regenerating tissues and potentially exploited for therapeutic purposes in regenerative medicine.

### Lessons from cellular models: Studying development in vitro

Cell culture of pluripotent ESCs (9) and 3D culture models have proven extremely useful for experimental manipulation of somatic and germline stem cells and organoids (34). Recent findings show that ground-state (derived using the so-called “2i” combination of GSK3 and MEK inhibitors), serum-grown, or “primed” mouse ESCs (mESCs) and human ESCs exhibit features



**Fig. 3. CAF-1 and heterochromatin as controllers of cell plasticity.** A schematic model highlights the role of the histone chaperone CAF-1, the replicative histone H3.1, and the presence of the H3K9me3 mark typical of heterochromatin. Chromatin assembled using the replicative histone variant, along with imposing the H3K9me3 mark, provides a barrier to stemness engaging cells in nonreversible cell fates.

of accessible chromatin in comparison with somatic cells, although the features of chromatin are less distinct than in earlier preimplantation stages (35). Recently, mESC lines with tagged H3.3 permitted the tracking of H3.3 dynamics before and after ESC differentiation (36, 37). Notably, in ESCs, a hyperdynamic (−1 position) H3.3 nucleosome marks gene promoters; upon differentiation, this nucleosome shifts downstream (to the +1 position) (36). These discrete dynamics in histone variant positioning suggest that fine-tuning at this level can regulate cell fate determination.

The lifetime of a particular histone variant is also a hallmark of stem cells, as shown in fly intestinal adult stem cells (ISCs). Notably, in ISCs, the specific centromeric variant CENP-A is retained for weeks in the self-renewing population—a mark of stemness (38). A long-lived form of CENP-A is thus characteristic of these populations of cells. In contrast, when these cells differentiate, in the daughter cells, the differentiating cell receives new CENP-A with the help of the fly version of the mammalian HJURP chaperone responsible for de novo CENP-A deposition, while the new daughter stem cell retains the parental CENP-

A. This asymmetric distribution of the parental CENP-A follows the fate of the stem cell. These reports provide attractive cellular models to track cell fate changes and further probe the role of chromatin organization and dynamics of H3 variants.

Cellular models have also illuminated the importance of heterochromatin in the genome. HP1 (heterochromatin protein 1)/Suv39h1 (histone methyltransferase) recruitment to active promoters leads to reversible gene silencing in mESCs by establishing heterochromatin, as shown in experiments using FIRE-Cas9 (Fkbp/Frb inducible recruitment for epigenome editing by Cas9) (39). Recent work in mESCs uncovered an unusual form of repression involving HP1 together with ADNP (activity-dependent neuroprotective protein) and CHD4 (chromatin remodeler) in a complex called ChAHP (40). ChAHP-mediated repression acts locally and does not rely on H3K9me3-modified nucleosomes. Ablation of this complex led to spontaneous differentiation along with precocious expression of lineage-specific genes (40).

Bivalent promoters are defined by the presence of a nucleosome combining both an active mark (H3K4me3) and a repressive mark (H3K27me3) on the same particle and thus poised to either become activated or kept repressed. They are more prevalent in cultured mESCs than in mouse early embryos and provide an entry point to explore how their chromatin modulation contributes to transcriptional output. Activation of bivalent genes occurs within minutes at target loci in mESCs when targeting chromatin remodeling via mSWI/SNF(BAF) complexes using the FIRE-Cas9 method to oppose the activity of Polycomb complexes (39). Conversely, the prevalence of Polycomb-mediated H3K27me3 at these promoters maintains them in a silent state. Intriguingly, H3K27 inheritance in mESCs may not be only self-sustained, as initially proposed, because these patterns can be established de novo (41). This is in contrast to germline-inherited Polycomb memory in flies and mice, as discussed above (25, 26). Hence, the maternally transmitted H3K27me3 that controls lineage-specific genes in vivo is not retained in the mESCs.

In common laboratory ESC cultures, most cells are pluripotent, with infrequent two-cell-like cells exhibiting characteristics of totipotency with increased plasticity. Quantitative polymerase chain reaction–based microfluidics single-cell expression profiling characterized these two-cell-like cells, and a small interfering RNA–based screening revealed key chromatin factors, including a noncanonical Polycomb PRC1 complex (PRC1.6) and the EP400-TIP60 complex (42). Higher-order chromatin also undergoes dynamic changes during differentiation of stem cells, as shown by high-resolution ultra-deep Hi-C mapping of the distinct signature in mESCs and

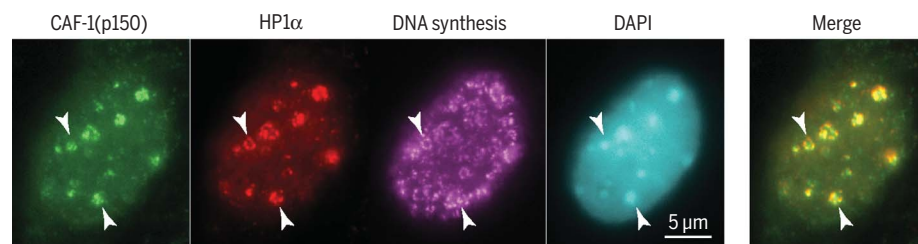


neural progenitor cells where distal gene bodies of active genes interact extensively (43). During neural differentiation, these preexisting long-range contacts between active TADs weaken while those between inactive regions become stronger. Cell type-specific enhancer-promoter contacts are formed in parallel to the expression of differentiation genes. Studies in ESCs have thus revealed unique roles of histone modifiers, variants, and chaperones in maintenance and/or changes in cellular identity that could also prove important for reprogramming.

### Chromatin plasticity in cell reprogramming and deregulation in pathologies

Manipulation in vitro affords the exciting possibility of reverting unipotent differentiated cells back to a pluripotent stem cell-like state, a process called reprogramming (11), as in the case of induced pluripotent stem cells (iPSCs) derived directly from skin cells with a limited set of transcription factors (44). However, the reprogramming efficiency remained limited, leading to a search for additional chromatin players involved in maintenance of somatic cell fate. The histone chaperone CAF-1, along with Setdb1, proved important in maintaining somatic cell identity (45). Depletion of CAF-1 augments the reprogramming efficiency of mouse embryonic fibroblasts (MEFs) (45). Interestingly, down-regulation of CAF-1(p150) in ESC leads to increased induction of 2C-like cells with greater potency than the ESC (30). Notably, in addition to in vitro reprogramming, natural reprogramming also occurs in intestine regeneration and skin renewal of the adult organism. This is also exemplified by the immune response, which underscores the paradigm of chromatin plasticity in differentiated cells. Using single-cell RNA sequencing, a recent study found that CD8<sup>+</sup> T cells lacking Suv39h1 have enhanced long-term memory and improved reprogramming capacity (46). The stability of the commitment into T helper 2 (T<sub>H</sub>2) lineage from CD4<sup>+</sup> T cells is maintained by the SUV39H1-H3K9me3-HP1 $\alpha$  pathway (46). It is remarkable that in distinct contexts regarding the degree of cell potency, both CAF-1 and the H3K9me3-heterochromatin pathways (i.e., Setdb1, Suv39h1, HP1 $\alpha$ ) restrict acquisition of increased potency and/or remodeling capacity (Fig. 3). Given the interactions and colocalization of CAF-1 with HP1 domains during S phase (47) (Fig. 4), whether CAF-1 and H3K9me3-heterochromatin pathways act independently or in concert will need further investigation. Furthermore, how these chromatin modulations connect with the function of specific transcription factors and changes in the cell cycle remains to be established. Nonetheless, this provides an example of different, overlapping chromatin layers influencing cell fate decisions, which should be explored for regeneration of cells in the adult organism.

Deregulation during normal aging and in many disease states represent natural in vivo cases of unscheduled chromatin alterations and un-



**Fig. 4. CAF-1 and heterochromatin marks within the cell.** During late S phase at sites of 5-bromo-2'-deoxyuridine incorporation that reveal replication sites, CAF-1 closely localizes in foci with HP1 $\alpha$ . A mouse embryonic fibroblast shows localization of the large subunit of CAF-1 (p150), HP1 $\alpha$ , DNA synthesis, and DNA. We visualized p150 and HP1 $\alpha$  by immunofluorescence staining, DNA synthesis by immunodetection of BrdU incorporation, and DNA by 4',6-diamidino-2-phenylindole (DAPI) staining as described in (47). A merge of the staining corresponding to detection of the large subunit of CAF-1 (p150) and HP1 $\alpha$  is shown. The arrowheads point to (pericentric) heterochromatin domains.

desirable somatic cell fate changes. This is exemplified in aging and neurodegenerative diseases (48). During advanced age-related macular degeneration (AMD), a major cause of vision loss in the elderly, chromatin accessibility (as determined by ATAC-seq) decreases globally in retinal pigmented epithelium (49). Cancer, where pathological mechanisms likely hijack existing chromatin plasticity to cause disease, further underlines the role of chromatin. Consider p53-deficient tumors that show a dependence on high levels of HJURP, the CENP-A chaperone, thereby suggesting that centromeric chromatin integrity is involved in tumor maintenance (50). HJURP is also an independent prognostic marker of luminal A breast carcinoma (51). These data link together chromosomal architecture and centromere function in cell fate maintenance. In addition to the centromeric histone variant, other histone H3 variants (H3.1 and H3.3) and histone chaperones have been implicated in malignancies. Widespread efforts are under way to examine the molecular etiology of oncohistones carrying point mutations in histone H3 Lys<sup>27</sup>, Gly<sup>34</sup>, and Lys<sup>36</sup> linked to distinct cancer types, respectively (52, 53). In these situations too, histone variant choice, histone chaperone, histone modifications, and heterochromatin are simultaneously involved in the dysregulation of cell identity.

A deeper understanding of the chromatin plasticity that wires these cell fate decisions is now guiding precision medicine and targeted therapeutics in a move toward cell precision medicine. Small-molecule inhibitors against EZH2 are being widely tested in trials along with other Polycomb subunit and BET bromodomain (chromatin reader domains that recognize acetylated histones) inhibitors that have shown therapeutic potential in diffuse intrinsic pontine glioma (DIPG) (53, 54). Trying to better target the cells that are most susceptible to the intervention at the chromatin level will be crucial. Moving forward, the ability to obtain patient-derived iPSCs and organoids to carry out transdifferentiation offers new openings to examine and address aging as well as to

treat diseases such as cancer and degenerative maladies.

### Conclusion

Research in developmental and stem cell biology has become increasingly enmeshed with the study of chromatin-based regulatory mechanisms. Although cell identity and chromatin organization are intimately linked, understanding whether chromatin plasticity is the cause or consequence of cell fate changes requires further investigation with dedicated model systems. With the help of multidisciplinary approaches looking at single-cell dynamics with high-resolution imaging in combination with high-throughput “omics” methods, there is now hope to answer questions about cellular heterogeneity and cell identity specification with unprecedented precision (55, 56). In regenerative medicine and disease treatment, targeting chromatin factors represents a promising avenue. Currently, immunotherapy and anticancer compounds combined with actionable chromatin targets are at the forefront of medicine. Multidisciplinary approaches will be instrumental to gain a comprehensive view of chromatin plasticity that may yield an understanding of how to switch between specific cellular states at will for therapeutic application.

### REFERENCES AND NOTES

1. D. Sitbon, K. Podsypanina, T. Yadav, G. Almouzni, *Cold Spring Harb. Symp. Quant. Biol.* **82**, 1–14 (2017).
2. A. V. Probst, E. Dunleavy, G. Almouzni, *Nat. Rev. Mol. Cell Biol.* **10**, 192–206 (2009).
3. E. E. M. Furlong, M. Levine, *Science* **361**, 1341–1345 (2018).
4. A. Bird, *Nature* **447**, 396–398 (2007).
5. A. Groth, W. Rocha, A. Verreault, G. Almouzni, *Cell* **128**, 721–733 (2007).
6. D. Filipescu, S. Müller, G. Almouzni, *Annu. Rev. Cell Dev. Biol.* **30**, 615–646 (2014).
7. R. C. Adam, E. Fuchs, *Trends Genet.* **32**, 89–100 (2016).
8. A. Burton, M. E. Torres-Padilla, *Nat. Rev. Mol. Cell Biol.* **15**, 723–734 (2014).
9. Y. Atlasi, H. G. Stunnenberg, *Nat. Rev. Genet.* **18**, 643–658 (2017).
10. E. Meshorer, T. Misteli, *Nat. Rev. Mol. Cell Biol.* **7**, 540–546 (2006).
11. K. Hochedlinger, R. Jaenisch, *Cold Spring Harb. Perspect. Biol.* **7**, a019448 (2015).

12. A. C. Mirabella, B. M. Foster, T. Bartke, *Chromosoma* **125**, 75–93 (2016).
13. G. Kelsey, O. Stegle, W. Reik, *Science* **358**, 69–75 (2017).
14. M. A. Eckersley-Maslin, C. Alda-Catalinas, W. Reik, *Nat. Rev. Mol. Cell Biol.* **19**, 436–450 (2018).
15. C. Luo, P. Hajkova, J. R. Ecker, *Science* **361**, 1336–1340 (2018).
16. S. Barral *et al.*, *Mol. Cell* **66**, 89–101.e8 (2017).
17. P. J. Murphy, S. F. Wu, C. R. James, C. L. Wike, B. R. Cairns, *Cell* **172**, 993–1006.e13 (2018).
18. C. W. Hanna *et al.*, *Nat. Struct. Mol. Biol.* **25**, 73–82 (2018).
19. B. Zhang *et al.*, *Nature* **537**, 553–557 (2016).
20. L. Gao *et al.*, *Cell* **173**, 248–259.e15 (2018).
21. L. Li *et al.*, *Nat. Cell Biol.* **20**, 847–858 (2018).
22. J. Wu *et al.*, *Nature* **557**, 256–260 (2018).
23. D. A. Cusanovich *et al.*, *Nature* **555**, 538–542 (2018).
24. J. A. Dahl *et al.*, *Nature* **537**, 548–552 (2016).
25. F. Zenk *et al.*, *Science* **357**, 212–216 (2017).
26. A. Inoue, L. Jiang, F. Lu, T. Suzuki, Y. Zhang, *Nature* **547**, 419–424 (2017).
27. X. Liu *et al.*, *Nature* **537**, 558–562 (2016).
28. A. H. Peters *et al.*, *Cell* **107**, 323–337 (2001).
29. C. Wang *et al.*, *Nat. Cell Biol.* **20**, 620–631 (2018).
30. T. Ishiuchi *et al.*, *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015).
31. A. Eymery, Z. Liu, E. A. Ozonov, M. B. Stadler, A. H. Peters, *Development* **143**, 2767–2779 (2016).
32. A. Eid, D. Rodriguez-Terrones, A. Burton, M. E. Torres-Padilla, *Genes Dev.* **30**, 2513–2526 (2016).
33. Z. Du *et al.*, *Nature* **547**, 232–235 (2017).
34. J. Drost, H. Clevers, *Nat. Rev. Cancer* **18**, 407–418 (2018).
35. J. Wu *et al.*, *Nature* **534**, 652–657 (2016).
36. S. Schlesinger *et al.*, *Nucleic Acids Res.* **45**, 12181–12194 (2017).
37. A. M. Deaton *et al.*, *eLife* **5**, e15316 (2016).
38. A. García Del Arco, B. A. Edgar, S. Erhardt, *Cell Rep.* **22**, 1982–1993 (2018).
39. S. M. G. Braun *et al.*, *Nat. Commun.* **8**, 560 (2017).
40. V. Ostapcuk *et al.*, *Nature* **557**, 739–743 (2018).
41. J. W. Højfeldt *et al.*, *Nat. Struct. Mol. Biol.* **25**, 225–232 (2018).
42. D. Rodriguez-Terrones *et al.*, *Nat. Genet.* **50**, 106–119 (2018).
43. B. Bonev *et al.*, *Cell* **171**, 557–572.e24 (2017).
44. K. Takahashi, S. Yamanaka, *Cell* **126**, 663–676 (2006).
45. S. Cheloufi *et al.*, *Nature* **528**, 218–224 (2015).
46. L. Pace *et al.*, *Science* **359**, 177–186 (2018).
47. J.-P. Quivy *et al.*, *EMBO J.* **23**, 3516–3526 (2004).
48. A. Berson, R. Nativio, S. L. Berger, N. M. Bonini, *Trends Neurosci.* **41**, 587–598 (2018).
49. J. Wang *et al.*, *Nat. Commun.* **9**, 1364 (2018).
50. D. Filipescu *et al.*, *Genes Dev.* **31**, 463–480 (2017).
51. R. Montes de Oca *et al.*, *Mol. Oncol.* **9**, 657–674 (2015).
52. K. Funato, V. Tabar, *Annu. Rev. Cancer Biol.* **2**, 337–351 (2018).
53. F. Mohammad, K. Helin, *Genes Dev.* **31**, 2313–2324 (2017).
54. D. Morel, G. Almouzni, J. C. Soria, S. Postel-Vinay, *Ann. Oncol.* **28**, 254–269 (2017).
55. S. J. Clark *et al.*, *Nat. Commun.* **9**, 781 (2018).
56. R. M. Harland, *Science* **360**, 967–968 (2018).

## ACKNOWLEDGMENTS

We thank our colleagues in the team for critical reading, and the many inputs for our thinking in the Labex DEEP. We apologize to colleagues whose important work could not be cited here. Because of space constraints and a focus on recent reports, we used reviews to provide background and reference to the primary research papers. **Funding:** Our support comes from la Ligue Nationale contre le Cancer (Equipe labellisée Ligue), ANR-11-LABX-0044\_DEEP and ANR-10-IDEX-0001-02 PSL, ANR-12-BSV5-0022-02 “CHAPINHIB,” ANR-14-CE16-0009 “Epicure,” ANR-14-CE10-0013 “CELLECTCHIP,” EU project 678563 “EPOCH28,” ERC-2015-ADG-694694 “ChromADICT,” ANR-16-CE15-0018 “CHRODYT,” ANR-16-CE12-0024 “CHIFT,” and ANR-16-CE11-0028 “REPLICAF.” **Competing interests:** None declared.

10.1126/science.aat8950



## REVIEW

# Dynamic DNA methylation: In the right place at the right time

Chongyuan Luo<sup>1,2\*</sup>, Petra Hajkova<sup>3,4\*†</sup>, Joseph R. Ecker<sup>1,2†</sup>

The classical model of cytosine DNA methylation (the presence of 5-methylcytosine, 5mC) regulation depicts this covalent modification as a stable repressive regulator of promoter activity. However, whole-genome analysis of 5mC reveals widespread tissue- and cell type-specific patterns and pervasive dynamics during mammalian development. Here we review recent findings that delineate 5mC functions in developmental stages and diverse genomic compartments as well as discuss the molecular mechanisms that connect transcriptional regulation and 5mC. Beyond the newly appreciated dynamics, regulatory roles for 5mC have been suggested in new biological contexts, such as learning and memory or aging. The use of new single-cell measurement techniques and precise editing tools will enable functional analyses of 5mC in gene expression, clarifying its role in various biological processes.

**D**NA methylation (the presence of 5-methylcytosine, 5mC) is one of the best studied epigenetic systems in mammals, and the mechanisms by which this covalent epigenetic mark are written, read, and perpetuated are well characterized (Figs. 1 and 2) (1, 2). Gene regulation by 5mC was initially studied by analyzing CpG islands (CGIs), which are stretches of DNA sequence enriched in CpG dinucleotides that are found at the majority of mammalian promoters (2). The presence of 5mC at a CGI is usually associated with long-term, stable gene repression (2). However, 5mC patterns outside of promoters and during dynamic biological processes were less understood before the development of epigenomic tools. Whole-genome 5mC profiling has uncovered the dynamics of 5mC at enhancers, gene bodies, and extended transcriptionally inactive partially methylated domains (PMDs), indicating regulatory roles for 5mC in development and disease (3–5). Although 5mC is broadly accepted as a repressive epigenetic mark, the mechanism of 5mC-mediated transcriptional repression is still not fully understood (1, 2). Several studies suggest that gene regulation by 5mC may involve functional interactions between 5mC and sequence-specific transcription factors (TFs) (6–9). Here we review recent insights regarding the dynamics of 5mC in development and disease. We further discuss the utility of single-cell methods and new epigenome-editing technologies for further analysis of 5mC function and the discovery of additional DNA base modifications in mammalian genomes.

## Dynamic DNA methylation in development

In the mammalian genome, 5mC is predominantly located within the CpG dinucleotide context. In most somatic tissues, the majority (>80%) of CpGs in the genome are methylated, with the exception of CGIs and other gene regulatory sequences that show reduced 5mC levels (3). Indeed, local 5mC depletion is a reliable signature of promoters and enhancers: CpG sites in promoter-associated CGIs are often less than 10% methylated, whereas distal regulatory sequences such as enhancers are commonly marked by levels of 5mC ranging from 10 to 50% (6). CGI methylation shows few dynamics across healthy tissues. By contrast, aberrant CGI hypermethylation is commonly observed in cancers (2). With respect to methylation patterns outside CGIs, direct comparison of whole-genome methylome profiles derived from numerous cell types has revealed that, under normal physiological conditions, 15 to 21% of CpGs vary in methylation status among human tissues (4, 5). These dynamically methylated CpGs cluster into more than 1 million tissue-specific differentially methylated regions (DMRs). DMRs are predominantly distal to transcriptional start sites and overlap with enhancers. Intragenic DMRs that stretch for kilobases (large DMRs) can overlap with tissue-defining “signature genes” as well as with predicted superenhancers, which are clusters of enhancers that function cooperatively (4, 10, 11). These tissue type-specific DMRs are defined by both lineage-specific gain and loss of 5mC during embryogenesis. A study that profiled 5mC dynamics during mouse late fetal development with daily sampling (12) revealed a temporal uncoupling of the loss and gain of 5mC during the development of all tissues analyzed: Enhancers predominantly lose 5mC during fetal development, whereas most de novo 5mC occurs after birth (Fig. 1) (12). In mouse brain cortex, synchronized de novo 5mC is observed during the second postnatal week, primarily targeting enhancers that are embryonically active (12). Although few analyses have been systematically performed across the lifespan of mammals,

existing data suggest that, overall, the methylome is stable beyond adolescence (13).

Contrary to what is observed in somatic cells, genome-wide changes of DNA methylation patterns occur during preimplantation mammalian development and in the early embryonic germ line (Fig. 1) (14–17). However, these global methylation changes are coordinated with genome-wide changes in histone modifications and reflect a global reprogramming of epigenetic information rather than locus-specific transcriptional regulation. For example, global erasure of DNA methylation in primordial germ cells does not lead to widespread transcriptional activation (16, 17), and numerous genes are expressed during spermatogenesis despite high levels of methylation at their promoters (18).

5mC patterns in differentiated tissues are highly specific and established primarily by site-specific remodeling at regulatory sequences. Sequential waves of gain and loss of 5mC in late fetal to early postnatal development contribute to the majority of 5mC diversity found in differentiated tissues.

## Transcription factors direct local 5mC remodeling

The landscape of 5mC in the mammalian genome is shaped by local depletion of 5mC at regulatory sequences (Fig. 2). The depletion of 5mC at CGIs is, in part, determined by the presence of CpG dinucleotide clusters that are specifically recognized and protected from methylation by proteins containing a CXXC zinc-finger domain (19). These proteins include components of the histone H3 lysine 4 (H3K4) methyltransferase complex (19). Trimethylated H3K4 (H3K4me3), in turn, inhibits the activity of the de novo DNA methyltransferase DNMT3A, thus preventing methylation at CGIs (20). Beyond CXXC domain-containing proteins, the depletion of 5mC at CGIs is also dependent on the binding of other specific TFs (19).

Unlike CGI promoters, enhancers have a CpG density comparable to the genomic average. Tissue- or cell type-specific regulatory sequences can be effectively identified by DMRs and are enriched in recognition sites for TFs with tissue-specific functions (4, 5, 21). We speculate that the large quantity of DMRs (20,000 to 100,000 per tissue type) reflects the collective local demethylation induced by specific TF binding. This model has been supported by experiments in mouse embryonic stem cells (mESCs) that use insulator protein CCCTC-binding factor (CTCF) and a transcriptional repressor of neural genes—RE1-silencing transcription factor (REST) (6). The presence of an intact CTCF motif is sufficient and necessary for local methylation depletion. The deletion of REST led to increased 5mC at regions of low methylation normally bound by REST. Notably, such gain of 5mC can be reversed by reintroducing REST.

It has been postulated that “pioneer” TFs, which can direct cellular reprogramming, are likely to induce local epigenetic remodeling (22). One example is ASCL1, whose forced expression

<sup>1</sup>Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. <sup>2</sup>Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. <sup>3</sup>MRC London Institute of Medical Sciences (LMS), Du Cane Road, W12 0NN London, UK. <sup>4</sup>Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, Du Cane Road, W12 0NN London, UK.

\*These authors contributed equally to this work.

†Corresponding author. Email: ecker@salk.edu (J.R.E.); petra.hajkova@lms.mrc.ac.uk (P.H.)

directly converts fibroblasts to postmitotic neurons and leads to reduced 5mC at most of its binding sites (22, 23). However, given the small number of TFs tested to date, it is unknown whether the ability to remodel 5mC is indeed restricted to TFs with pioneer activity. In addition, the mechanism of TF-induced local 5mC loss is not understood. TFs may directly recruit cytosine demethylation machinery containing ten-eleven translocation methylcytosine dioxygenase (TET, see below) (24) or, alternatively, the demethylation is mediated by intermediate steps of chromatin remodeling or histone modification.

### Transcription factors read DNA methylation

The depletion of 5mC at active regulatory elements suggests that the absence of this repressive mark is essential for active transcription regulation. However, our understanding of the mechanism by which 5mC affects transcription machinery remains incomplete. 5mC readers, including proteins containing a methyl-CpG-binding domain such as MECP2 or MBD proteins, can recognize 5mC and recruit coregulators to methylated DNA (25). In addition, studies suggest that numerous TFs can also directly sense the state of cytosine methylation through the modulation of their binding affinity and specificity (26).

The development of high-throughput assays for analyzing TF-DNA interactions has enabled the identification of many TFs with binding specificity that is altered by the presence of 5mC in recognition sequences (26). A protein microarray study that investigated more than 1500 TFs found more than 40 TFs that can bind methyl-

ated motifs (27). Pull-down experiments of methylated and unmethylated random sequences with a recombinant TF protein followed by sequencing [methyl-SELEX (systematic evolution of ligands by exponential enrichment)] concluded that the binding of about 60% of TFs was influenced by 5mC, and the binding of comparable numbers of TFs was enhanced or inhibited by 5mC (28). These studies unexpectedly uncovered that certain interactions between TFs and low-affinity binding sites can be enhanced by the presence of 5mC. Similarly, DNA affinity purification sequencing (DAP-seq) of more than 500 *Arabidopsis* (plant) TFs determined that >75% of TFs are sensitive to 5mC (29), suggesting that the regulation of TF binding by 5mC is conserved across animal and plant kingdoms.

Although in vitro studies reveal that the majority of TF-DNA interactions can be modulated by 5mC, it is still uncertain how broadly this mechanism contributes to in vivo gene regulatory programs. There are only a handful of reports of 5mC regulating TF binding in a physiological context. In two well-characterized examples, aberrant gain of 5mC at the binding site of insulator protein CTCF is observed in the *Igf2/H19* locus in Beckwith-Wiedemann (fetal overgrowth) syndrome and in the *PDGFRA* locus of isocitrate dehydrogenase mutant gliomas (8, 9). In both cases, hypermethylation of the binding site prevents CTCF binding and results in the disruption of chromatin topological domains, leading to ectopic enhancer-promoter interactions. Intriguingly, 5mC only affects CTCF binding at a specific subset of recognition sequences (30). This selective 5mC sensitivity results from the different biophysical properties of CTCF's 11 zinc fingers—

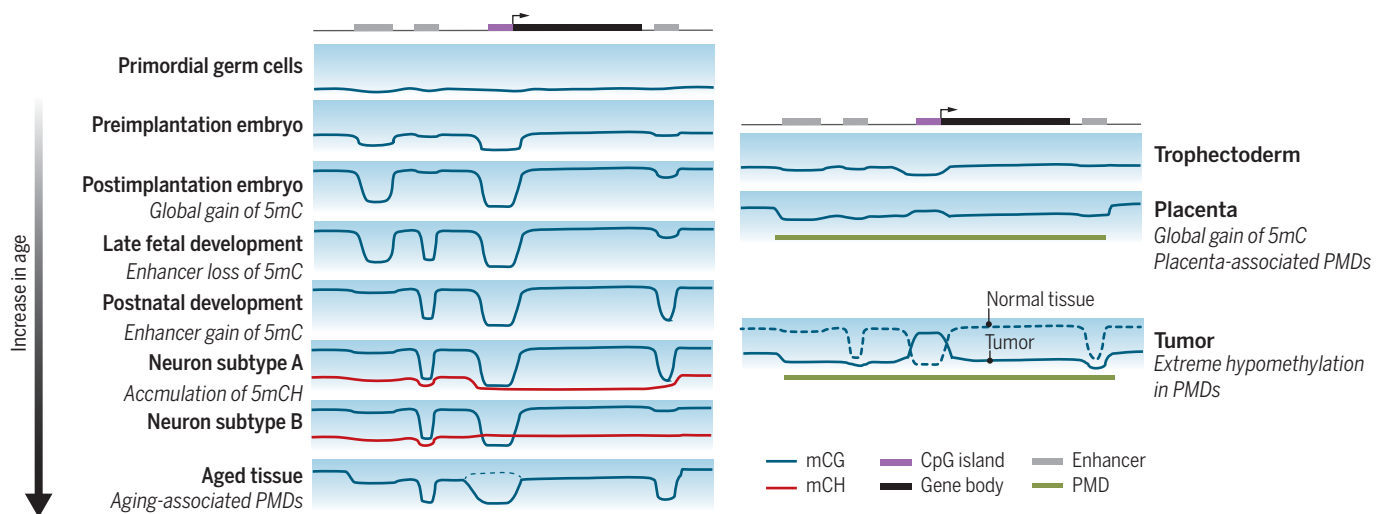
each motif has a different sensitivity for methylation of specific cytosines (31).

We speculate that the in vivo regulation of TF binding by 5mC is more common than what has been reported in the literature. However, such mechanisms may only affect TF function at a subset of sites that show developmental- or disease-related 5mC dynamics and thus might have been previously overlooked. In addition, the bidirectional regulation between TF binding and 5mC presence poses a challenge in determining whether TF binding is the cause or consequence of 5mC remodeling. We anticipate that new epigenetic editing tools for specific manipulation of 5mC (see below) may greatly facilitate functional studies of TF-5mC interactions.

### Intragenic DNA methylation and transcription

Compared with its well-established repressive function at regulatory elements, less is known about 5mC regulation and function(s) at intragenic regions (Fig. 2). Studies with mESCs and a heterologous yeast system found that the establishment of intragenic 5mC is a controlled process carried out by de novo DNA methyltransferase DNMT3B and cotranscriptionally regulated histone modification H3K36me3 (32, 33). Specifically, DNMT3B preferentially binds to H3K36me3 and deposits 5mC within gene bodies.

Although promoter 5mC shows a negative correlation with transcription, early studies of intragenic 5mC function posited the opposite finding (2). Indeed, for certain tissue types, active genes are more methylated than repressed genes. However, this correlation appears to be tissue specific. The interaction between DNMT3B and H3K36me3



**Fig. 1. DNA methylation signatures of developmental stages and tumors.** This figure summarizes 5mC levels measured from heterogeneous collections of cell types, tissues, or organs. Low global 5mC levels are observed in primordial germ cells and preimplantation embryos. A global gain of 5mC occurs during postimplantation development. In late fetal development, local demethylation occurs at regulatory sequences; in postnatal development, a gain of 5mC occurs at embryonically active regulatory sequences, and a loss of 5mC occurs at regulatory sequences active in differentiated tissues. Neuron subtypes A and B show brain-specific accumulation of mCH, the abundance of

which is inversely correlated with transcription activity. In aged tissues, PMDs in late-replication regions resulted from progressive methylation loss; the dashed line indicates that a small percentage of CGIs show hypermethylation in aged tissues (67). The trophoblast has a low global 5mC level similar to that observed in the preimplantation embryo. Placenta and tumors show pervasive PMDs and CGI hypermethylation. The solid and dashed lines in the tumor panel indicate mCG levels in tumor and matching normal tissues, respectively. The color gradient in the background indicates an increase in the level of 5mC, from the bottom to the top of each panel.



leads to greater levels of intragenic 5mC for active genes in mESCs (32). For some tissue types, the positive correlation between 5mC and transcription can be explained by the stratification of genes by chromatin states. At many developmental genes enriched in H3K27me<sub>3</sub>, intragenic 5mC antagonizes polycomb repression and thus is positively correlated with transcription (34). In PMD-associated tissues such as cultured fibroblasts, genes located in PMDs are both expressed and methylated at low levels (3). Many signature genes of human tissues overlap with intragenic large DMRs, suggesting that intragenic 5mC may regulate tissue-specific gene expression (4). In addition, intragenic 5mC has also been postulated to regulate alternative promoter use and alternative splicing and prevent spurious internal transcription initiation (2, 35, 36).

Human embryonic stem cells (hESCs) and mammalian brain tissues contain abundant non-CG methylation (5mCH, where H represents A, C, or T), although the underlying three-base DNA sequence context (5mCAG versus 5mCAC, respectively) is distinct (3, 13). Intragenic 5mCH is positively correlated with gene expression in hESCs but is negatively correlated with gene expression in the brain (13). In neurons, intragenic 5mCH quantitatively modulates gene expression by interacting with MECP2 and recruiting the nuclear receptor corepressor complex to chromatin (37–40). MECP2 contains a methyl-CpG binding domain, and mutations in MECP2 have been linked to Rett syndrome, an autism spectrum disorder (41). Independent studies have reported that MECP2 binds to both 5mCA and 5mCG sites with high affinity (37, 39, 40). CA sites are

the most frequently methylated non-CG sites (13). Repression dependent on both MECP2 and 5mCA is particularly critical for long genes (with a length greater than 100 kb), many of which have neuronal functions (39).

Intragenic 5mC (including 5mCH) performs fine-tuning and supportive roles in transcription regulation but is nevertheless functionally important, as suggested by known links between disease and intragenic 5mCA and MECP2 repression (39).

### Functions of cytosine hydroxymethylation

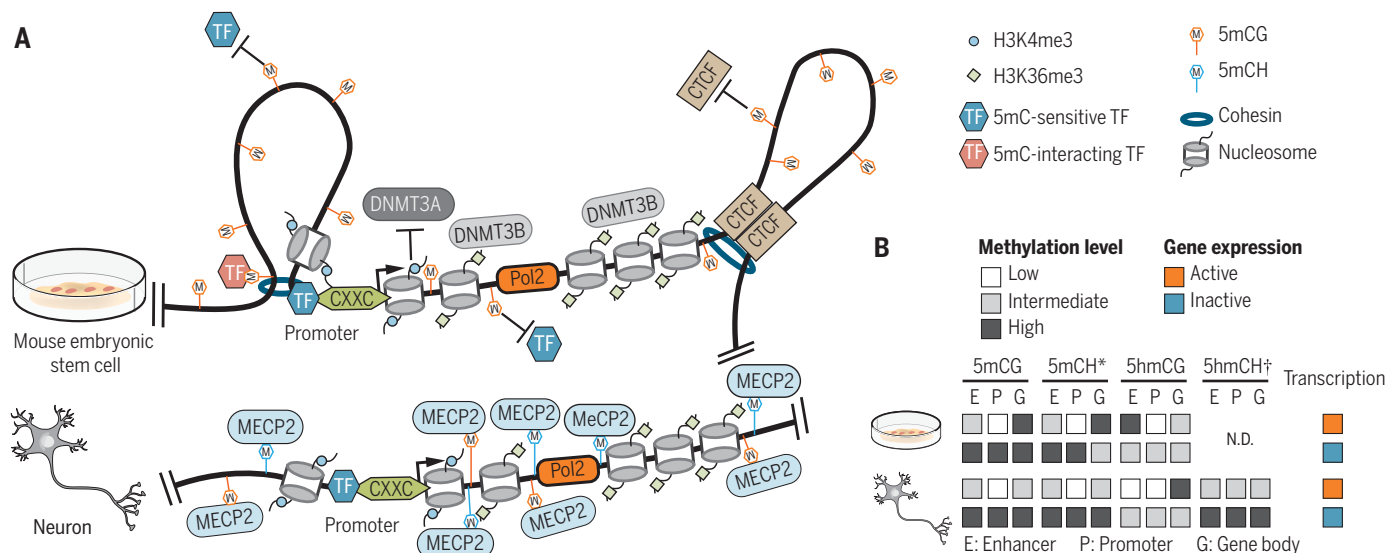
The dynamic changes in 5mC patterns have raised questions about the possible mechanisms of 5mC removal. In this context, 5-hydroxymethylcytosine (5hmC) is the most abundant oxidative derivative of 5mC. 5hmC is generated by the TET family of proteins and is an intermediate product of 5mC demethylation mediated by TET in combination with thymine DNA glycosylase-mediated base excision repair (24). Although TET enzymes are highly expressed during the developmental periods of global DNA demethylation, their role in the observed methylation dynamics seems auxiliary and primarily antagonizes the activity of de novo methylation (17, 42). However, TETs are likely to play important roles in local 5mC dynamics at the regulatory sequences. TET binding is enriched in promoters and enhancers and TET is found in complexes with TFs, suggesting involvement in 5mC dynamics after TF binding (24).

Studies also suggest that the majority of 5hmC can be relatively stable, arguing against 5hmC as being solely a short-lived demethylation intermediate and potentially pointing toward an epigenetic or regulatory role for this modification (43).

This is also supported by the abundance of 5hmC in the brain (up to 20% of the total amount of 5mCG and 5hmCG) (13, 40). Although sequencing-based approaches have only found 5hmC in CG dinucleotides in mESCs and brain cortical tissue (24), 5hmCH accounted for 38% of the total 5mCH and 5hmCH modifications in cerebellar granule cells, suggesting that 5mCH is also targeted for TET-mediated oxidation (40). Interestingly, 5hmCG and 5hmCA (the main context of non-CG methylation) interact with MECP2 differently and lead to opposite gene regulatory outputs. Both 5mCA and 5hmCA can be bound by MECP2 (39, 40), whereas the oxidation of 5mCG to 5hmCG abolishes the MECP2 interaction and disrupts the function of 5mCG as a repressive mark (Fig. 2B).

### Aberrant methylation patterns in diseases and aged tissues

Global changes of methylation patterns are observed in pathology. Tumor methylomes are characterized by global hypomethylation and localized hypermethylation at specific CGIs, many of them associated with tumor suppressor genes (2). Analysis of the whole-genome methylomes of 39 primary tumors in mouse and human shows a pronounced loss of methylation at nuclear lamina-associated PMDs located in late replicating parts of the genome (44). Intriguingly, limited PMD hypomethylation can be detected in healthy tissue and increases with age (44). This loss in DNA methylation observed during aging may be due to the insufficient maintenance of DNA methylation across an increasing number of cell divisions (44), which ties similar methylation changes to both cancer and aging (44, 45). An increase in PMD abundance



**Fig. 2. The function and regulation of DNA methylation in diverse genomic contexts.** (A) Gene promoter CGIs are protected from 5mC by CXXC domain-containing proteins. De novo DNA methyltransferase DNMT3A activity is inhibited by H3K4me<sub>3</sub>. At distal regulatory elements, 5mC can either prohibit the binding of 5mC-sensitive TFs or interact with TFs that show a preference for methylated binding sequence. DNA methylation at CTCF binding sites regulates chromatin conformation by modulating CTCF binding. In mESCs (top), DNMT3B is recruited to

transcribed regions by H3K36me<sub>3</sub> and establishes gene-body methylation. In postmitotic neurons (bottom), 5mCH is bound by MeCP2. Pol2, RNA polymerase II. Nucleosomes were omitted from the bottom panel for simplicity. (B) Transcription outputs are correlated with 5mC and 5hmC levels of enhancers (E), promoters (P), and gene bodies (G). \*5mCH patterns in hESCs are shown; the 5mCH level is substantially lower in mESCs than in hESCs. †To date, a significant level of 5hmCH was only found in cerebellar granule cells. N.D., not detected.

is also found during the early development of placenta along with comparable CGI hypermethylation, both of which are also observed in cancers (46). The similarity between tumor and placenta methylomes suggests that tumorigenesis may represent a latent developmental epigenomic program.

Despite global DNA hypomethylation, a limited number of genomic loci gain DNA methylation during aging and in cancer. Interestingly, these are sequences that are transcriptionally repressed by the polycomb system (and lack DNA methylation) under normal physiological conditions (47). Similar gains of methylation at distinct polycomb-repressed sequences have also been observed, although to a lesser degree, during placental development (46), suggesting an interesting and mechanistically important interplay between DNA methylation and polycomb-regulated epigenetic systems.

In instances in which DNA methylation is globally reduced as a part of a normal developmental program, no aberrant transcriptional activation is observed, suggesting an alternative or compensatory mechanism of gene expression regulation. The situation is, however, different under pathological conditions in which the global loss of DNA methylation is often associated with an aberrant transcriptional outcome.

### Single-cell techniques uncover the diversity of DNA methylation patterns

Developing tissues contain heterogeneous cell populations that are derived from different cell lineages and exhibit various degrees of maturation. Although computational approaches for tissue deconvolution are improving (48), bulk tissue analyses cannot easily distinguish molecular signatures contributed by distinct cell populations. Improvements in single-cell methylome approaches now provide the power to distinguish cellular heterogeneity within a tissue or cell population and identify epigenome fluctuations in a given cell type over time.

Since single-cell reduced representation bisulfite sequencing (scRRBS) and single-cell bisulfite sequencing (scBS-seq) were first reported (49), methodological improvements have increased both throughput and resolution. Datasets containing thousands of single-cell methylomes, such as neuron classification in the human and mouse frontal cortices, have been generated with a well-based nuclei collection method—single-nucleus methylcytosine sequencing (snmC-seq)—or a combinatorial indexing-based method—single-cell combinatorial indexing for methylation analysis (sci-MET) (10, 50). Unlike single-cell RNA sequencing, which captures transcriptomic information, single-nucleus methylome data allowed prediction of cell type-specific regulatory elements with 5mC signatures. Building on single-cell methylome methods, single-cell multi-omics approaches have been developed that can analyze both the transcriptomic and epigenomic profiles from a single cell (49). Studies using these techniques have identified cell-to-cell epigenetic variation in early mouse and human embryos (51, 52), demonstrating the enhancement of resolution enabled by multi-omic single-cell methods.

In addition to sequencing-based single-cell methylome strategies, a fluorescent *in vivo* reporter of genomic methylation enables microscopic examination of the 5mC status at a single genomic locus with cellular resolution (53). In an alternative strategy, the fluorescent protein gene reporter can be directly inserted into the methylation-sensitive locus, which provides a direct readout of the methylation status in living cells or a whole animal (54). Although single-cell methods are just starting to be applied to the analysis of 5mC functions, studies have already identified 5mC heterogeneity both across or within cell types, providing information that cannot be accessed with bulk-level analyses.

### Toward site-specific manipulation of DNA methylation

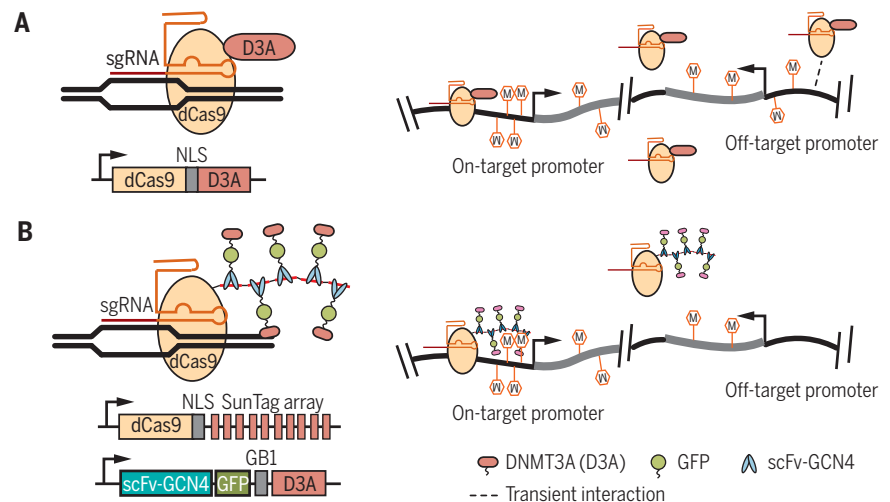
Functional analyses of 5mC have traditionally relied on pharmacological or genetic perturbations that affect 5mC levels genome-wide (30, 32). Although discoveries regarding 5mC function and regulation have been made with genome-wide perturbation strategies, these methods are associated with pleiotropic phenotypes and inevitably confound the interpretation of results (35). Modern genome-engineering tools have opened up the possibility for a more precise site-specific intervention that would potentially avoid pleiotropic effects of global 5mC disruption (55).

Strategies for targeting 5mC to specific genomic locations have used different programmable DNA binding domains, including zinc fingers, transcription activator–like effectors and cleavage-deficient dCas9. dCas9-TET and dCas9-DNMT3A fusions have been shown to induce effective remodeling of 5mC (Fig. 3) (56). Most studies take a

candidate approach to estimate the extent of off-target manipulation. However, off-target *de novo* methylation activity can be widespread across the entire genome (57) and is not easily detected with steady-state 5mC measurements. Local amplification of regulator concentration through protein (SunTag) or RNA scaffolding (e.g., Casilio) has been incorporated into epigenetic editing to enhance the on-target/off-target ratio (Fig. 3). This scaffold allows multiple copies of epigenetic effectors (e.g., DNMT3A catalytic domain) to be tethered to a genomic site, providing more robust on-target modulation and reduced non-specific effects (58).

### Beyond cytosine base modifications: 6mA, the new kid on the block

5mC and its oxidation derivatives had been considered the only type of mammalian DNA modification until reports of the presence of *N*<sup>6</sup>-methyladenine (6mA) in DNA of higher eukaryotes (59). Compelling multimodal evidence has been presented for the existence of 6mA in mouse and human genomes. These findings are supported by the identification of both methyltransferase and demethylase for deposition and removal of 6mA, respectively (60, 61). Interestingly, the genome-wide distribution of 6mA is species-specific and lacks a conserved pattern across the eukaryotic species analyzed to date (59). Reported differences in 6mA distributions between mouse and human may represent tissue-specific patterns rather than species differences, because the tissues used for the mouse and human studies were unique to each species [mESCs and blood DNA, respectively (60, 61)]. Functional analysis of 6mA in mammalian genomes may be most productive



**Fig. 3. Comparison of direct fusion- and SunTag protein scaffold-based 5mC editing tools.**

(A) Direct fusion-based editing. (B) The SunTag system allows local amplification of effector (DNMT3A) concentration and independent control of dCas9 and DNMT3A expression levels to minimize off-target *de novo* methylation caused by untethered DNMT3A. The SunTag-based 5mC editing tool (dCas9Sun-D3A) shows comparable levels of on-target methylation at on-target promoters as the direct fusion-based editing tool (dCas9-D3A), but much lower off-target methylation levels. sgRNA, single-guide RNA; NLS, nuclear localization sequence; GB1, solubility tag (protein G B1 domain); scFv-GCN4, single-chain variable fragment antibody against GCN4. The schematics in (A) and (B) are based on 5mC editing tools developed by Liu *et al.* (56) and Pflueger *et al.* (58), respectively.



when applied to tissue with abundant 6mA, to be determined with surveys of 6mA across tissues and developmental stages.

## Outlook

Recent studies have indicated intriguing, possibly regulatory, roles for 5mC dynamics in diverse biological processes and disease states. Animal models, such as mouse fear conditioning, suggest a role for 5mC in regulating learning and memory (62). Epigenome-wide association studies have linked 5mC signatures to many diseases (63). However, many studies use correlative analysis and report only moderate 5mC differences detected from complex tissues. Thus, although interesting, these studies should be considered as exploratory in nature. In addition, meta-analysis of 5mC microarray data has identified a small number (~300 to 400) of CpG sites that are associated with age across multiple tissues, suggesting the existence of a methylome-based clock (45). Further work is needed to link these clock methylation variants with other molecular information (e.g., gene expression) and aging-related phenotypes.

In mitotically inactive mammalian tissues, moderate changes in 5mC dynamics likely reflect cellular heterogeneity either across cell types or within a cell type. In this context, we anticipate that single-cell multi-omic strategies will enhance the specificity and sensitivity of the analysis of 5mC patterns, allowing the determination of cellular identity in tissues (cell atlas) and measurement of alterations pertinent to experimental and environmental perturbations. Effectively, the 5mC dynamics of all cell types in the tissue can be reconstructed with single-cell profiling without cell-type labeling or enrichment strategies, contributing to organism-scale cell-type reference atlas initiatives as well as profiling of disease tissues at a cell-type resolution. (64).

Site-specific methylome editing will serve as a key technique for the study of 5mC function and should pave the way for the development of epigenetic-based strategies for cellular engineering and therapies. Specific editing tools will be crucial for moving beyond correlation, allowing separation of cause from consequence. An important use for methylome editing will be to

aid understanding of the impact of 5mC on transcription and its relationship to other types of epigenetic modifications—a subject actively debated. For example, a recent methylation-editing study demonstrated that deposition of 5mC at thousands of gene promoters led to only modest transcriptional alterations (65, 66). With the availability of facile methods for single-cell methylome profiling and editing tools that allow manipulation of 5mC in cells at the right place and time, we anticipate that, in the coming decade, discoveries of new biological functions (or the debunking of proposed ones) for DNA methylation will proceed at an unprecedented pace.

## REFERENCES AND NOTES

1. D. Schübeler, *Nature* **517**, 321–326 (2015).
2. P. A. Jones, *Nat. Rev. Genet.* **13**, 484–492 (2012).
3. R. Lister et al., *Nature* **462**, 315–322 (2009).
4. M. D. Schultz et al., *Nature* **523**, 212–216 (2015).
5. M. J. Ziller et al., *Nature* **500**, 477–481 (2013).
6. M. B. Stadler et al., *Nature* **480**, 490–495 (2011).
7. S. Domcke et al., *Nature* **528**, 575–579 (2015).
8. A. C. Bell, G. Felsenfeld, *Nature* **405**, 482–485 (2000).
9. W. A. Flavahan et al., *Nature* **529**, 110–114 (2016).
10. C. Luo et al., *Science* **357**, 600–604 (2017).
11. W. A. Whyte et al., *Cell* **153**, 307–319 (2013).
12. Y. He, M. Hariharan, D. U. Gorkin, D. E. Dickel, C. Luo, R. G. Castanon, J. R. Nery, A. Y. Lee, B. A. Williams, D. Trout, H. Amrhein, R. Fang, H. Chen, B. Li, A. Visel, L. Pennacchio, B. Ren, J. Ecker, bioRxiv 166744 [Preprint]. 21 July 2017. <https://doi.org/10.1101/166744>.
13. R. Lister et al., *Science* **341**, 1237905 (2013).
14. Z. D. Smith et al., *Nature* **484**, 339–344 (2012).
15. Z. D. Smith et al., *Nature* **511**, 611–615 (2014).
16. S. Seisenberger et al., *Mol. Cell* **48**, 849–862 (2012).
17. P. W. S. Hill et al., *Nature* **555**, 392–396 (2018).
18. S. S. Hammoud et al., *Cell Stem Cell* **15**, 239–253 (2014).
19. Z. D. Smith, A. Meissner, *Nat. Rev. Genet.* **14**, 204–220 (2013).
20. X. Guo et al., *Nature* **517**, 640–644 (2015).
21. G. C. Hon et al., *Nat. Genet.* **45**, 1198–1206 (2013).
22. O. L. Wapinski et al., *Cell* **155**, 621–635 (2013).
23. C. Luo, Y. L. Qian, O. L. Wapinski, R. Castanon, J. R. Nery, S. M. Cullen, M. A. Goodell, H. Chang, M. Wernig, J. R. Ecker, bioRxiv 371427 [Preprint]. 18 July 2018. <https://doi.org/10.1101/371427>.
24. H. Wu, Y. Zhang, *Cell* **156**, 45–68 (2014).
25. Q. Du, P.-L. Luu, C. Stirzaker, S. J. Clark, *Epigenomics* **7**, 1051–1073 (2015).
26. H. Zhu, G. Wang, J. Qian, *Nat. Rev. Genet.* **17**, 551–565 (2016).
27. S. Hu et al., *eLife* **2**, e00726 (2013).
28. Y. Yin et al., *Science* **356**, eaaj2239 (2017).
29. R. C. O'Malley et al., *Cell* **165**, 1280–1292 (2016).
30. M. T. Maurano et al., *Cell Reports* **12**, 1184–1195 (2015).
31. H. Hashimoto et al., *Mol. Cell* **66**, 711–720.e3 (2017).
32. T. Baubec et al., *Nature* **520**, 243–247 (2015).
33. M. Morselli et al., *eLife* **4**, e06205 (2015).
34. H. Wu et al., *Science* **329**, 444–448 (2010).
35. F. Neri et al., *Nature* **543**, 72–77 (2017).
36. A. Teissandier, D. Bourchis, *EMBO J.* **36**, 1471–1473 (2017).
37. J. U. Guo et al., *Nat. Neurosci.* **17**, 215–222 (2014).
38. L. Chen et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5509–5514 (2015).
39. H. W. Gabel et al., *Nature* **522**, 89–93 (2015).
40. M. Mellén, P. Ayata, N. Heintz, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7812–E7821 (2017).
41. J. Guy, H. Cheval, J. Selfridge, A. Bird, *Annu. Rev. Cell Dev. Biol.* **27**, 631–652 (2011).
42. R. Amouroux et al., *Nat. Cell Biol.* **18**, 225–233 (2016).
43. M. Bachman et al., *Nat. Chem.* **6**, 1049–1055 (2014).
44. W. Zhou et al., *Nat. Genet.* **50**, 591–602 (2018).
45. S. Horvath, K. Raj, *Nat. Rev. Genet.* **19**, 371–384 (2018).
46. Z. D. Smith et al., *Nature* **549**, 543–547 (2017).
47. Y. Schlesinger et al., *Nat. Genet.* **39**, 232–236 (2007).
48. A. E. Jaffe, R. A. Irizarry, *Genome Biol.* **15**, R31 (2014).
49. G. Kelsey, O. Stegle, W. Reik, *Science* **358**, 69–75 (2017).
50. R. M. Mulqueen et al., *Nat. Biotechnol.* **36**, 428–431 (2018).
51. L. Li et al., *Nat. Cell Biol.* **20**, 847–858 (2018).
52. S. Rulands et al., *Cell Syst.* **7**, 63–76.e12 (2018).
53. Y. Stelzer, C. S. Shivavilla, F. Soldner, S. Markoulaki, R. Jaenisch, *Cell* **163**, 218–229 (2015).
54. M. Van de Pette et al., *Cell Reports* **18**, 1090–1099 (2017).
55. J. Pulecio, N. Verma, E. Mejía-Ramírez, D. Huangfu, A. Raya, *Cell Stem Cell* **21**, 431–447 (2017).
56. X. S. Liu et al., *Cell* **167**, 233–247.e17 (2016).
57. C. Galonska et al., *Nat. Commun.* **9**, 597 (2018).
58. C. Pflueger et al., *Genome Res.* **28**, 1193–1206 (2018).
59. G.-Z. Luo, C. He, *Nat. Struct. Mol. Biol.* **24**, 503–506 (2017).
60. C.-L. Xiao et al., *Mol. Cell* **71**, 306–318.e7 (2018).
61. T. P. Wu et al., *Nature* **532**, 329–333 (2016).
62. J. J. Day, A. J. Kennedy, J. D. Sweatt, *Annu. Rev. Pharmacol. Toxicol.* **55**, 591–611 (2015).
63. D. S. Paul, S. Beck, *Trends Mol. Med.* **20**, 541–543 (2014).
64. A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, *eLife* **6**, e27041 (2017).
65. E. E. Ford, M. R. Grimmer, S. Stolzenburg, O. Bogdanovic, A. de Mendoza, P. J. Farnham, P. Blancfort, R. Lister, bioRxiv 170506 [Preprint]. 17 August 2017. <https://doi.org/10.1101/170506>.
66. K. Korthauer, R. A. Irizarry, bioRxiv 381145 [Preprint]. 1 August 2018. <https://doi.org/10.1101/381145>.
67. M. Berdasco, M. Esteller, *Aging Cell* **11**, 181–186 (2012).

## ACKNOWLEDGMENTS

We apologize to those colleagues whose work we could not cite owing to space limitations. We thank B. A. Wang, J. A. Law, and J. R. Dixon (Salk Institute); S. C. Huang (New York University); and B. Ren (University of California, San Diego) for insightful discussions. Work in the Hajkova lab is supported by MRC funding (MC\_US\_A652\_5PY70) and by an ERC grant (ERC-CoG-648879—dynamic modifications). Research in the Ecker lab is supported by NIH awards 5R21HG009274, 5U24DK112348, 5R01ES025585, 5R21MH112161, 5R01MH112763, and 1U19MH114831. Additional support to J.R.E. is provided by the Office of Naval Research N00014-16-1-3159 and the Cure Alzheimer's Fund. J.R.E. is an investigator of the Howard Hughes Medical Institute. **Competing interests:** The authors declare no competing interests.

10.1126/science.aaf6806

## REVIEW

# Developmental enhancers and chromosome topology

Eileen E. M. Furlong<sup>1\*</sup> and Michael Levine<sup>2,3\*</sup>

Developmental enhancers mediate on/off patterns of gene expression in specific cell types at particular stages during metazoan embryogenesis. They typically integrate multiple signals and regulatory determinants to achieve precise spatiotemporal expression. Such enhancers can map quite far—one megabase or more—from the genes they regulate. How remote enhancers relay regulatory information to their target promoters is one of the central mysteries of genome organization and function. A variety of contrasting mechanisms have been proposed over the years, including enhancer tracking, linking, looping, and mobilization to transcription factories. We argue that extreme versions of these mechanisms cannot account for the transcriptional dynamics and precision seen in living cells, tissues, and embryos. We describe emerging evidence for dynamic three-dimensional hubs that combine different elements of the classical models.

**T**ranscriptional enhancers are short segments of DNA that activate gene expression in an orientation-independent manner in response to intrinsic and external signals. A typical human protein-coding gene contains multiple enhancers, each bound by a specific combination of sequence-specific transcription factors (TFs). These factors can activate appropriate target genes over long distances, and sometimes even across chromosomes. In this review, we discuss insights provided by emerging technologies in imaging and genomics. These methods are beginning to provide elegant visualization of enhancer-promoter communication and are transforming our understanding about how enhancers work.

## Enhancers work within the context of chromatin domains

Developmental enhancers mediate localized patterns of gene expression in space and time. For example, the ZRS enhancer regulates expression of the *Sonic hedgehog* (*Shh*) gene within a specific region (the ZPA) of developing limb buds in vertebrate embryos (1). Dominant mutations in the ZRS cause familial forms of polydactyly—individuals with supernumerary toes or fingers. The ZRS maps approximately one megabase from the *Shh* transcription start site, but the basis for such long-range gene control remains unknown (1). Even in the compact *Drosophila* genome, there are examples of enhancers working over distances of 70 to 100 kb, including the *cut* wing margin enhancer and the *svb* epidermal enhancer (2). These examples are likely not exceptional—recent large-scale approaches suggest that 20 to 30% of *Drosophila* enhancers may act as distal elements, skipping intervening genes (3–5). How-

ever, the average enhancer is not so remote, typically mapping ~20 to 50 kb in vertebrates and 4 to 10 kb in *Drosophila* from their target genes. Even at these distances, it seems unlikely that TFs directly touch the RNA polymerase II (Pol II) machinery at core promoters without being brought into physical proximity. How is this achieved?

Data from both imaging and Hi-C (genome-wide chromosome conformation capture assays) indicate that metazoan genomes are organized into a series of topological associating domains

## “How remote enhancers relay regulatory information to their target promoters is one of the central mysteries of genome organization and function.”

[TADs; discussed in detail in (6, 7)], which bring distant cis-regulatory elements into proximity, such as enhancers and promoters. The boundaries of TADs are often delineated by clusters of binding motifs for insulator proteins, such as CTCF, and the promoters of actively transcribed genes, such as tRNAs (8–10). How TADs are formed and affect gene expression remains unclear. In vertebrates, loop extrusion is the current prevailing model for TAD formation (11). This involves the loading of one or more cohesin complexes that form tripartite rings around chromatin and actively extrude a chromatin loop from one direction until they reach a barrier that blocks their activity. CTCF is likely the main barrier protein in vertebrates, although its binding is quite dynamic, with residence times of 1 to 2 min (12). In flies, embryos lacking CTCF develop normally (13), suggesting that either other barrier proteins exist, or that TADs can be formed by other mechanisms.

The yeast condensin complex can also extrude DNA loops at the remarkable rate of ~1.5 kb per second in vitro (14). Although in vivo rates have yet to be determined, this suggests that even the remote ZRS enhancer could be brought into proximity with the *Shh* promoter in about 10 min. As discussed below, loop extrusion incorporates features of two classical models of enhancer activity: looping (e.g., TAD loop domains) and tracking (e.g., convergent movements of cohesin complexes loaded at flanking insulators).

## TADs impose regulatory constraints on developmental enhancers

How TADs and TAD boundaries impinge on enhancer function remains an open question. There is considerable evidence that TAD boundaries act as insulators to preclude inappropriate enhancer-promoter interactions. An enhancer located in one TAD preferentially interacts with “local” promoters rather than those located in neighboring TADs (15). Compelling evidence comes from genetic studies in mice that removed TAD boundaries (16) or CTCF binding sites within boundaries (17, 18), or created chromosomal inversions or duplications that fuse adjacent TADs (19). These manipulations can cause developmental disorders as a result of inappropriate enhancer-promoter interactions (16, 19). Such a scenario is seen for a rare genetic disorder affecting human limb development (16). A large TAD containing the *Epha4* gene is flanked by two smaller TADs containing *Wnt6* and *Ihh* genes on one side and *Pax3* on the other. Chromosomal rearrangements that disrupt the TAD boundaries result in new interactions of *Wnt6*, *Ihh*, and *Pax3* with enhancers located in the *Epha4* TAD (16).

Depletion of CTCF (20) or cohesin (21, 22) leads to a dramatic diminishment of TAD structures in vertebrates; however, this has only modest effects on gene expression. Hundreds—not thousands—of genes are affected, and less than half exhibit elevated expression, suggesting spurious gains in enhancer-promoter interactions in the absence of a boundary (21). It therefore appears that TADs constrain the action of just a subset of enhancers. We note, however, that it is possible that many genes may display less precise spatial or temporal expression in the absence of TADs when measured with quantitative single-cell methods.

## Promoter specificity, distance, and competition influence enhancer interactions

As discussed above, diminishment of TAD organization in trans, via depletion of CTCF or cohesin, leads to relatively mild changes in gene expression. Yet, the removal of individual TAD boundaries in cis can cause dramatic phenotypes. What is the basis for this apparent discrepancy? The latter might arise from the occurrence of promiscuous enhancers located near TAD boundaries (Fig. 1A). Once a boundary is removed, there is a good probability for activation of neighboring genes. Indeed, genetic studies indicate that some enhancers can interact with any promoter in their

<sup>1</sup>European Molecular Biology Laboratory (EMBL), Genome Biology Unit, D-69117, Heidelberg, Germany. <sup>2</sup>Lewis-Sigler Institute for Integrative Genomics, Princeton, NJ 08540, USA.

<sup>3</sup>Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA.

\*Corresponding author. Email furlong@embl.de (E.E.M.F.); msl2@princeton.edu (M.L.)



vicinity (15), whereas others have an inherent preference for particular promoter types, e.g., TATA or DPE (23) (Fig. 1B). For example, a shared enhancer in the mouse *Mrf4-Myf5* locus differentially regulates linked *Mrf4* and *Mrf5* promoters in intercalated myotome and fetal muscle fibers, respectively (24). This enhancer mediates both patterns when linked to a heterologous promoter, suggesting promoter specificity within the endogenous locus. Thus, it appears that some enhancers have the capacity to interact with many sequences throughout an entire TAD, but productively engage just a subset of available promoters. How is this specificity achieved?

One potential mechanism is promoter competition, whereby shared enhancers are sequestered by the closer or stronger promoter when presented with multiple choices (Fig. 1C). Competition has been documented in a variety of developmental processes. A notable recent example concerns the proto-oncogene *c-Myc* (25). *c-Myc* overexpression is a major cause of many human cancers. Expression is normally attenuated by *PVT1*, a long noncoding RNA located just downstream. The *PVT1* promoter is located between the *c-Myc* promoter and a number of remote 3' enhancers that regulate expression in different tissues, including lymphocytes. The proximal *PVT1* promoter sequesters these shared enhancers to attenuate *c-Myc* expression.

Proximal promoters generally have a competitive advantage over more distal promoters, but this can be negated by insulators and TAD boundaries. The higher-order topologies present in TADs increase the proximity, and therefore probability, of long-range enhancer-promoter interactions. For example, both the ZRS enhancer and distant *Shh* promoter are located within a common TAD, which reduces the effective distance separating them. When the TAD structure is altered by a genetic inversion, ZRS-*Shh* interactions are diminished (26). However, *Shh* expression was restored by reducing the genomic distance separating the enhancer and promoter within the inversion chromosome. Thus, promoter specificity, distance, and competition all influence the functional consequences of disrupting TAD boundaries (Fig. 1).

### Models of enhancer-promoter communication

Even with the contraction in space afforded by TADs, the distances between enhancers and target promoters can be large, as seen for ZRS-*Shh* interactions. How can enhancers convey regulatory information across such distances? Various models for enhancer-promoter communication have been proposed over the years (Fig. 2), including tracking (or scanning), linking (or chaining), looping,

and mobilization to transcription factories. Although the field has largely converged on looping, it is interesting to consider other models in light of genome topologies such as TADs.

The tracking model proposed that Pol II bound to upstream regulatory elements could move along DNA, pulling the enhancer with it until coming into contact with a proximal promoter. The motor force of Pol II elongation was seen as the key mediator of this enhancer mobilization (Fig. 2A). Proposed evidence for tracking came from inserting "road blocks" (insulators) between enhancers and promoters. Long non-coding RNAs (lncRNAs) emanating from enhancers directed toward their target promoters were suggested to be a manifestation of this process within the  $\beta$ -globin locus (27). It is conceivable that loop extrusion—cohesin-driven spooling of DNA—could foster tracking of distant enhancers.

The linking model is an extension of observations seen for the Lambda repressor, whereby protein-protein oligomers bridge distal elements and target promoters (Fig. 2B). One proposed linking factor in metazoans is the *Drosophila* Chip protein (28), which was proposed to oligomerize from the enhancer to the promoter. However, more recent studies with the vertebrate ortholog, Lbd1, suggests that it might actually form targeted loops through homodimerization when bound at enhancers and promoters (29).

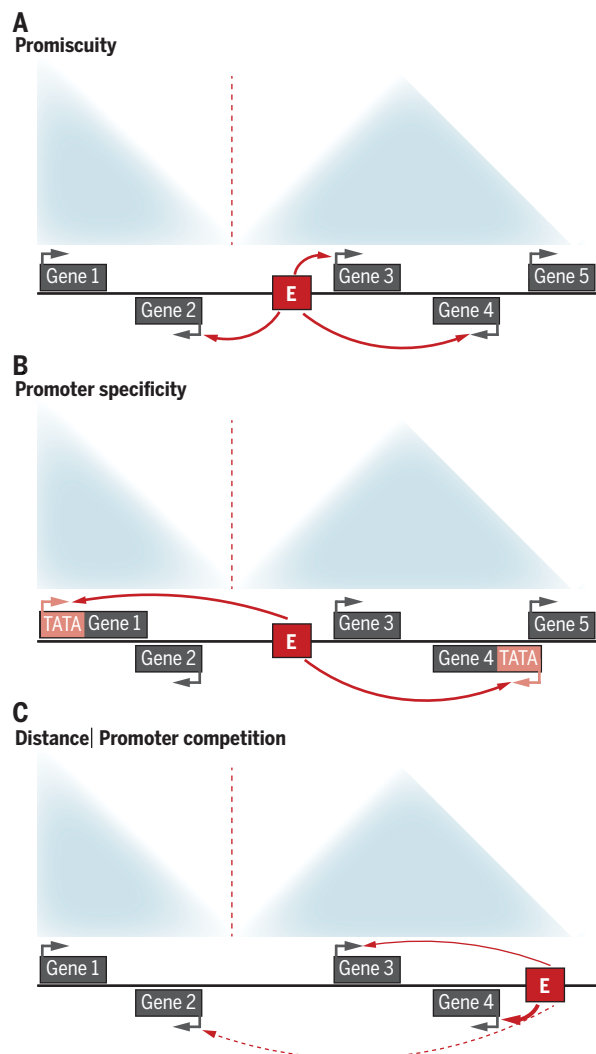
The looping model proposes that factors bound to two distinct sites would physically interact with each other, resulting in extrusion of the intervening DNA (Fig. 2C). As TF binding is dynamic, so too is loop formation. In this mechanism, the intervening DNA is passive during the formation of loops, in contrast to tracking or linking models where it could play an active

role. The first evidence for looping came from *Escherichia coli* over 30 years ago (30), where optimal interactions required proteins to be located on the same side of the helix (Fig. 2C). The common feature of looping and tracking is the reliance on an adenosine 5'-triphosphate (ATP)-driven motor that can move along DNA, e.g., cohesin and/or condensin complexes in the case of loops and TADs, and Pol II elongation for tracking.

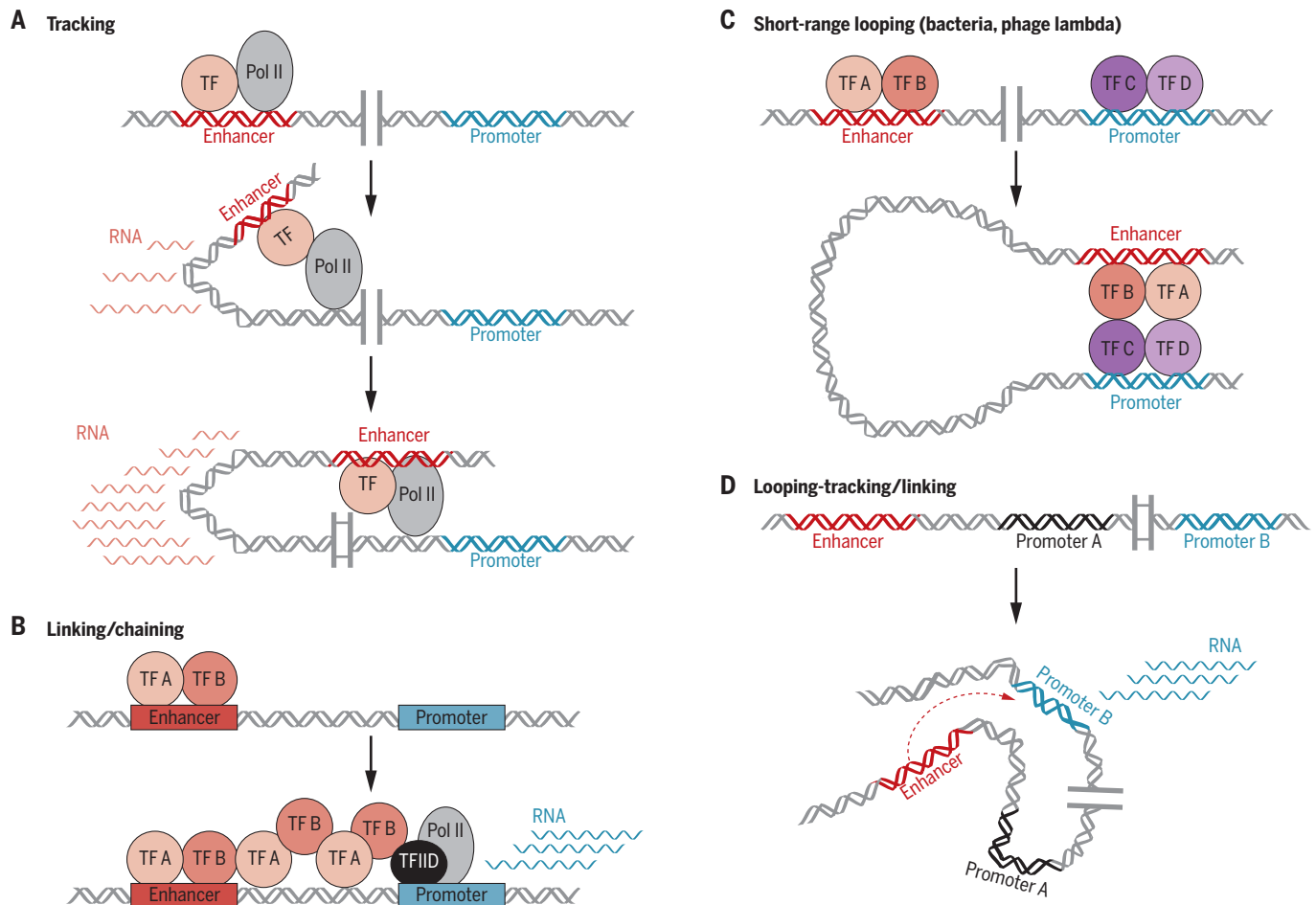
### Mechanisms of loop formation

In vertebrates, a typical TAD contains several transcription units and dozens of enhancers. How do the right enhancers interact with the right promoters within the context of TADs? Sequence-specific binding of CTCF and the recruitment of cohesin can contribute to targeted enhancer-promoter interactions. The first evidence for this came from a genetic screen in *Drosophila*, which identified mutants that perturb long-range activation of *cut* and *Ubx* (31). This identified Nipped-B (NIPBL in vertebrates), a protein required for cohesin loading. CTCF-cohesin complexes have since been implicated in the formation of long-range loops in many contexts, acting positively or negatively to regulate enhancer-promoter interactions (32, 33) (Fig. 2C). However, there are also many examples where enhancers skip over CTCF bound-regions to regulate specific target genes (8). Thus, CTCF can sometimes function as a barrier or insulator, forming loops that block inappropriate enhancer-promoter interactions, but sometimes not.

One explanation for this apparent specificity is an orientation dependence of CTCF binding sites. The formation of loops is facilitated by a convergent arrangement of such sites (10). The first hints of this directionally came from the 5' boundary element in the chicken



**Fig. 1. Properties of enhancers influence the impact of TAD boundaries.** After boundary deletion (dashed lines), enhancers (E) have different abilities to regulate promoters in neighboring TADs depending on their (A) promiscuity, (B) promoter specificity, (C) distance to other promoters.



**Fig. 2. Models of enhancer-promoter communication.** (A) Pol II binds to an enhancer and tracks along chromatin (synthesizing RNA), pulling the enhancer with it. (B) TFs bound to a regulatory element oligomerize, chaining to the promoter. (C) Looping (in bacteria, lambda) requires protein-protein interactions between factors on the same face of the helix. (D) Long-range loops can bring enhancers close to a promoter, but not in direct proximity. Tracking or linking could bridge the distance.

$\beta$ -globin locus and the ICR in the mouse *H19-Igf2* locus, both of which depend on CTCF and function in an orientation-dependent manner (34). Inverting the orientation of specific CTCF sites can alter enhancer-promoter loops, causing changes in gene expression (18, 35). This orientation-dependent requirement for CTCF interactions is reminiscent of the requirement of helical phasing in bacterial, bacteriophage, and SV40 elements. In principle, CTCF orientation can help ensure that the correct enhancers loop to the right promoters, but such a mechanism cannot account for all aspects of enhancer specificity within TADs.

CTCF and cohesin are expressed in most or all tissues and are generally involved in constitutive looping interactions (9, 32). However, cohesin can partner with other factors to mediate tissue-specific enhancer-promoter proximity. *Sox2* expression in embryonic stem (ES) cells, for example, depends on tissue-specific interactions between its promoter and a distal enhancer that is mediated by cohesin-Mediator complexes (9). However, the relatively mild changes in gene expression resulting from CTCF and cohesin depletion sug-

gests that additional proteins also mediate dynamic enhancer-promoter loops within TADs. YY1 is one example (36), which is broadly, if not ubiquitously expressed. Tissue-specific enhancer-promoter loops generally depend on tissue-specific TFs, such as GATA1 and Klf1 in erythrocytes (37, 38). At the  $\beta$ -globin locus, GATA1 mediates promoter loops independently of cohesin by interacting with Lbd1 (38).

Given the density of enhancers, insulators, and TF binding sites throughout the genome, it is difficult to envisage how tracking or linking models could produce specific activation of the appropriate target gene. Enhancers often skip genes to interact with preferred target promoters, and many enhancers (~50% in *Drosophila*) are located in introns or 3' regulatory regions. In such cases, enhancer tracking or oligomerization of linking proteins would likely occlude the transcription of associated coding sequence. Strict versions of the tracking model also fail to explain transvection—gene activation across paired chromosomes. In principle, tracking or linking mechanisms could work over short distances, particularly when considered in the con-

text of TADs and specific loops within TADs (Fig. 2D). This would bring enhancers in close proximity with other enhancers and promoters to generate a three-dimensional “hub” for the propagation of short-range signals. In such a scenario, the enhancer need not physically touch the promoter, but rather, physical proximity may be sufficient for the regulation of promoter activity (Fig. 2D, see below).

### Are loops sufficient for gene activation?

The emerging picture from both chromatin-capture and imaging techniques is that multiple enhancers and promoters are organized in complex nonbinary topologies (4, 8, 10). This view clashes with the simplest models of TF-directed looping of a specific enhancer to a specific promoter. Here, we consider different types of enhancer-promoter topologies and their roles in transcription activation.

Enhancer-promoter proximity is globally correlated with gene activity—for example, active enhancers are generally found near active promoters (and other active enhancers) (4, 8, 10, 39). However, proximity appears to be sufficient for



activating some genes but not others (Fig. 3A). The locus control region (LCR), for example, forms contacts with the globin promoters in erythroid cells where the genes are active, but not in the brain where they are inactive (40). Moreover, forcing a loop between the LCR enhancer and  $\beta$ -globin promoter is sufficient to activate gene expression (29), indicating that proximity acts as a trigger. By contrast, enhancers at other loci are already in proximity with their target promoter prior to gene expression (4, 41–43). For example, the T helper type 2 locus contains three genes

(interleukin-3, -4, and -5) that are coordinately expressed in a subset of T lymphocytes. The type 2 LCR comes into proximity with the *IL-3*, -4, and -5 promoters during the specification of different T cell lineages, prior to their activation (41). Only at later stages do the genes become activated, but the LCR-promoter topology remains unchanged. Similarly, a study of 100 loci during *Drosophila* development indicates that many embryonic enhancers are in preformed topologies with their target promoters prior to gene activation (4). Most of these genes contain promoter-proximal

paused Pol II (4), suggesting that they are primed for rapid induction.

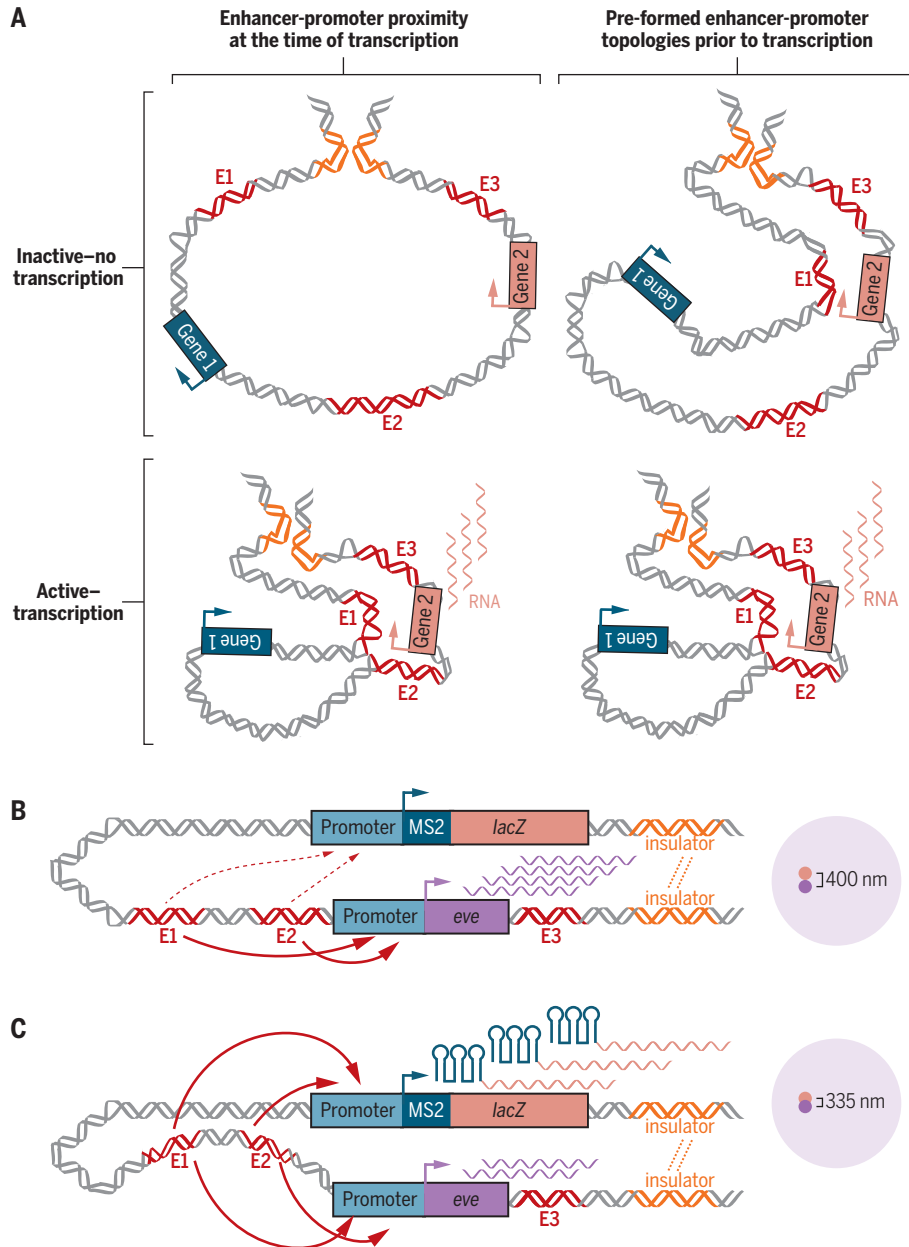
What triggers activation? In some cases, this may be due to subtle, dynamic movements of an enhancer or promoter within a preformed topology (4, 42). At other loci, the recruitment of a TF to a prelooped enhancer may trigger activation, as seen at the T helper type 2 locus in mammals (41). These preformed loops, or “hubs,” might be assembled by TFs present at earlier developmental stages preceding activation. GATA3 and STAT6 are thought to perform this role in the T helper type 2 locus.

We suggest that gene activation is a two-step process at many loci: Enhancers and promoters come into proximity (local compaction) to prime expression, and then subtle topological changes trigger activation (Fig. 3). Recent live imaging supports this; the *Drosophila eve* locus contains a series of 5' and 3' enhancers located close to the promoter that mediate expression in segmentation stripes along the embryo. A *lacZ* reporter gene containing MS2 RNA stem loops was linked to the *eve* promoter and Homie insulator and positioned ~140 kb from the endogenous *eve* locus (44). In the absence of looping (e.g., insulator-insulator interactions), the *lacZ* reporter gene is located ~700 nm from the endogenous locus. This distance is reduced to less than 400 nm upon insulator pairing (Fig. 3B). However, this proximity is not sufficient for activation in the majority of cells. Active foci of transcription exhibit an even tighter association; ~330 nm rather than ~400 nm (Fig. 3C). It therefore appears that looping is necessary but not sufficient for transcriptional activation of the reporter gene. Similar results were obtained in a transvection assay across homologous chromosomes. Only half of the paired alleles display expression (45).

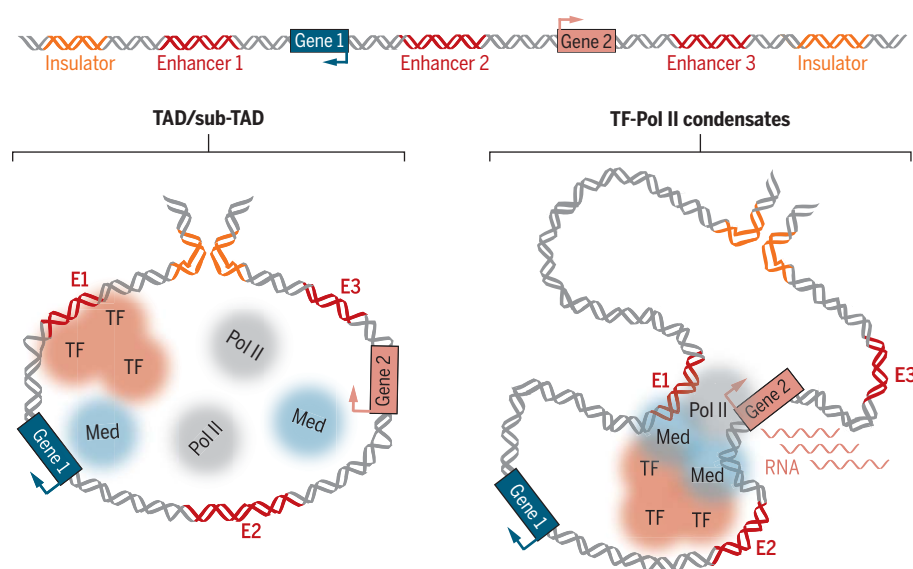
These observations are reminiscent of the *Shh* locus—the long-range ZRS enhancer and *Shh* promoter are contained within a common looped topology throughout the limb bud. Super-resolution microscopy indicates that the enhancer and promoter are in even closer proximity in the posterior limb where *Shh* is active, but not in anterior regions where it is inactive (46). In principle, preformed topologies should reduce the search time and space of enhancer-promoter interactions. Spatial confinement, for example, is the major parameter that determines interaction frequency between distal V(D)J regions comprising immunoglobulin genes in B lymphocytes (47).

### Transcription hubs

In addition to tracking, linking, and looping, “transcription factories” was an influential model for enhancer activity (48). According to this view, TFs bound to distal enhancers mobilize the associated gene to discrete foci that reside in fixed locations within the nucleus. Most of the evidence for factories was obtained by using fixed tissues in cultured cells, including embryonic blood cells. These assays revealed discrete nuclear foci containing the phosphorylated form of Pol II, in addition to many coexpressed genes that are located in different chromosomal positions, and



**Fig. 3. Two types of topologies at complex loci.** (A) Left: Enhancer (E)–promoter proximity at the time of gene expression. Right: Preformed (compacted) topologies prior to gene expression. (B) Insulator:insulator pairing brings transgenic *eve*-promoter and endogenous enhancers (E) in proximity—from ~700 to ~400 nm, without *lacZ*-reporter transcription. (C) Further compaction (~335 nm) occurs during reporter transcription.



**Fig. 4. Hub and condensates model.** Preformed topologies increase local TF, coactivator, and Pol II concentration (hubs or microenvironments), where different enhancers (E) dynamically share common resources.

even on different chromosomes (48, 49). Because many fewer Pol II foci were detected [40 to 200 per cell (48)] compared to the number of actively transcribed genes per nucleus, the factory model proposed that multiple coexpressed genes move in and out of preassembled factories. With advances in live imaging, we now know that the system is much more dynamic. For example, super-resolution live imaging revealed highly dynamic and transient clusters of Pol II (50). These clusters do not reside in fixed locations within the nucleus, but are instead formed de novo upon transcriptional stimulation, persisting for short periods, on the order of a minute.

A dynamic variant of the transcription factory model (hubs) is gaining momentum as it incorporates features of all classical models of enhancer-promoter interactions, explains many observations reported for transcription factories, and accounts for more contemporary observations such as transcriptional bursting. According to this model, prelooped topologies serve as hubs or traps for the accumulation of Pol II and other complexes required for gene expression (Fig. 4). Liquid-liquid phase transitions were proposed to facilitate this process (51) because many TFs, coactivators, and components of the basal transcription machinery contain intrinsically disordered domains that can foster such interactions. Studies of the assembly of germline determinants (P-granules) in *Caenorhabditis elegans* indicate that different RNA and protein subunits associate through such phase transitions (52). Live-imaging assays permitted direct visualization of coalescing P-granule “droplets” in early *C. elegans* embryos.

According to this “hub and condensate” model for enhancer-promoter communication, enhancers need not directly touch their target promoters, but merely come into proximity, within 100 to

300 nm (45). The coalescence or aggregation of multiple Mediator complexes, preinitiation complexes, and Pol II could serve to bridge enhancers to their target promoters over such distances (Fig. 4). Obviously, this model is somewhat speculative, although recent studies (53–55) provide direct visualizations of dynamic Pol II, TF, and Mediator condensates at sites of active transcription. Moreover, active enhancers have a higher diffusion rate than inactive enhancers (56), which could be interpreted as less-restricted movement of TFs within liquid condensates.

## Conclusion

Single-cell genomics, genome editing, single-molecule live imaging, and super-resolution methods are animating classical “snapshots” of enhancer-promoter communication. The first “movies” suggest that each of the classical mechanisms—tracking, linking, looping, factories—could contribute to the overall process; however, none is sufficient. A holistic understanding will require the simultaneous visualization of enhancers, promoters, nascent transcripts, TFs, Pol II, and associated cofactors at specific loci within higher-order chromosomal topologies. In addition, complex genetic loci contain multiple enhancers and often use alternative promoters. Limitations in detection methods and resolution currently restrict the visualization of all the moving parts. However, the remarkable advances that we have witnessed in recent years give great promise for a new synthesis in our understanding of complex developmental and disease processes.

## REFERENCES AND NOTES

1. L. A. Lettice et al., *Development* **141**, 1715–1725 (2014).
2. N. Frankel et al., *Nature* **466**, 490–493 (2010).
3. E. Z. Kvon et al., *Nature* **512**, 91–95 (2014).

4. Y. Ghavi-Helm et al., *Nature* **512**, 96–100 (2014).
5. E. Cannavò et al., *Nature* **541**, 402–406 (2017).
6. J. R. Dixon, D. U. Gorkin, B. Ren, *Mol. Cell* **62**, 668–680 (2016).
7. J. Dekker, E. Heard, *FEBS Lett.* **589** (20PartA), 2877–2884 (2015).
8. A. Sanyal, B. R. Lajoie, G. Jain, J. Dekker, *Nature* **489**, 109–113 (2012).
9. J. E. Phillips-Cremmins et al., *Cell* **153**, 1281–1295 (2013).
10. S. S. Rao et al., *Cell* **159**, 1665–1680 (2014).
11. G. Fudenberg et al., *Cell Rep.* **15**, 2038–2049 (2016).
12. A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, X. Darzacq, *eLife* **6**, e25776 (2017).
13. M. C. Gambetta, E. E. M. Furlong, *Genetics* **210**, 129–136 (2018).
14. M. Ganji et al., *Science* **360**, 102–105 (2018).
15. O. Symmons et al., *Genome Res.* **24**, 390–400 (2014).
16. D. G. Lupiáñez et al., *Cell* **161**, 1012–1025 (2015).
17. V. Narendra et al., *Science* **347**, 1017–1021 (2015).
18. Y. Guo et al., *Cell* **162**, 900–910 (2015).
19. M. Franke et al., *Nature* **538**, 265–269 (2016).
20. E. P. Nora et al., *Cell* **169**, 930–944.e22 (2017).
21. S. S. P. Rao et al., *Cell* **171**, 305–320.e24 (2017).
22. W. Schwarzer et al., *Nature* **551**, 51–56 (2017).
23. T. Juven-Gershon, J. Y. Hsu, J. T. Kadonaga, *Genes Dev.* **22**, 2823–2830 (2008).
24. T. H. Chang et al., *Dev. Biol.* **269**, 595–608 (2004).
25. S. W. Cho et al., *Cell* **173**, 1398–1412.e22 (2018).
26. O. Symmons et al., *Dev. Cell* **39**, 529–543 (2016).
27. S. Kong, D. Bohl, C. Li, D. Tuan, *Mol. Cell. Biol.* **17**, 3955–3965 (1997).
28. P. Morcillo, C. Rosen, M. K. Baylies, D. Dorsett, *Genes Dev.* **11**, 2729–2740 (1997).
29. W. Deng et al., *Cell* **149**, 1233–1244 (2012).
30. T. M. Dunn, S. Hahn, S. Ogden, R. F. Schleif, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5017–5020 (1984).
31. R. A. Rollins, P. Morcillo, D. Dorsett, *Genetics* **152**, 577–593 (1999).
32. K. S. Wendt et al., *Nature* **451**, 796–801 (2008).
33. E. Splinter et al., *Genes Dev.* **20**, 2349–2354 (2006).
34. A. C. Bell, A. G. West, G. Felsenfeld, *Cell* **98**, 387–396 (1999).
35. E. de Wit et al., *Mol. Cell* **60**, 676–684 (2015).
36. A. S. Weintraub et al., *Cell* **171**, 1573–1588.e28 (2017).
37. S. Schoenfelder et al., *Nat. Genet.* **42**, 53–61 (2010).
38. I. Krivega, A. Dean, *Nucleic Acids Res.* **45**, 8255–8268 (2017).
39. B. Bonev et al., *Cell* **171**, 557–572.e24 (2017).
40. R. J. Palstra et al., *Nat. Genet.* **35**, 190–194 (2003).
41. C. G. Spiliarakis, R. A. Flavell, *Nat. Immunol.* **5**, 1017–1027 (2004).
42. G. Andrey et al., *Science* **340**, 1234167 (2013).
43. A. J. Rubin et al., *Nat. Genet.* **49**, 1522–1528 (2017).
44. H. Chen et al., *Nat. Genet.* **50**, 1296–1303 (2018).
45. B. Lim, T. Heist, M. Levine, T. Fukaya, *Mol. Cell* **70**, 287–296.e6 (2018).
46. I. Williamson, L. A. Lettice, R. E. Hill, W. A. Bickmore, *Development* **143**, 2994–3001 (2016).
47. J. S. Lucas, Y. Zhang, O. K. Dudko, C. Murre, *Cell* **158**, 339–352 (2014).
48. D. A. Jackson, F. J. Iborra, E. M. Manders, P. R. Cook, *Mol. Biol. Cell* **9**, 1523–1536 (1998).
49. C. S. Osborne et al., *Nat. Genet.* **36**, 1065–1071 (2004).
50. I. I. Cisse et al., *Science* **341**, 664–667 (2013).
51. D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, P. A. Sharp, *Cell* **169**, 13–23 (2017).
52. C. P. Brangwynne et al., *Science* **324**, 1729–1732 (2009).
53. S. Chong et al., *Science* **361**, eaar2555 (2018).
54. B. R. Sabari et al., *Science* **361**, eaar3958 (2018).
55. W. K. Cho et al., *Science* **361**, 412–415 (2018).
56. B. Gu et al., *Science* **359**, 1050–1055 (2018).

## ACKNOWLEDGMENTS

We are grateful to members of the Furlong and Levine labs, and to J. Jaynes, for comments and discussions. We apologize to colleagues whose work we could not discuss. **Funding:** We thank the European Union’s Horizon 2020 research and innovation programme for financial support [grant agreement 664918 (MRG-GRAMMAR) to E.E.M.F., NIH GM118147 to M.L.]. **Competing interests:** The authors declare no competing interests.

10.1126/science.aau0320



## REVIEW

# RNA modifications modulate gene expression during development

Michaela Frye<sup>1,2\*</sup>, Bryan T. Harada<sup>3,4</sup>, Mikaela Behm<sup>1</sup>, Chuan He<sup>3,4,5\*</sup>

RNA modifications have recently emerged as critical posttranscriptional regulators of gene expression programs. They affect diverse eukaryotic biological processes, and the correct deposition of many of these modifications is required for normal development. Messenger RNA (mRNA) modifications regulate various aspects of mRNA metabolism. For example, *N*<sup>6</sup>-methyladenosine (m<sup>6</sup>A) affects the translation and stability of the modified transcripts, thus providing a mechanism to coordinate the regulation of groups of transcripts during cell state maintenance and transition. Similarly, some modifications in transfer RNAs are essential for RNA structure and function. Others are deposited in response to external cues and adapt global protein synthesis and gene-specific translation accordingly and thereby facilitate proper development.

Understanding normal tissue development and disease susceptibility requires knowledge of the various cellular mechanisms that control gene expression in multicellular organisms. Much work has focused on investigation of lineage-specific transcriptional networks that govern stem cell differentiation (1). Yet gene expression programs are dynamically regulated during development and require the coordination of both mRNA metabolism and protein synthesis. The deposition of chemical modifications onto RNA has emerged as a basic mechanism to modulate cellular transcriptomes and proteomes during lineage fate decisions in development.

Many of the more than 170 modifications present in RNA have been known for decades, but only in the past several years have sufficiently sensitive tools and high-resolution genome-wide techniques been developed to identify and quantify these modifications in low-abundance RNA species such as mRNA (2, 3). Some RNA modifications have been shown to affect normal development; these modifications can control the

turnover and/or translation of transcripts during cell-state transitions and therefore play important roles during tissue development and homeostasis. In particular, the *N*<sup>6</sup>-methyladenosine (m<sup>6</sup>A) modification of mRNA is an essential regulator of mammalian gene expression (4, 5). Other modifications such as 5-methylcytosine (m<sup>5</sup>C) and *N*<sup>1</sup>-methyladenosine (m<sup>1</sup>A) are currently best described for their functional roles in noncoding RNAs but have also been studied in mRNA (4, 6).

We summarize here recent studies that elucidate the roles of RNA modifications in modulating gene expression throughout cell differentiation and animal development. Because of space limitations, we will focus on m<sup>6</sup>A in mRNA and m<sup>5</sup>C in tRNA as notable examples. RNA editing and RNA tail modifications, which have been comprehensively reviewed previously, will not be included.

## Types of RNA modifications

### Modifications in mRNA

In addition to the 5' cap and 3' polyadenylation, mRNAs contain numerous modified nucleosides,

including base isomerization to produce pseudouridine (Ψ); methylation of the bases to produce m<sup>6</sup>A, m<sup>1</sup>A, and m<sup>5</sup>C; methylation of the ribose sugar to install 2'-O-methylation (N<sub>m</sub>, m<sup>6</sup>A<sub>m</sub>); and oxidation of m<sup>5</sup>C to 5-hydroxymethylcytosine (hm<sup>5</sup>C) (4). Of these, one of the most abundant and well-studied mRNA modifications is m<sup>6</sup>A. Of all transcripts encoded by mammalian cells, 20 to 40% are m<sup>6</sup>A methylated, and methylated mRNAs tend to contain multiple m<sup>6</sup>A per transcript (2, 3). m<sup>6</sup>A and other RNA modifications are also present in long noncoding and microRNAs.

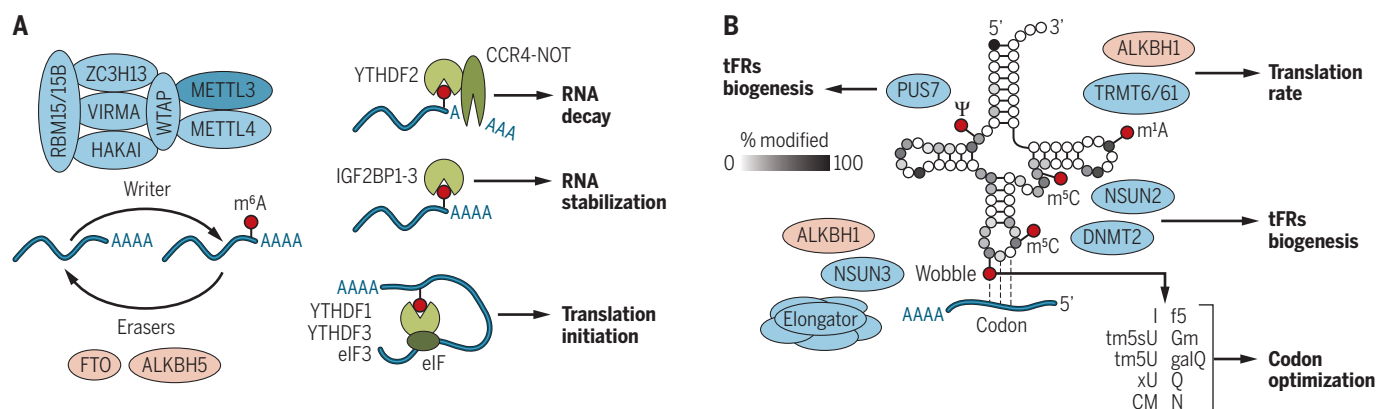
The biological functions of m<sup>6</sup>A are mediated by writer, eraser, and reader proteins (Fig. 1A) (4). m<sup>6</sup>A is installed by a multiprotein writer complex that consists of the METTL3 catalytic subunit, and many other accessory subunits (4). Two demethylases, FTO and ALKBH5, act as erasers (7, 8). m<sup>6</sup>A can both directly and indirectly affect the binding of reader proteins on methylated mRNAs to regulate the metabolism of these transcripts (4). For example, YTHDF2 binds to m<sup>6</sup>A in mRNA and targets the transcripts for degradation (4, 9), and YTHDF1, YTHDF3, and eIF3 promote translation of m<sup>6</sup>A-containing transcripts (4, 10, 11). The list of m<sup>6</sup>A readers that regulate mRNA homeostasis is still growing (12, 13), and the functions of m<sup>6</sup>A could depend on recognition by cell type-specific reader proteins. Reader and eraser proteins for other modifications are less well described.

### Modifications in tRNA and ribosomal RNA

In addition to mRNA, the faithful translation of the genetic code is orchestrated by at least two

<sup>1</sup>Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK. <sup>2</sup>German Cancer Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. <sup>3</sup>Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA. <sup>4</sup>Howard Hughes Medical Institute, The University of Chicago, Chicago, IL 60637, USA. <sup>5</sup>Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA.

\*Corresponding author. Email: m.frye@gen.cam.ac.uk (M.F.); chuanhe@uchicago.edu (C.H.)



**Fig. 1. Regulation of gene expression by RNA modifications.** (A) m<sup>6</sup>A is installed by a multicomponent writer complex with the catalytic subunit METTL3 and removed by the demethylase enzymes FTO and ALKBH5. m<sup>6</sup>A reader proteins can specifically bind m<sup>6</sup>A transcripts and effect different outcomes for methylated mRNAs. (B) RNA modifications in human eukaryotic tRNAs according to Modomics ([http://genesilico.pl/trnamodviz/jit\\_viz/select\\_tRNA](http://genesilico.pl/trnamodviz/jit_viz/select_tRNA)). xU, other modified uracil (U); N, unknown modified. How often a base is modified is shown by the grayscale. Only examples of writers (TRMT6/61, DNMT2, NSUN2, NSUN3, PUS7, and Elongator) and erasers (ALKBH1) are shown and how they affect translation. Modifications at the wobble base are most diverse.

more types of RNAs, tRNA and ribosomal RNA (rRNA). Human rRNAs contain a set of chemical modifications that often cluster at functionally important sites of the ribosome, such as the peptidyltransferase center and the decoding site (14). Modification in tRNAs are the most diverse, with cytoplasmic and mitochondrial tRNAs carrying more than 100 different modifications (Fig. 1B). A human tRNA can contain between 11 and 13 different modifications that are deposited at different steps during its maturation and could directly affect translation (15). The modifications range from simple methylation and isomerization events—including  $m^5C$ ,  $m^1A$ ,  $\Psi$ , 5-methyluridine ( $m^5U$ ), 1- and 7-methylguanosine ( $m^7G$ ,  $m^7G$ ), and inosine (I)—to complex multiple-step chemical modifications (Fig. 1B) (15). The function of a modification depends on both its location in the tRNA and its chemical nature. For example,  $m^5C$  is site-specifically deposited by at least three enzymes—NSUN2, NSUN3, and DNMT2 (Fig. 1B)—and all three enzymes influence tRNA metabolism differently. Modifications at the wobble position are the most diverse and often optimize codon usage during gene-specific translation (Fig. 1B) (16, 17).

### RNA modifications in development mRNA modifications in development

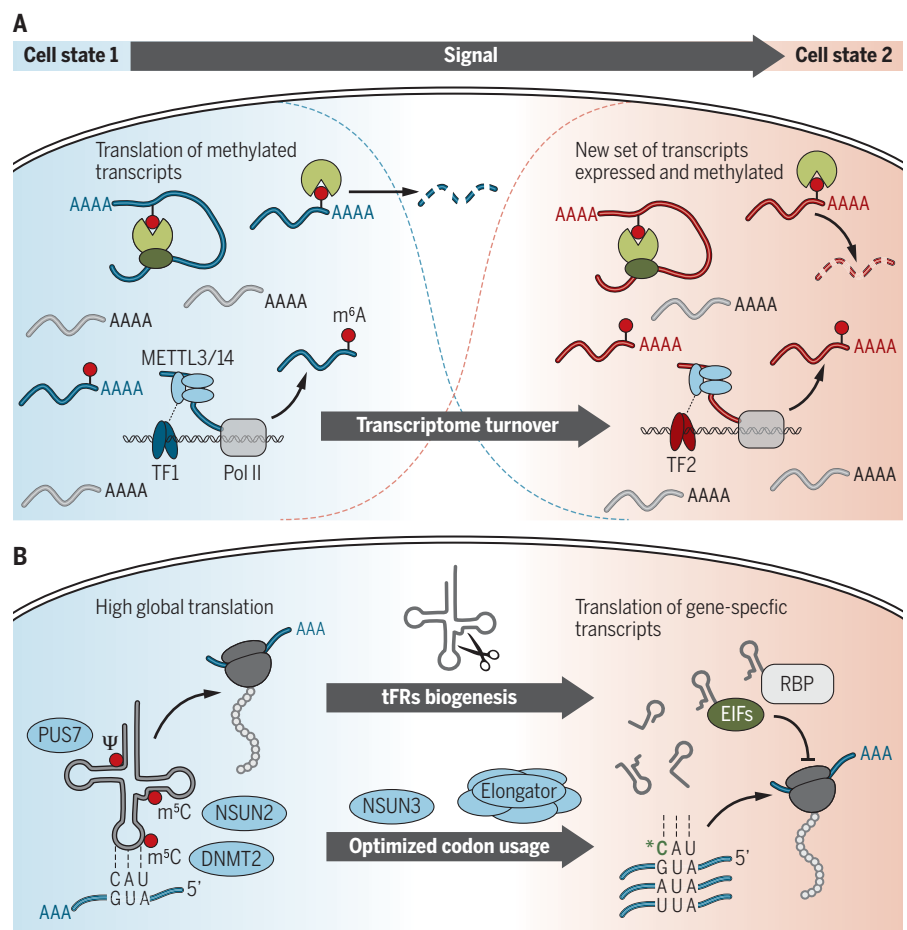
A wealth of recent studies identified an essential role for  $m^6A$  during development, and many of them highlighted a role for  $m^6A$  in the regulation of transcriptome switching during embryonic and adult stem cell differentiation (4). An early clue that  $m^6A$  is essential for development was the observation that removal of the  $m^6A$  writer enzyme *Mettl3* is embryonic lethal in mice (5). *Mettl3*<sup>-/-</sup> embryos appear normal before implantation but begin to show defects after implantation and are absorbed by embryonic day 8.5. Examination of gene expression from these embryos and from embryonic stem cells (ESCs) depleted of *Mettl3* suggested impaired exit from pluripotency because, for example, expression of the pluripotency factor *Nanog* was sustained (5, 18). Transcripts that encode certain pluripotency factors are methylated (5, 18, 19), which affects the turnover of these transcripts during differentiation. At least some of these transcripts are cotranscriptionally methylated through the recruitment of the  $m^6A$  writer complex by cell-state specific transcription factors such as *Smad2* and *Smad3* (20). Therefore,  $m^6A$  marks transcripts that encode important developmental regulators to facilitate their turnover during cell fate transitions and thereby enables cells to properly switch their transcriptomes from one cellular state to another (Fig. 2A).

This paradigm has also been used to explain the differentiation of other cell types. Conditional knockout of *Mettl3* in CD4<sup>+</sup> T cells prevents the proliferation and differentiation of naïve T cells through stabilization of *Socs* family genes (21). Loss of *Mettl14* (an essential component of the METTL3/14 methyltransferase complex) in the brain delays cortical neurogenesis and is associated with slower cell-cycle progression and

impaired decay of transcripts that are involved in lineage specification of cortical neural stem cells (22). Similarly, deletion of *Ythdf2* delays mouse neuronal development through impaired proliferation and differentiation of neural stem and progenitor cells (23).  $m^6A$ -mediated RNA decay also regulates various stages of zebrafish development. For example, during the maternal-to-zygotic transition, embryos that lack *Ythdf2* exhibit impaired clearance of maternal transcripts, delaying embryonic development (24). Loss of *Mettl3* blocks the endothelial-to-hematopoietic transition in zebrafish because of loss of the *Ythdf2*-mediated decay of genes that specify endothelial cell fate, such as *Notch1a* and *Rhoca* (25).

Although these studies highlight the functional roles of the YTHDF2-mediated clearance of mRNAs, loss of *Ythdf2* only partially accounts for

phenotypes associated with loss of *Mettl3*. For example, loss of *Mettl3* impairs priming of mammalian ESCs, yet *Ythdf2* knockout embryos are able to exit pluripotency (23, 26). Similarly, *Mettl3* deletion in zebrafish is lethal owing to severe hematopoietic defects, but adult *Ythdf2* knockout fish seem to be normal (24, 25). Work on gametogenesis highlights the importance of other  $m^6A$  eraser and reader proteins in development because loss of *Mettl3*, *Mettl14*, *Alkbh5*, *Ythdf2*, and *Ythdc2* are all associated with impaired fertility and defects in spermatogenesis and/or oogenesis (8, 26–32). These defects were associated with the altered abundance, translation efficiency, and splicing of methylated transcripts that encode regulators of gametogenesis. Work in *Drosophila* suggests important roles for  $m^6A$  in mediating splicing because deletion of *Ime4*, the *Mettl3*



**Fig. 2. RNA modifications regulate cell differentiation and development. (A)** Model for the roles of  $m^6A$  in cell differentiation. In the naïve, undifferentiated state, cell state-specific master transcription factors recruit the METTL3 complex to methylate transcripts that encode cell fate factors. Translation of these methylated factors may aid in the maintenance of cell state and prevent differentiation. When cells initiate differentiation and switch their transcriptional program, reader proteins mediate the turnover of the methylated transcripts to facilitate transcriptome switching. **(B)** Modification by NSUN2, DNMT2, and PUS7 protects tRNAs from cleavage and production of tRFs, which enables high global translation. In a different cell state, tRFs can affect global and gene-specific protein translation by displacing distinct RBPs and are therefore important players in stem cell differentiation. Wobble tRNA modifications—for example, by NSUN3 and Elongator—enhance the versatility of tRNA anticodon to recognize mRNA to optimize codon usage and translation of cytoplasmic and mitochondrial mRNAs during differentiation.

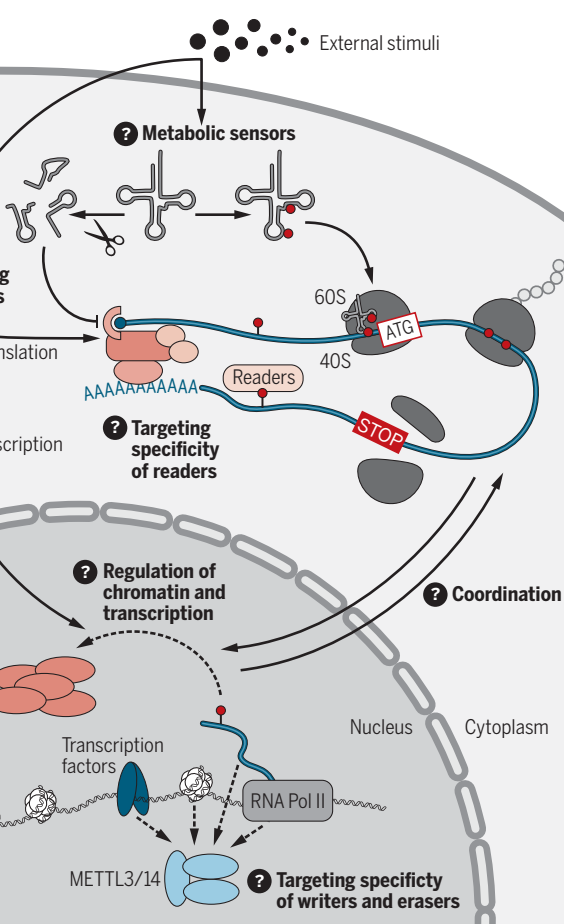
homolog, and other m<sup>6</sup>A writer complex subunits reduces viability of females owing to inappropriate splicing of *Sex lethal* (Sxl), an important regulator of dosage compensation and sex determination (33, 34).

Together, these studies demonstrate that the functional network that coordinates mRNA methylation is highly complex and highlight the requirement of m<sup>6</sup>A for the proper execution of stem cell differentiation programs (Fig. 2A). Transcripts that maintain a cell state are most likely cotranscriptionally decorated with m<sup>6</sup>A through the recruitment of the writer complex by cell state-specific transcription factors. Whereas m<sup>6</sup>A promotes the decay of these transcripts, active transcription may maintain them at steady-state levels, with other readers potentially aiding in mediating their processing and translation. Upon receiving the signal(s) for cells to differentiate and repress transcription of these factors, m<sup>6</sup>A coordinates the timely decay of these transcripts, which allows cells to differentiate. Although other posttranscriptional mechanisms aid in the promotion of cell-state switching, m<sup>6</sup>A writers and readers being required for many of these transitions suggests that m<sup>6</sup>A regulates gene expression in ways that cannot be substituted by other similar mechanisms.

### tRNA modifications in development

Although RNA modifications are highly diverse and found in all RNA species, the recent discoveries underpin an emerging common theme: RNA modifications coordinate translation of transcripts that encode functionally related proteins when cells respond to differentiation or other cellular and environmental cues. Loss of tRNA modifying enzymes can delay stem cell differentiation, often only in distinct tissues. For instance, knockout of *Nsun2* delays stem cell differentiation in the brain and skin (35, 36). Depletion of the pseudouridine synthase PUS7 impairs hematopoietic stem cell commitment, and loss of *Dnmt2* delays endochondral ossification (37, 38). Knockout of *Elp3*, a core component of Elongator that modifies the tRNA wobble position, is embryonic lethal (39).

Several recent studies reveal that the dynamic deposition of tRNA modifications is a fast and efficient way for cells to adapt the protein translation machinery to external stimuli (38, 40–42). For example, self-renewing stem cells must be resilient to external differentiation cues and main-



**Fig. 3. Future directions for research into gene regulation by RNA modifications.** There are several unresolved questions in the field: Mechanistically, how do external stimuli regulate RNA modification to affect protein translation rates and transcription? How do RNA modifying enzymes act as metabolic sensors? How do RNA modifications directly or indirectly regulate chromatin regulatory complexes to affect chromatin state or transcription? What factors—such as transcription factors, chromatin, RNA, RBPs, or components of the RNA polymerase II complex—recruit m<sup>6</sup>A writer and eraser enzymes to their targets? What factors regulate and determine the target specificity of readers? How are the protein synthesis and transcription machineries coordinated by RNA modifications?

tain protein synthesis at a low rate, yet their differentiation requires high levels of protein synthesis to produce committed progenitors (40, 43, 44). The deposition of RNA modifications into tRNAs represents an efficient way to adapt energy requirements to specific cell states.

Recent studies discovered that tRNA modifications regulate protein translation rates during development via tRNA-derived small noncoding RNA fragments (tRFs) (6, 38). Loss of NSUN2-mediated methylation at the variable loop increases the affinity to the endonuclease angiogenin, promotes cleavage of tRNAs into tRFs, and inhibits global protein synthesis (35, 40). Similarly, the Ψ writer PUS7 modifies tRNAs and thereby influences the formation of tRFs, which then target the translation initiation complex (38). Loss of DNMT2-mediated methylation at the anticodon loop (C38) causes both tRNA-specific fragmentation and codon-specific mistranslation (37). Thus,

altered tRNA modification patterns shape tRF biogenesis and determine their intracellular abundances (Fig. 2B). tRFs could act on global and gene-specific protein translation by displacing distinct RNA-binding proteins (RBPs) and are therefore important players in stem cell differentiation (38, 40), sperm maturation (45), retrotransposon silencing (46), intergenerational transmission of paternally acquired metabolic disorders (47), and breast cancer metastasis (48).

Wobble tRNA modifications enhance the versatility of tRNA anticodons to recognize mRNA to optimize codon usage and translation of cytoplasmic and mitochondrial mRNAs (Fig. 2B) (16, 49–51). Mitochondria are crucial players in stem cell activation, fate decisions, tissue regeneration, aging, and diseases (52). Mitochondrial translation can be affected by mitochondrial tRNA and mRNA modifications. For example, mammalian mitochondria use folate-bound one-carbon (1C) units to methylate tRNA through the serine hydroxymethyltransferase 2 (SHMT2). SHMT2 provides methyl donors to produce the taurinomethyluridine base at the wobble position of distinct mitochondrial tRNAs. Loss of the catalytic activity of SHMT2 impairs oxidative phosphorylation and mitochondrial translation (53).

Stem cell differentiation requires the constant and dynamic adaption of energy supply to fuel protein synthesis. A highly efficient and fast trigger to adapt global and gene-specific protein translation rates to external stimuli is the dynamic deposition and removal of modifications in tRNAs.

### RNA modifications in disease tRNA modifications in disease

Complex human pathologies that are directly linked to tRNA modifications include cancer, type 2 diabetes, neurological disorders, and mitochondrial-linked disorders (54). The human brain is particularly sensitive to defects in tRNA modifications (55), and the cellular defects are commonly caused by impaired translational efficiency and misfolded proteins, leading to a deleterious activation of the cellular stress response.

Similar to normal tissues, tumor cells are challenged by a changing microenvironment—for example, through hypoxia, inflammatory cell infiltration, and exposure to cytotoxic drug treatments (40). Thus, tumor cell populations rely on the correct deposition of tRNA modifications to switch their transcriptional and translation programs dynamically in response to external stimuli.



For instance, mouse skin tumors that lack the NSUN2-mediated m<sup>5</sup>C modification repress global protein synthesis, leading to an enlarged undifferentiated tumor-initiating cell population (40). However, the up-regulation of NSUN2 and methylation of tRNAs is strictly required for cell survival in response to chemotherapeutic drug treatment, and NSUN2-negative tumors fail to regenerate after exposure to cytotoxic drug treatments (40). Thus, tumor-initiating cell populations require the tight control of protein synthesis for accurate cell responses and to maintain the bulk tumor.

Similarly, modifications found in other non-coding RNAs are likely to play important roles in their biogenesis and function. For instance, the biogenesis of rRNA is known to be substantially affected by various modifications, the defect of which could contribute to human ribosomopathies (56).

### mRNA modifications in disease

mRNA modifications also contribute to the survival and growth of tumor cells, further highlighting the importance of mRNA modifications in the regulation of cell fate decisions. The METTL3 and METTL14 subunits of the m<sup>6</sup>A writer complex are highly expressed in human hematopoietic stem and progenitor cells (HSPCs), and the expression of these two subunits declines during differentiation of HSPCs along the myeloid lineage (57, 58). Overexpression of METTL3 inhibits cell differentiation and increases cell growth (57, 58). Consistent with a role in maintaining self-renewal programs, METTL3 and METTL14 are overexpressed in acute myeloid leukemia (AML), and AML cells are sensitive to depletion of METTL3 and METTL14 (57–59). These effects could be mediated by changes in the methylation of cell state-specific transcripts such as *MYC*, *MYB*, *BCL2*, *PTEN*, and *SP1* that help to maintain self-renewal and prevent differentiation (57–59). The stabilization of certain m<sup>6</sup>A methylated transcripts in AML cells may be mediated by the IGF2BP1-3 family of m<sup>6</sup>A reader proteins rather than the YTHDF1-3 family (13, 58).

An opposite role for m<sup>6</sup>A in leukemogenesis was found in certain subtypes of AML with increased expression of the demethylase FTO, resulting in decreased m<sup>6</sup>A and elevated levels of oncogene transcripts (60). Inhibition of FTO reduces AML cell proliferation and viability in these cell types (60, 61). The mechanisms and pathways for the writers and eraser to affect AML are likely distinct. Whereas elevated writer expression blocks differentiation of HPSCs to contribute to AML initiation and cell survival, elevated FTO mostly affects AML proliferation. This distinction is exemplified by the dual role of the oncometabolite R-2HG; its inhibition of TET2 contributes to AML initiation but also inhibits FTO in a subset of AML, leading to repressed proliferation (61). Decreased m<sup>6</sup>A is also associated with some solid tumors, likely promoting their proliferation. For example, in breast cancer, hypoxia was shown to induce the

overexpression of ALKBH5, an m<sup>6</sup>A eraser, and ZNF217, a transcription factor that can inhibit METTL3, resulting in reduction of the m<sup>6</sup>A methylation and decay of transcripts such as *Nanog* (62, 63). Similarly, overexpression of ALKBH5 or down-regulation of METTL3 or METTL14 promotes the tumorigenicity of glioblastoma cells through stabilization of pro-proliferative transcripts such as *FOXMI* (64, 65). In endometrial cancer, reduced m<sup>6</sup>A promotes cell proliferation through misregulation of transcripts encoding regulators of the AKT pathway (66). Additional mechanisms for how m<sup>6</sup>A alters gene expression to help drive cancer progression are likely to be discovered in the future.

### Future perspectives

Although these studies demonstrate the roles of RNA modifications in various developmental processes, our understanding of how RNA modifications contribute to these processes remains incomplete, especially at the mechanistic level (Fig. 3). The development of new tools that can determine the transcriptome-wide distribution of RNA modifications at nucleotide resolution with quantitative information about the modification fraction would greatly help in these endeavors. Further, it will be essential to understand the intrinsic and extrinsic factors that determine the specificity of the RNA modification writers, readers, and erasers and how these proteins are regulated in different cell types across development. For many RNA modifications, there is only very little information available on how these modifications recruit or repel RBPs, yet this information is essential to understand how RNA modifications modulate the RNA processing or protein translation machineries. In addition, how cells adjust RNA modifications and adapt the protein synthesis machinery in response to metabolic requirements remains largely unclear—in particular, how these changes in translation could have cell type-specific effects. Last, recent studies have suggested that m<sup>6</sup>A could directly or indirectly influence chromatin state and transcription through regulation of chromatin regulatory complexes and long noncoding RNAs (67, 68). The potential roles of m<sup>6</sup>A and other RNA modifications in shaping chromatin states may provide additional mechanisms for explaining how these modifications contribute to gene regulation in development.

### REFERENCES AND NOTES

- E. E. M. Furlong, M. Levine, *Science* **361**, 1341–1345 (2018).
- D. Dominissini et al., *Nature* **485**, 201–206 (2012).
- K. D. Meyer et al., *Cell* **149**, 1635–1646 (2012).
- I. A. Roundtree, M. E. Evans, T. Pan, C. He, *Cell* **169**, 1187–1200 (2017).
- S. Geula et al., *Science* **347**, 1002–1006 (2015).
- M. Frye, S. Blanco, *Development* **143**, 3871–3881 (2016).
- G. Jia et al., *Nat. Chem. Biol.* **7**, 885–887 (2011).
- G. Zheng et al., *Mol. Cell* **49**, 18–29 (2013).
- X. Wang et al., *Nature* **505**, 117–120 (2014).
- K. D. Meyer et al., *Cell* **163**, 999–1010 (2015).
- X. Wang et al., *Cell* **161**, 1388–1399 (2015).
- R. R. Edupuganti et al., *Nat. Struct. Mol. Biol.* **24**, 870–878 (2017).
- H. Huang et al., *Nat. Cell Biol.* **20**, 285–295 (2018).

- K. E. Sloan et al., *RNA Biol.* **14**, 1138–1152 (2017).
- P. Schimmel, *Nat. Rev. Mol. Cell Biol.* **19**, 45–58 (2018).
- R. Schaffrath, S. A. Leidel, *RNA Biol.* **14**, 1209–1222 (2017).
- G. Hanson, J. Collier, *Nat. Rev. Mol. Cell Biol.* **19**, 20–30 (2018).
- P. J. Batista et al., *Cell Stem Cell* **15**, 707–719 (2014).
- Y. Wang et al., *Nat. Cell Biol.* **16**, 191–198 (2014).
- A. Bertero et al., *Nature* **555**, 256–259 (2018).
- H. B. Li et al., *Nature* **548**, 338–342 (2017).
- K. J. Yoon et al., *Cell* **171**, 877–889.e17 (2017).
- M. Li et al., *Genome Biol.* **19**, 69 (2018).
- B. S. Zhao et al., *Nature* **542**, 475–478 (2017).
- C. Zhang et al., *Nature* **549**, 273–276 (2017).
- I. Ivanova et al., *Mol. Cell* **67**, 1059–1067.e4 (2017).
- K. Xu et al., *Cell Res.* **27**, 1100–1114 (2017).
- Z. Lin et al., *Cell Res.* **27**, 1216–1230 (2017).
- P. J. Hsu et al., *Cell Res.* **27**, 1115–1127 (2017).
- M. N. Wojtas et al., *Mol. Cell* **68**, 374–387.e12 (2017).
- A. S. Bailey et al., *eLife* **6**, e26116 (2017).
- D. Jain et al., *eLife* **7**, e30919 (2018).
- U. Haussmann et al., *Nature* **540**, 301–304 (2016).
- T. Lence et al., *Nature* **540**, 242–247 (2016).
- S. Blanco et al., *EMBO J.* **33**, 2020–2039 (2014).
- S. Blanco et al., *PLOS Genet.* **7**, e1002403 (2011).
- F. Tuorto et al., *EMBO J.* **34**, 2350–2362 (2015).
- N. Guzzi et al., *Cell* **173**, 1204–1216.e26 (2018).
- H. Yoo, D. Son, Y. J. Jang, K. Hong, *Biochem. Biophys. Res. Commun.* **478**, 631–636 (2016).
- S. Blanco et al., *Nature* **534**, 335–340 (2016).
- S. Delaunay et al., *J. Exp. Med.* **213**, 2503–2523 (2016).
- D. D. Nedialkova, S. A. Leidel, *Cell* **161**, 1606–1618 (2015).
- E. Llorens-Bobadilla et al., *Cell Stem Cell* **17**, 329–340 (2015).
- R. A. Signer, J. A. Magee, A. Salic, S. J. Morrison, *Nature* **509**, 49–54 (2014).
- U. Sharma et al., *Science* **351**, 391–396 (2016).
- A. J. Schorn, M. J. Gutbrod, C. LeBlanc, R. Martienssen, *Cell* **170**, 61–71.e11 (2017).
- Y. Zhang et al., *Nat. Cell Biol.* **20**, 535–540 (2018).
- H. Goodarzi et al., *Cell* **161**, 790–802 (2015).
- L. Van Haute et al., *Nat. Commun.* **7**, 12039 (2016).
- S. Nakano et al., *Nat. Chem. Biol.* **12**, 546–551 (2016).
- S. Haag et al., *EMBO J.* **35**, 2104–2119 (2016).
- H. Zhang, K. J. Menzies, J. Auwerx, *Development* **145**, dev143420 (2018).
- R. J. Morscher et al., *Nature* **554**, 128–132 (2018).
- A. G. Torres, E. Battle, L. Ribas de Pouplana, *Trends Mol. Med.* **20**, 306–314 (2014).
- A. Bednářová et al., *Front. Mol. Neurosci.* **10**, 135 (2017).
- M. M. Parks et al., *Sci. Adv.* **4**, eaao0665 (2018).
- L. P. Vu et al., *Nat. Med.* **23**, 1369–1376 (2017).
- H. Weng et al., *Cell Stem Cell* **22**, 191–205.e9 (2018).
- I. Barbieri et al., *Nature* **552**, 126–131 (2017).
- Z. Li et al., *Cancer Cell* **31**, 127–141 (2017).
- R. Su et al., *Cell* **172**, 90–105.e23 (2018).
- C. Zhang et al., *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2047–E2056 (2016).
- C. Zhang et al., *Oncotarget* **7**, 64527–64542 (2016).
- Q. Cui et al., *Cell Reports* **18**, 2622–2634 (2017).
- S. Zhang et al., *Cancer Cell* **31**, 591–606.e6 (2017).
- J. Liu et al., *Nat. Cell Biol.* **20**, 1074–1083 (2018).
- D. P. Patil et al., *Nature* **537**, 369–373 (2016).
- Y. Wang et al., *Nat. Neurosci.* **21**, 195–206 (2018).

### ACKNOWLEDGMENTS

**Funding:** B.T.H. is supported by National Cancer Institute fellowship F32 CA221007. C.H. is supported by the National Institutes of Health (HG008935 and GM071440). C.H. is an investigator of the Howard Hughes Medical Institute. M.F. is supported by a Cancer Research UK Senior Fellowship (C10701/A15181), the European Research Council (ERC: 310360), and the Medical Research Council UK (MR/M01939X/1). Part of this work was carried out in the framework of the European COST action EPITRAN 16120. **Competing interests:** C.H. is a scientific founder of Accent Therapeutics and a member of its scientific advisory board. M.F. consults for Storm Therapeutics. All other authors declare no competing financial interests.

10.1126/science.aau1646

# RESEARCH

## *Hox* gene function in the sea anemone *Nematostella vectensis*

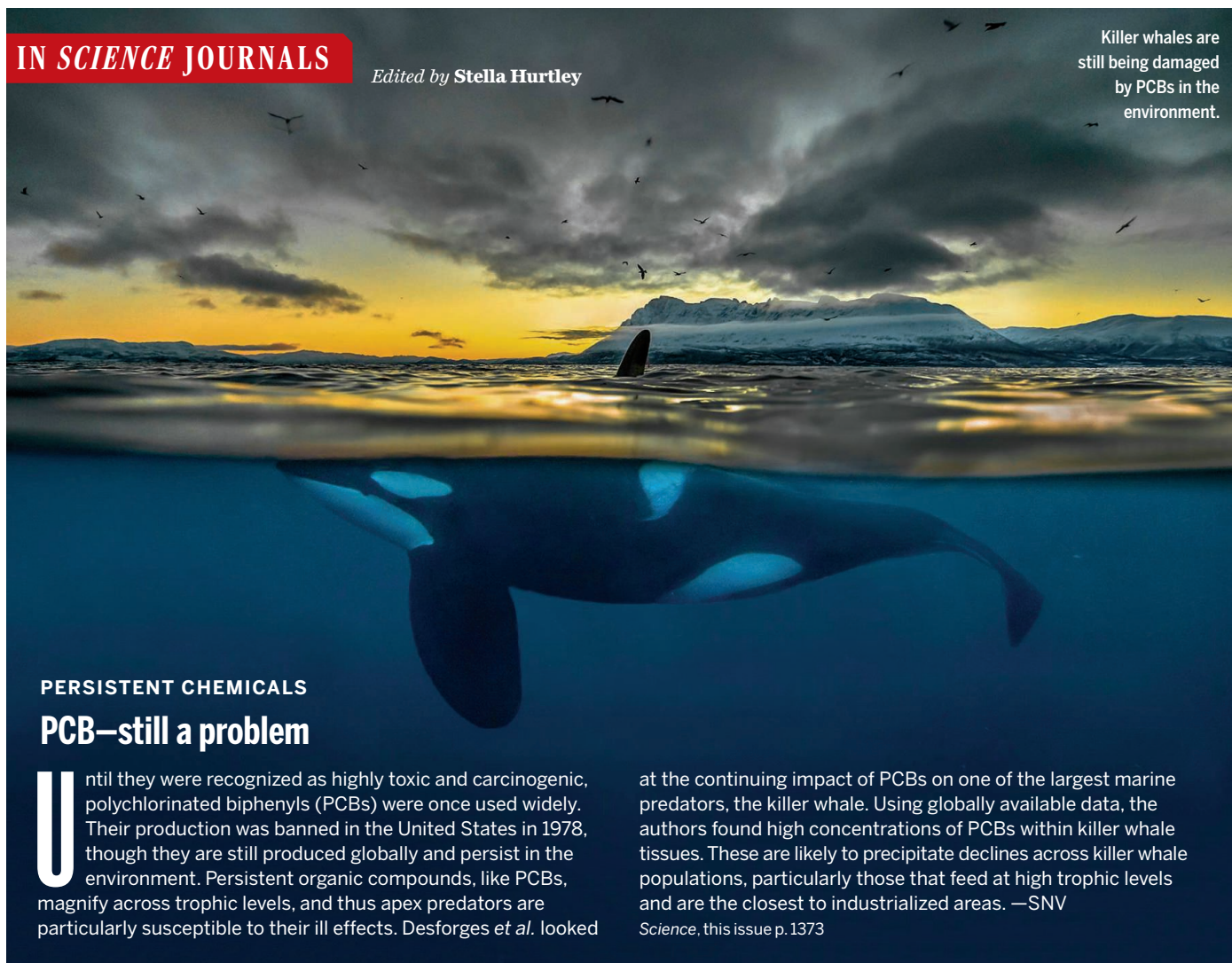
He et al., p. 1377



## IN SCIENCE JOURNALS

Edited by Stella Hurtley

Killer whales are still being damaged by PCBs in the environment.



### PERSISTENT CHEMICALS

#### PCB—still a problem

Until they were recognized as highly toxic and carcinogenic, polychlorinated biphenyls (PCBs) were once used widely. Their production was banned in the United States in 1978, though they are still produced globally and persist in the environment. Persistent organic compounds, like PCBs, magnify across trophic levels, and thus apex predators are particularly susceptible to their ill effects. Desforges *et al.* looked

at the continuing impact of PCBs on one of the largest marine predators, the killer whale. Using globally available data, the authors found high concentrations of PCBs within killer whale tissues. These are likely to precipitate declines across killer whale populations, particularly those that feed at high trophic levels and are the closest to industrialized areas. —SNV

*Science*, this issue p. 1373

### SINGLE-CELL GENOMICS

#### Single-cell chromatin and RNA analysis

Single-cell analyses have begun to provide insight into the differences among and within the individual cells that make up a tissue or organism. However, technological barriers owing to the small amount of material present in each single cell have prevented parallel analyses. Cao *et al.* present sci-CAR, a pooled barcode

method that jointly analyzes both the RNA transcripts and chromatin profiles of single cells. By applying sci-CAR to lung adenocarcinoma cells and mouse kidney tissue, the authors demonstrate precision in assessing expression and genome accessibility at a genome-wide scale. The approach provides an improvement over bulk analysis, which can be confounded by differing cellular subgroups. —LMZ

*Science*, this issue p. 1380

### MOLECULAR BIOLOGY

#### How cells ensure symmetric inheritance

Parental histones with modifications are recycled to newly replicated DNA strands during genome replication, but do the two sister chromatids inherit modified histones equally? Yu *et al.* and Petryk *et al.* found in mouse and yeast, respectively, that modified histones are segregated to both DNA daughter strands

in a largely symmetric manner (see the Perspective by Ahmad and Henikoff). However, the mechanisms ensuring this symmetric inheritance in yeast and mouse were different. Yeasts use subunits of DNA polymerase to prevent the lagging-strand bias of parental histones, whereas in mouse cells, the replicative helicase MCM2 counters the leading-strand bias. —SYM

*Science*, this issue p. 1386, p. 1389; see also p. 1311

CREDITS: (TOP) HE ET AL.; (BOTTOM): PHOTO: AUDUN RIKARDSEN



## BIOPHYSICS

### Manta rays' unique filtration system

Manta rays feed by opening their mouths as they swim, catching tiny plankton while simultaneously expelling water through their gills. How their filtering apparatus allows food that is smaller than their filter pore size to be captured, especially without clogging, remains elusive. Divi *et al.* examined the fluid flow around the filter-feeding mouth of the manta ray using physical and computational modeling. Flow separation happened behind the leading edge of each filter lobe, causing a large vortex within each pore. This meant that tiny particles were repelled away from the filters instead of passing through. Contact forces caused the particles to “ricochet” away from the filter pore and backward to the faster-moving freestream flow; thus, the particles collected above the filter, rather than on top of it (which would cause it to clog). This unique filtration system may hold interesting industrial applications. —PJB

*Sci. Adv.* 10.1126/sciadv.aat9533 (2018).

## FIBROSIS

### PD-1<sup>+</sup> T cells prompt pulmonary fibrosis

Although immunological T cells expressing programmed cell death 1 (PD-1) are sometimes described as exhausted, they are not too exhausted to wreak havoc in a variety of settings. Celada *et al.* examined cells from patients with sarcoidosis or idiopathic pulmonary fibrosis and saw an increase in PD-1<sup>+</sup> CD4<sup>+</sup> T cells relative to healthy controls. These cells were mostly T helper 17 cells and were able to induce fibroblasts to produce collagen in vitro. Blocking PD-1 in the coculture system prevented this induction and associated cytokine production from the T cells. Furthermore, blocking PD-1 in a mouse model of pulmonary

fibrosis reduced fibrosis symptoms. —LP  
*Sci. Transl. Med.* 10, eaar8356 (2018).

## BIOSYNTHESIS

### How algae turn tides toxic

Algal blooms can devastate marine mammal communities through the production of neurotoxins that accumulate within the food web. Brunson *et al.* identified a cluster of genes associated with biosynthesis of the neurotoxin domoic acid in a marine diatom (see the Perspective by Pohnert *et al.*). In vitro experiments established a series of enzymes that create the core structure of the toxin. Knowledge of the genes involved in domoic acid production will allow for genetic monitoring of algal blooms and aid in identifying conditions that trigger toxin production. —MAF

*Science*, this issue p. 1356;  
see also p. 1308

## ULTRAFAST OPTICS

### Making ultrafast cycles of light

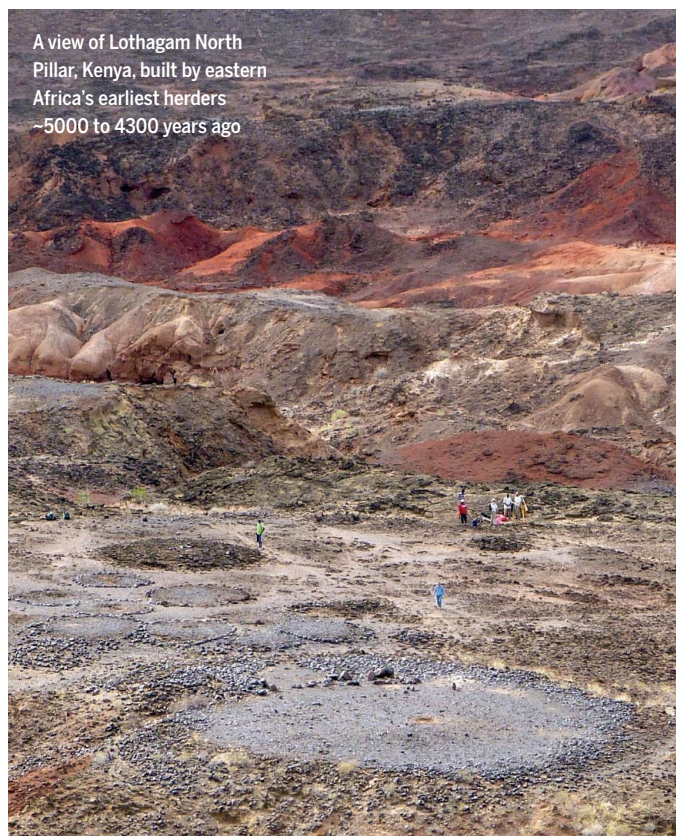
The ability to generate coherent optical frequency combs has had a huge impact on precision metrology, imaging, and sensing applications. On closer inspection, the broadband “white light” generated through the interaction of femtosecond mode-locked laser pulses is composed of billions or trillions of precisely spaced wavelengths of light. Carlson *et al.* demonstrate an alternative to the mode-locked laser approach—the electro-optic modulation of a continuous-wave laser light source can also generate optical frequency combs (see the Perspective by Torres-Company). The electro-optic modulation techniques can operate at much higher repetition rates than mode-locked lasers, which means they could potentially yield even more precise measurements. —ISO

*Science*, this issue p. 1358;  
see also p. 1316

## IN OTHER JOURNALS

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

A view of Lothagam North Pillar, Kenya, built by eastern Africa's earliest herders ~5000 to 4300 years ago



## ARCHAEOLOGY

### A massive cemetery at Lake Turkana

The oldest monumental architecture in eastern Africa was built about 5000 years ago. Apparently, these are commemorative structures built by pastoralists. Hildebrand *et al.* report a detailed excavation and ground-penetrating radar survey at the Lothagam North Pillar Site in Kenya. A massive mortuary was revealed, containing the skeletal remains of nearly 600 individuals buried over several centuries. Associated personal ornamentation included beads, colored stone, tusks, and a headpiece with over 400 teeth from more than 100 gerbils. There appeared to be no differentiation by hierarchy, gender, or age. The authors suggest that this site and others like it would have been stable landmarks at which mobile communities could gather and interact. —AMS

*Proc. Natl. Acad. Sci. U.S.A.* 115, 8942 (2018).

## STEM CELLS

### The blood, by night and day

Many developmental events have now been associated with the daily cycle. But what is happening physiologically

during the hours of light and dark? Bone turnover, immunity, and stem cell migration from the bone marrow are certainly affected. Golan *et al.* observed two peaks in bone marrow hematopoietic stem and progenitor cells (HSPCs) after



## EVOLUTION

## Father, do not eat thy son

**B**eing a parent is costly in many ways. In hard times, some animal parents resort to desertion, abortion, or even cannibalism. Such behaviors may allow adults to regain energy reserves and bank them for reproduction under more favorable conditions. Matsumoto *et al.* studied filial cannibalism in caregiving male blennies (*Rhabdoblennius nitidus*). These fish cannot court additional females while guarding eggs because their androgen supplies are suppressed by the presence of the eggs. Hormone synthesis only resumes in the fathers when all the eggs are removed, but not necessarily eaten, from a blenny's nest. Thus, the blennies eschew cannibalism in favor of infanticide to rapidly reset their reproductive physiology. In this way, they gain another chance at courtship and potentially a bigger clutch of eggs. —CA

*Curr. Biol.* **28**, 2831 (2018).

*Rhabdoblennius nitidus*  
males destroy their  
eggs to reset fertility.



transient increases of norepinephrine and tumor necrosis factor. The nighttime peak increases melatonin secretion to maintain and replenish HSPCs and to reduce differentiation. By contrast, the morning peak increases differentiation and vascular permeability for blood replenishment. Hence, light and dark provide different signals that are crucial for blood cell activity and overall health. —BAP

*Cell Stem Cell* **10**, 1016/  
j.stem.2018.08.002 (2018).

## WORKFORCE POPULATION

### Connecting the bio to the tech

With so many investments being made in training current STEM students to enter a 21st-century job market, why are some biotech companies struggling to fill entry-level positions? Thompson *et al.* surveyed academic and industry stakeholders and found disconnects on several levels. On the university side, faculty and career counselors are unfamiliar with the career opportunities, partly because biotech companies have done a relatively poor job of

marketing entry-level careers. Several obstacles limiting the placement of STEM graduates arise from this, including students not being job ready at graduation and students pursuing education beyond what is needed, leaving them overqualified. The authors make recommendations for encouraging and implementing partnerships among academia, industry, government agencies, and undergraduate institutions as a way forward. —MMc

*CBE Life Sci. Educ.* **17**, es12 (2018).

## PLASMONICS

### Controlling gold helices with DNA

The plasmonic properties of gold nanoparticles can be exploited by creating arrays that control their spacing and orientation to induce chiral optical properties. Lan *et al.* attached gold nanorods to V-shaped hinged DNA origami structures, thereby allowing the DNA origami to self-assemble along their edges and create staircase helical structures. By interchanging the spacer that sets the hinge angle of the V with a longer spacer, they converted the helix into

an extended structure with a different circular-dichroism spectrum. Alternatively, the authors show that a fixed spacer can be used and additional DNA strands attached at the hinge and at the ends of the V structures. These strands could then be reversibly displaced by ones that cause the helix to switch handedness. —PDS

*J. Am. Chem. Soc.* **140**, 11763 (2018).

## SCIENCE LEARNING

### Better communication without politics

How can preconceived beliefs be overcome so that people can avoid misperceptions about climate change? Guilbeault *et al.* developed an experimental system in which 2400 conservative and liberal individuals were placed into networks with either no cues as to the political orientation of other participants or visible designations. Groups were asked to interpret a graph showing trends in Arctic sea ice. At first, nearly 40% of conservatives incorrectly interpreted the data, saying that amounts of Arctic sea ice would be increasing by 2025; 26% of liberals made the same mistake. After

network interactions in the absence of political identifiers, more than 85% of individuals of either ideology correctly saw the trend toward decreasing Arctic sea ice. Knowledge of the political identification or affiliation of others prevented social learning. —BJ

*Proc. Natl. Acad. Sci. U.S.A.* **10**, 1073/  
pnas.1722664115 (2018).

## LONGEVITY

### Fasting to improve health span

Patterns of eating, as well as caloric content, are important determinants of health span. Mitchell *et al.* investigated how dietary composition and feeding patterns influence longevity in a large cohort of mice over 54 weeks. Caloric restriction (a 30% reduction in daily intake) or single-meal feeding (resulting in fasting during each day but no caloric restriction) increased life span and delayed the onset of age-associated pathologies in mice compared with ad libitum feeding. It is possible that daily fasting, even without caloric restriction, may improve health span in humans. —GKA

*Cell Metab.* **10**, 1016/  
j.cmet.2018.08.011 (2018).

ALSO IN *SCIENCE* JOURNALSEdited by **Stella Hurtley****EPIGENOMICS****Dissecting the epigenomic footprint**

Genome-wide epigenetic marks regulate gene expression, but the amount and function of variability in these marks are poorly understood. Working with human-derived samples, Onuchic *et al.* examined disease-associated genetic variation and sequence-dependent allele-specific methylation at gene regulatory loci. Regulatory sequences within individual chromosomal DNA molecules showed full or no methylation at specific sites corresponding to “on” and “off” switches. Interestingly, methylation did not occur on each DNA molecule, resulting in a variable fraction of methylated chromosomes. This stochastic type of gene regulation was more common for rare genetic variants, which may suggest a role in human disease. —LMZ

*Science*, this issue p. 1354**CANCER****Waking up in a trap**

Cancer patients who have undergone successful treatment can experience relapse of their disease years or even decades later. This is because cancer cells that have disseminated beyond the primary tumor site enter a state of dormancy, where they remain viable but not proliferating. Eventually, by mechanisms that are poorly understood, these clinically undetectable cells “wake up” and form actively growing metastases. Studying mouse models, Albregues *et al.* found that sustained lung inflammation and the accompanying formation of neutrophil extracellular traps (NETs) could convert dormant cancer cells to aggressive lung metastases (see the Perspective by Aguirre-Ghiso). Awakening of these cells was associated with NET-mediated remodeling of the extracellular matrix and could

be prevented by an antibody against the remodeled version of a matrix protein called laminin-111. —PAK

*Science*, this issue p. 1353;

see also p. 1314

**ARCHAEOLOGY****Classic Maya civilization in detail**

Lidar (a type of airborne laser scanning) provides a powerful technique for three-dimensional mapping of topographic features. It is proving to be a valuable tool in archaeology, particularly where the remains of structures may be hidden beneath forest canopies. Canuto *et al.* present lidar data covering more than 2000 square kilometers of lowland Guatemala, which encompasses ancient settlements of the Classic Maya civilization (see the Perspective by Ford and Horn). The data yielded population estimates, measures of agricultural intensification, and evidence of investment in landscape-transforming infrastructure. The findings indicate that this Lowland Maya society was a regionally interconnected network of densely populated and defended cities, which were sustained by an array of agricultural practices that optimized land productivity and the interactions between rural and urban communities. —AMS

*Science*, this issue p. 1355;

see also p. 1313

**ORGANIC CHEMISTRY****Two ways out of an oxetane**

Oxetanes are highly reactive four-membered rings that contain three carbon atoms and an oxygen atom. Recently, they were implicated as transient intermediates in Lewis acid-catalyzed intramolecular metathesis reactions of ketones with olefins. Ludwig *et*

*al.* now report that by replacing the Lewis acid with a strong Brønsted acid, they can change the course of the oxetane ring-opening. In a so-called interrupted metathesis, the oxygen atom migrates and then departs through dehydration, while the remaining carbon framework cyclizes to form tetrahydrofluorene compounds. —JSY

*Science*, this issue p. 1363**ORGANIC CHEMISTRY****Arenes and amides from a single source**

Pharmaceutical synthesis often requires the formation of adjacent carbon-carbon and carbon-nitrogen bonds. Monos *et al.* present a method that delivers the carbon and nitrogen components in a single reagent, specifically, an aryl ring tethered through sulfur dioxide to an amide. A light-activated catalyst primes an olefin to react with the nitrogen, which in turn leads to migration of the aryl ring and loss of the sulfur bridge. The efficient room-temperature process is applicable to a variety of different arenes, including heterocycles. —JSY

*Science*, this issue p. 1369**EVO-DEVO****Hox code in segmentation and patterning**

*Hox* genes encode conserved transcription factors that are best known for their role in governing anterior-posterior body patterning in diverse bilaterian animals. He *et al.* used a combination of CRISPR mutagenesis and short hairpin RNA-based gene knockdowns to interrogate *Hox* gene function in a cnidarian, the sea anemone *Nematostella vectensis* (see the Perspective by Arendt). Four homeobox-containing genes constitute a molecular network that coordinately controls the morphogenesis of radial

endodermal segments and the patterning of tentacles. Thus, an ancient *Hox* code may have evolved to regulate both tissue segmentation and body patterning in the bilaterian-cnidarian common ancestor. —BAP

*Science*, this issue p. 1377;

see also p. 1310

**ANTIBODIES****Giving antibodies a boost**

Persistent immune activation during chronic infections is often associated with increased generation and deposition of immune complexes. The actions of antibody-based drugs can thus be severely impaired in individuals with chronic infections. Using lymphocytic choriomeningitis virus (LCMV) as a model of chronic infection, Wieland *et al.* examined how to enhance antibody functions in this setting. The ability of antibodies to deplete target cells was dependent on antigen expression levels. Furthermore, afucosylation of antibodies directed against CD4 and CD8 $\alpha$  enhanced their ability to deplete CD4<sup>+</sup> and CD8<sup>+</sup> T cells in mice persistently infected with LCMV. Whether afucosylation can be universally used to enhance antibody functions during chronic infections remains to be seen. —AB

*Sci. Immunol.* **3**, eaao3125 (2018).**BIOCHEMISTRY****Using kinase inhibitors on pseudokinases**

Pseudokinases are structurally similar to kinases but lack catalytic activity. Instead, pseudokinases often promote protein degradation by acting as scaffolds for ubiquitin ligases and their substrates. The pseudokinases TRIB1 and TRIB2 are implicated in leukemia and other cancers. Jamieson *et al.* and Foulkes *et al.* found that these proteins bound to clinically approved kinase inhibitors,



which induced structural changes that led to their degradation upon interacting with ubiquitin ligases. Thus, these drugs might be repurposed to block the function of TRIBs in cancer patients. —LKF

*Sci. Signal.* **11**, eaau0597, eaat7951 (2018).

## RESEARCH ARTICLE SUMMARY

## CANCER

# Neutrophil extracellular traps produced during inflammation awaken dormant cancer cells in mice

Jean Albrengues, Mario A. Shields, David Ng, Chun Gwon Park, Alexandra Ambrico, Morgan E. Poindexter, Priya Upadhyay, Dale L. Uyeminami, Arnaud Pommier, Victoria Küttner, Emilis Bružas, Laura Maiorino, Carmelita Bautista, Ellese M. Carmona, Phyllis A. Gimotty, Douglas T. Fearon, Kenneth Chang, Scott K. Lyons, Kent E. Pinkerton, Lloyd C. Trotman, Michael S. Goldberg, Johannes T.-H. Yeh, Mikala Egeblad\*

**INTRODUCTION:** Most cancer patients die from cancer that recurs after spreading to a different tissue, rather than from their original tumor. After successful treatment of the original tumor, cancer cells that have disseminated to other sites can undergo dormancy, remaining viable but not proliferating. In breast, prostate, and other cancers, cancer cells can remain dormant and clinically undetectable for years and even decades before recurring, or awakening, as metastatic cancer. Little is known about what might initiate cancer awakening, and this

in turn reduces our opportunities to prevent metastasis.

**RATIONALE:** Epidemiological studies have suggested that inflammation is linked to a higher risk of breast cancer recurrence after a period of clinical dormancy. Smoking, which causes chronic lung inflammation, is also associated with a higher risk of recurrence. However, whether inflammation can cause awakening is not clear. Inflammatory cells, such as neutrophils, can provide many differ-

ent signals that promote cancer progression. Neutrophils can kill harmful microorganisms by the release of neutrophil extracellular traps (NETs) into the extracellular space. NETs are scaffolds of DNA with associated cytotoxic proteins and proteases [e.g., neutrophil elastase (NE) and matrix metalloproteinase 9 (MMP9)]. NETs induced by bacteria or by cancer cells can promote metastasis, but the mechanism by which this occurs is not known. In this study, we tested whether NETs formed during lung inflammation could induce awakening.

**RESULTS:** We found that sustained experimental lung inflammation—induced by either

## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aao4227>

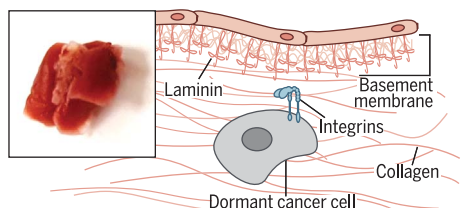
tobacco smoke exposure or nasal instillation of lipopolysaccharide (LPS)—converted dormant cancer cells to aggressive lung metastases in mice. Both types of sustained inflam-

mation also caused the formation of NETs. Inhibiting NET formation or digesting the NETs' DNA scaffold prevented conversion of single disseminated cancer cells to growing metastases in mouse models of breast and prostate cancer. The NET DNA bound to the extracellular matrix (ECM) protein laminin, thus bringing two NET-associated proteases, NE and MMP9, to their substrate. This in turn facilitated a sequential cleavage of laminin, first by NE and then by MMP9. The NET-mediated proteolytic remodeling of laminin revealed an epitope that triggered proliferation of dormant cancer cells through integrin activation and FAK/ERK/MLCK/YAP signaling. We generated a blocking antibody against NET-remodeled laminin, and this antibody prevented or reduced tobacco smoke exposure- or LPS-induced inflammation from awakening dormant cancer cells in mice.

**CONCLUSION:** Our data implicate NETs and NET-mediated ECM remodeling as critical mediators of inflammation-induced awakening in mouse models of dormancy. We propose that NETs awaken cancer by concentrating neutrophil proteases at the ECM protein laminin, allowing for sequential proteolytic remodeling of laminin and leading to integrin-mediated signaling in the cancer cells. Our findings set the stage for epidemiological studies to test possible links among inflammation or smoking, NETs, and recurrence after dormancy in human patients. If such links can be established, we envision that approaches similar to the ones used in mouse models in our study could be used to target NETs and their downstream effectors to reduce the risk of cancer recurrence in human patients. ■

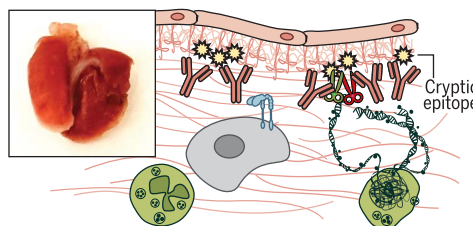
The list of author affiliations is available in the full article online.  
\*Corresponding author. Email: [egeblad@cshl.edu](mailto:egeblad@cshl.edu)  
Cite this article as J. Albrengues et al., *Science* 361, eaao4227 (2018). DOI: 10.1126/science.aao4227

## Undetectable disseminated dormant cancer

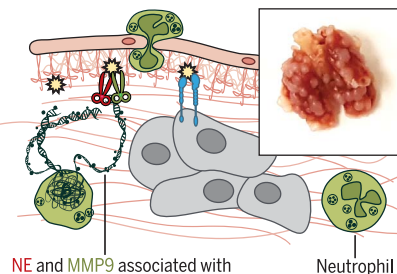


Inflammation caused by e.g. tobacco smoke exposure

## Awakening is prevented



## Awakening and cancer growth



NE and MMP9 associated with Neutrophil Extracellular Traps

Neutrophil

**NETs formed during lung inflammation awaken dormant cancer cells.** Lung inflammation—caused, for instance, by tobacco smoke exposure—leads to NET formation. Two proteases on the NETs sequentially cleave the extracellular matrix protein laminin. This cleavage generates an epitope that activates integrin signaling and proliferation of the previously dormant cancer cells. Antibodies blocking the NET-generated laminin epitope prevent inflammation-induced awakening.



## RESEARCH ARTICLE

## CANCER

# Neutrophil extracellular traps produced during inflammation awaken dormant cancer cells in mice

Jean Albrengues<sup>1</sup>, Mario A. Shields<sup>1</sup>, David Ng<sup>1</sup>, Chun Gwon Park<sup>2,3</sup>, Alexandra Ambrico<sup>1</sup>, Morgan E. Poindexter<sup>4</sup>, Priya Upadhyay<sup>4</sup>, Dale L. Uyeminami<sup>4</sup>, Arnaud Pommier<sup>1</sup>, Victoria Küttner<sup>1</sup>, Emilis Bružas<sup>1,5</sup>, Laura Maiorino<sup>1,5</sup>, Carmelita Bautista<sup>1</sup>, Ellese M. Carmona<sup>2,3</sup>, Phyllis A. Gimotty<sup>6</sup>, Douglas T. Fearon<sup>1,7,8</sup>, Kenneth Chang<sup>1</sup>, Scott K. Lyons<sup>1</sup>, Kent E. Pinkerton<sup>4</sup>, Lloyd C. Trotman<sup>1</sup>, Michael S. Goldberg<sup>2,3</sup>, Johannes T.-H. Yeh<sup>1</sup>, Mikala Egeblad<sup>1\*</sup>

Cancer cells from a primary tumor can disseminate to other tissues, remaining dormant and clinically undetectable for many years. Little is known about the cues that cause these dormant cells to awaken, resume proliferating, and develop into metastases. Studying mouse models, we found that sustained lung inflammation caused by tobacco smoke exposure or nasal instillation of lipopolysaccharide converted disseminated, dormant cancer cells to aggressively growing metastases. Sustained inflammation induced the formation of neutrophil extracellular traps (NETs), and these were required for awakening dormant cancer. Mechanistic analysis revealed that two NET-associated proteases, neutrophil elastase and matrix metalloproteinase 9, sequentially cleaved laminin. The proteolytically remodeled laminin induced proliferation of dormant cancer cells by activating integrin  $\alpha 3 \beta 1$  signaling. Antibodies against NET-remodeled laminin prevented awakening of dormant cells. Therapies aimed at preventing dormant cell awakening could potentially prolong the survival of cancer patients.

Most cancer patients die not from their original primary tumor but from metastases that arise in distant tissues. Often, metastatic disease occurs after a prolonged period of dormancy, when disseminated cancer cells are present but clinically undetectable (1). Disseminated cancer cells can remain dormant for years or even decades before recurring, or awakening, as metastatic cancer. T cells and natural killer cells can eliminate the disseminated cancer cells as they start proliferating, preventing them from reaching clinically detectable levels (2–5). In contrast, increased extracellular matrix (ECM) deposition and sprouting angiogenesis have been shown to trigger awakening and metastasis in experimental models (6, 7).

It is still unclear what triggers a change in the balance between signals that keep dissem-

inated tumor cells from growing and those that cause awakening and outright metastases. In breast cancer survivors, elevated plasma levels of C-reactive protein, a nonspecific marker of chronic inflammation, are associated with reduced disease-free survival (8), suggesting that inflammation may play a role in the switch between dormancy and metastasis. Inflammation has many causes: For example, smoking induces chronic inflammation in the lung, but the association between smoking and breast cancer risk has been controversial. Nevertheless, two recent, large, pooled analysis studies showed that current smoking or prior heavy smoking was significantly associated with an elevated risk of breast cancer recurrence and death from breast cancer (9, 10). In mice, tobacco smoke exposure increased lung metastasis by a factor of 2 (11).

Inflammation is commonly mediated by neutrophils [also called polymorphonuclear leukocytes (PMNs)], and these cells are critical for cancer cell awakening in experimental models (12). Still, it remains unclear how neutrophils cause awakening.

Neutrophils are well known for their ability to kill harmful microorganisms. They do so via (i) phagocytosis, whereby bacteria or fungi are engulfed and digested; (ii) degranulation of cytotoxic enzymes and proteases into the extracellular space; or (iii) the formation of neutrophil extracellular traps (NETs)—scaffolds of chroma-

tin with associated cytotoxic enzymes and proteases that are released into the extracellular space where they can trap microorganisms (13). NETs are generated through a signaling process that involves citrullination of histones by the protein arginine deiminase (PAD) 4 enzyme, chromatin decondensation, and disintegration of the nuclear membrane. Contents from the neutrophil's secretory granules—including neutrophil elastase (NE), cathepsin G (CG), and matrix metalloproteinase 9 (MMP9)—associate with the decondensed chromatin. Finally, the plasma membrane ruptures, and the protease-associated chromatin fibers are released into the extracellular space (14, 15). A growing body of evidence indicates a role for NETs, not just in infections but also in noninfectious inflammatory diseases (13), thrombosis (16, 17), and impaired wound healing in diabetes (18). NETs formed in response to systemic bacterial infection or after surgical stress promote cancer dissemination (19, 20). Using mouse models, we set out to identify how NET structures facilitate metastasis after a period of dormancy.

## Inflammation-activated neutrophils drive cancer cell awakening

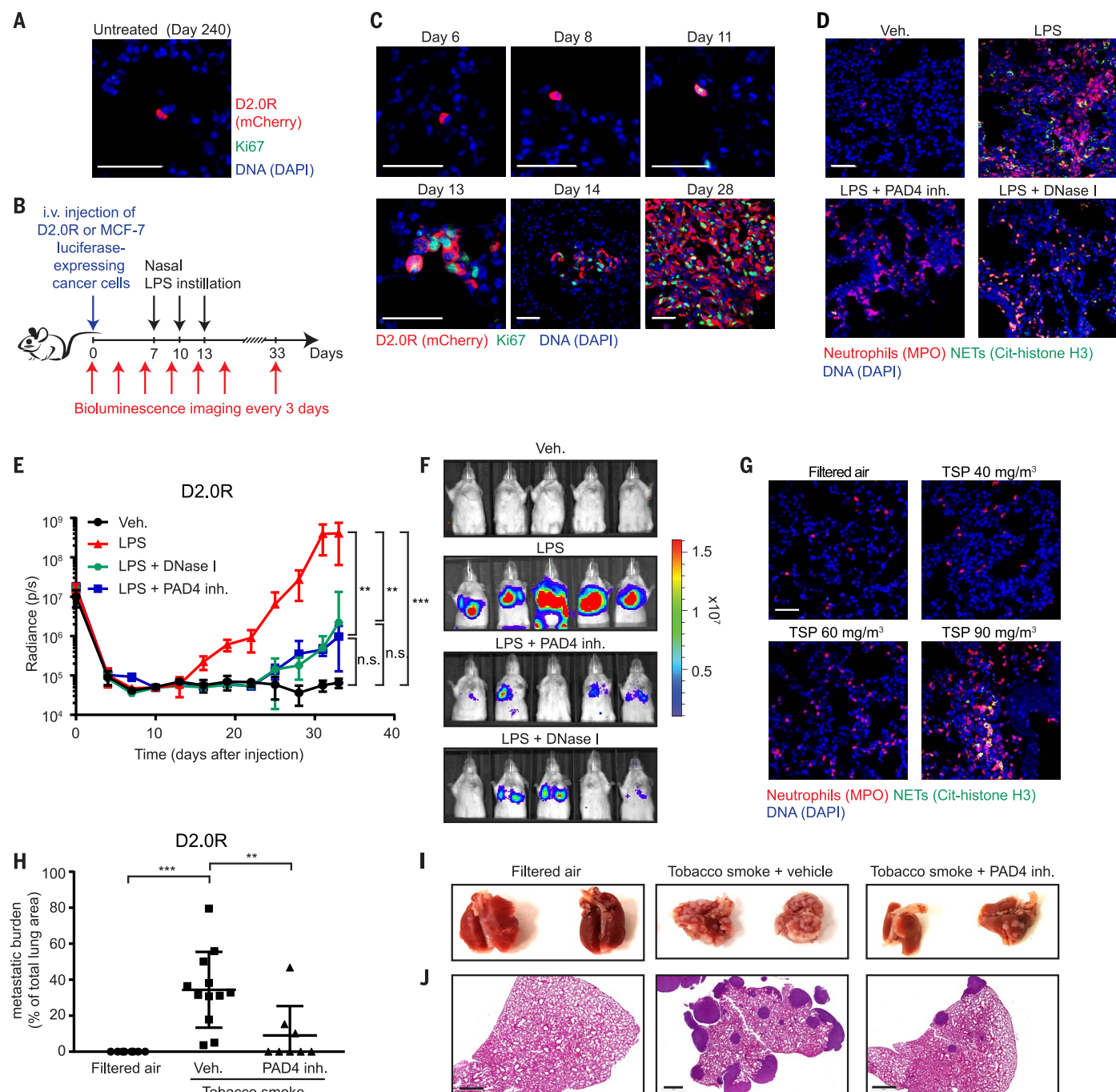
To determine whether local inflammation in the lung could directly drive awakening of disseminated, dormant cancer cells, we studied two models of dormancy. We injected luciferase- and mCherry-expressing breast cancer cells (murine D2.0R and human MCF-7 cell lines) intravenously into syngeneic BALB/c or nude mice, respectively. Tumors did not form, even 240 days after injection. Instead, single, nonproliferative cancer cells were found in the lungs (Fig. 1A and fig. S1, A and B) (7, 21). To explore the effect of lung inflammation on dormancy, we nasally instilled lipopolysaccharide [(LPS), also called endotoxin, a potent inducer of inflammation] into mice bearing dormant cancer cells. One LPS instillation, which models a short infection, did not awaken the dormant D2.0R and MCF-7 cancer cells; however, three injections, which model sustained, bacterially induced lung inflammation, led to aggressive lung metastasis (Fig. 1B and fig. S1, C to H). Small clusters of cancer cells appeared between the second and third instillations of LPS, indicating their escape from dormancy, whereas cancer cells remained as single, nonproliferative cells in control mice (Fig. 1, A to C, and fig. S1, A and B). The same results were observed when starting LPS instillation 1 month, instead of 7 days, after intravenous injection of D2.0R and MCF-7 cells (fig. S1, I to N), showing that dormant cells remained sensitive to external stimuli even at later time points.

LPS instillation caused marked neutrophil recruitment (movies S1 and S2), after both one and three instillations (fig. S2, A to C). Still, the relationship between the awakening of D2.0R proliferation and neutrophil recruitment and activation remained unclear. To assess these dynamics, we used confocal intravital lung imaging (22) with a fluorescence ubiquitination-based

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. <sup>2</sup>Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

<sup>3</sup>Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02215, USA. <sup>4</sup>Center for Health and the Environment, University of California, Davis, Davis, CA 95616, USA. <sup>5</sup>Watson School of Biological Sciences, Cold Spring Harbor, NY 11724, USA. <sup>6</sup>Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. <sup>7</sup>Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, UK. <sup>8</sup>Meyer Cancer Center, Weill Cornell Medical College, New York, NY 10021, USA.

\*Corresponding author. Email: egeblad@cshl.edu



**Fig. 1. NETs promote dormant cancer cell awakening after sustained lung inflammation.** (A and B) D2.0R cells stayed dormant in the lungs of mice for months. (A) Representative micrographs of D2.0R cells (red, mCherry) in Ki67 (green)– and DAPI (blue)–stained lung sections from untreated mice at day 240. Scale bar, 50  $\mu$ m. (B) Schematic showing experimental design. (C) Sustained lung inflammation promoted awakening. Representative immunostaining of lungs for D2.0R cells (red, mCherry) and Ki67 (green) with DAPI (blue), at indicated time after injection. Scale bars, 50  $\mu$ m. (D) Nasal LPS instillation induced NETs in lungs. Images show representative immunostaining for myeloperoxidase (MPO, red), citrullinated histone H3 (Cit-histone H3, green), and DAPI (blue) in the lungs of mice treated as indicated. Scale bar, 50  $\mu$ m. Veh., vehicle; inh., inhibitor. (E and F) Targeting NETs reduced inflammation-induced awakening. (E) Mice with D2.0R cells, treated as indicated, were

monitored by bioluminescence imaging (BLI) ( $n = 10$  mice for vehicle and LPS groups;  $n = 5$  mice for DNase I and PAD4 inhibitor groups; means  $\pm$  SD). (F) Representative BLI images at day 33. (G) Exposure to tobacco smoke (TSP, total suspended particles) induced NETs in the lungs. Images show representative immunostaining for myeloperoxidase (red), citrullinated histone H3 (green), and DAPI (blue) in the lungs of mice treated as indicated. Scale bar, 50  $\mu$ m. (H to J) Targeting NETs reduced smoking-induced awakening. (H) Mice with D2.0R cells were treated as indicated, and lung metastatic burden was quantified from hematoxylin and eosin (H&E) staining ( $n = 8$  mice for filtered air and tobacco smoke + PAD4 groups;  $n = 12$  mice for tobacco smoke + vehicle group; mean  $\pm$  SD). (I) Lungs from mice treated as indicated at day 30. (J) Representative images of H&E staining of lungs from animals described in (H) at day 30. Scale bars, 700  $\mu$ m. \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; n.s., not significant.

cell cycle indicator (FUCCI) system, which fluorescently labels cells red in G<sub>0</sub> and G<sub>1</sub> phases and green in S, G<sub>2</sub>, and M phases (fig. S2D) (23). Neutrophil infiltration and activation were tracked using the LysM-enhanced green fluorescent protein (EGFP) transgene and the NE 680 FAST probe, which fluoresces after NE cleavage.

In control mice instilled with phosphate-buffered saline (PBS), NE activity was low, and all D2.0R cells were red in the G<sub>0</sub> and G<sub>1</sub> cell cycle phases at days 8, 11, and 21 (movies S3 to S5). In contrast, beginning 5 days after the first LPS instillation (day 11), D2.0R cells became yellow, indicating that they had entered the G<sub>1</sub>/S transition of the cell cycle, and this transition correlated with high neutrophil recruitment and NE activity (movie S6 and fig. S2, E and F). We detected small clusters of proliferating cells at day 14 and established proliferative lung metastasis at day 21 (movies S7 and S8 and fig. S2F).

We next investigated whether neutrophils contributed to awakening of dormant cancer cells. Indeed, neutrophil depletion completely prevented inflammation-induced awakening of dormant D2.0R and MCF-7 cancer cells (fig. S3, A to I). Thus, sustained inflammation induced by LPS causes dormant cancer cells to reenter the cell cycle, and this effect requires neutrophils.

### NETs awaken dormant cancer cells in mice

After LPS instillation, we detected numerous NETs in the lungs within 4 hours, persisting 24 hours later; in contrast, NETs were absent in normal lung tissue (Fig. 1D and fig. S3, J and K). NETs were also present in the plasma after LPS treatment, as assessed by double-stranded DNA (dsDNA) assays or an enzyme-linked immunosorbent assay (ELISA) against DNA-bound NE (fig. S3, L and M).

To determine whether NETs contributed to cancer cell awakening, we blocked NET formation with a PAD4 inhibitor or digested the NET DNA scaffold with deoxyribonuclease (DNase) I (free or coated on nanoparticles) (Fig. 1D and fig. S4, A to F). Both treatments prevented or decreased LPS-induced awakening of dormant D2.0R and MCF-7 cancer cells (Fig. 1, E and F, and fig. S4, G to M). Results were similar when LPS instillation was started 1 month after intravenous injection of the cancer cells (fig. S1, I to N), showing that NETs remained a powerful external stimulus of awakening, even at later time points. DNase I was most effective as a preventative treatment, immediately before LPS instillation, but also reduced end-point metastatic burden when administered at day 14, after the appearance of small clusters of cancer cells in the lungs (fig. S4N). In all cases, the NET-targeting treatments reduced neutrophil recruitment (fig. S4, O and P), suggesting that NETs promote further inflammation.

The cancer cell dormancy models described above were generated by intravenous injection of cancer cells, so we used two additional approaches to explore whether NETs could also induce proliferation after natural dissemina-

tion of cancer cells. First, we used the RapidCaP prostate cancer model (24), where prostate cancer develops from normal epithelial tissues within an intact organ. The model has a low incidence of lung metastasis (24), and we observed no lung metastasis in control mice during the observation period, despite the presence of single, disseminated cancer cells in the lungs. In contrast, LPS instillations resulted in lung metastasis in three of five mice. However, zero of five mice developed macroscopic metastasis when NET formation was blocked with the PAD4 inhibitor, even after three additional LPS instillations (fig. S5, A to C;  $P = 0.02$ ). In the second approach, we allowed MCF-7 cells to form primary mammary tumors and spontaneously disseminate. We then resected those primary tumors. Under these conditions also, LPS-induced NETs awakened the cancer cells (fig. S5, D to H).

Repeated LPS instillation models a sustained bacterially induced inflammation. Smoking similarly induces chronic lung inflammation and has been associated with increased risk of breast cancer recurrence (9, 10). To examine NETs and cancer cell dormancy under these conditions, we exposed mice to three different concentrations of tobacco smoke for 3 weeks. This resulted in a dose-dependent increase in neutrophil infiltration in the lungs. NETs formed in the lungs at the highest tobacco smoke exposure level (a level almost equivalent to moderate, active smoking and well above secondhand exposure to tobacco smoke) (Fig. 1G and fig. S5, I and J). NETs, as well as LPS, were also found in the plasma of the tobacco smoke-exposed mice (fig. S5, K to M), suggesting that tobacco exposure causes systemic exposure to both agents.

We next tested whether the tobacco smoke could awaken dormant D2.0R cells. Indeed, aggressive metastasis developed when mice were exposed to tobacco smoke at the level that induced NET formation, and this was prevented by PAD4 inhibitor treatment. Cells remained dormant in mice exposed to filtered air (Fig. 1, H to J, and fig. S5N). Thus, NETs formed during inflammation induced the awakening of dormant cancer cells in multiple mouse models.

### NETs awaken slow-cycling cells in vitro in ECM models

To determine how NETs induced awakening of dormant cancer cells, we turned to three-dimensional (3D) culture systems, where the isolated effect of NETs could be tested in the absence of the many cell types present in lungs. Both D2.0R and MCF-7 cells become slow cycling when cultured on basement membrane matrix (matrigel) (25), whereas metastatic D2.0A1 cells, isolated from the same mammary lesion as the D2.0R cells, proliferated (fig. S6, A and B). To generate NETs, we cultured freshly isolated neutrophils and stimulated them with LPS, phorbol 12-myristate 13-acetate (PMA), or *N*-formyl-methionyl-leucyl-phenylalanine (fMLP) (Fig. 2A and fig. S6, C and D). We found that NETs were also induced by coculturing neutro-

phils with metastatic D2.0A1 cells but not with the dormant D2.0R cells NETs (fig. S6C). This is similar to our previous findings using another pair of cell lines (22), the metastatic and NET-inducing 4T1 and the nonmetastatic and non-NET-inducing 4T07 cell lines. Finally, to compare the effects of NETs with that of neutrophil degranulation, we induced degranulation by culturing with complement 5a (c5a). We next tested the effect of the NET-containing conditioned media (CM) from the different neutrophil culture conditions on luciferase-expressing D2.0R and MCF-7 cancer cells cultured on matrigel (Fig. 2A). NET-containing CM led to the awakening and proliferation of dormant cancer cells, whereas CM from degranulated or inactivating conditions had no effect (Fig. 2, B to D, and figs. S6E and S7, A to C). LPS, PMA, fMLP, and c5a added directly to the cancer cells had no effect on awakening in our 3D culture model systems, nor did CM from D2.0A1 or D2.0R cancer cells (fig. S7, D and E). Treatment of D2.0R cells with LPS before intravenous injection also did not lead to metastasis (fig. S7, F and G), further suggesting that it was the NETs in the CM and not LPS that activated the slow-cycling cells. When the neutrophil cultures were treated with a PAD4 inhibitor or DNase I during NET-activating conditions, NETs did not form or were digested (fig. S6, C and D), and the CM no longer induced awakening of the dormant cancer cells (Fig. 2D and fig. S7B).

Given that nicotine can induce NETs in vitro (26), we next tested the effect of cigarette smoke extract (CSE) on neutrophils. We found that CSE induced NETs in vitro in a dose-dependent manner and that this NET-containing CM could also awaken the slow-cycling cells (Fig. 2, E and F, and fig. S7, H and I). Because we had detected LPS in the plasma of tobacco smoke-exposed mice, we assessed the effect of tauroldine, a compound that neutralizes the effect of LPS (27). Tauroldine and PAD4 inhibition both reduced CSE-induced formation of NETs, and the CM was no longer able to awaken the cancer cells (Fig. 2, E and F, and fig. S7, H and I), suggesting that LPS present in CSE drives smoking-induced NETs. CSE had no effect on awakening when it was added directly to the cancer cells in vitro, and it caused cancer cell death at the highest concentrations (fig. S7J). Altogether, these results show that NETs induced by multiple means promote cancer cell awakening in vitro in the presence of an artificial ECM and in the absence of lung cells, other immune cells, and vasculature.

### NET-associated proteases awaken cancer cells

Degradation of the ECM, activation of cell surface receptors, and the release and activation of cytokines and growth factors by proteases are necessary steps in metastasis (28, 29). We hypothesized that NET-associated protease activity was responsible for the effects of NETs on awakening. Using our 3D culture systems, we analyzed the effects of inhibiting three major

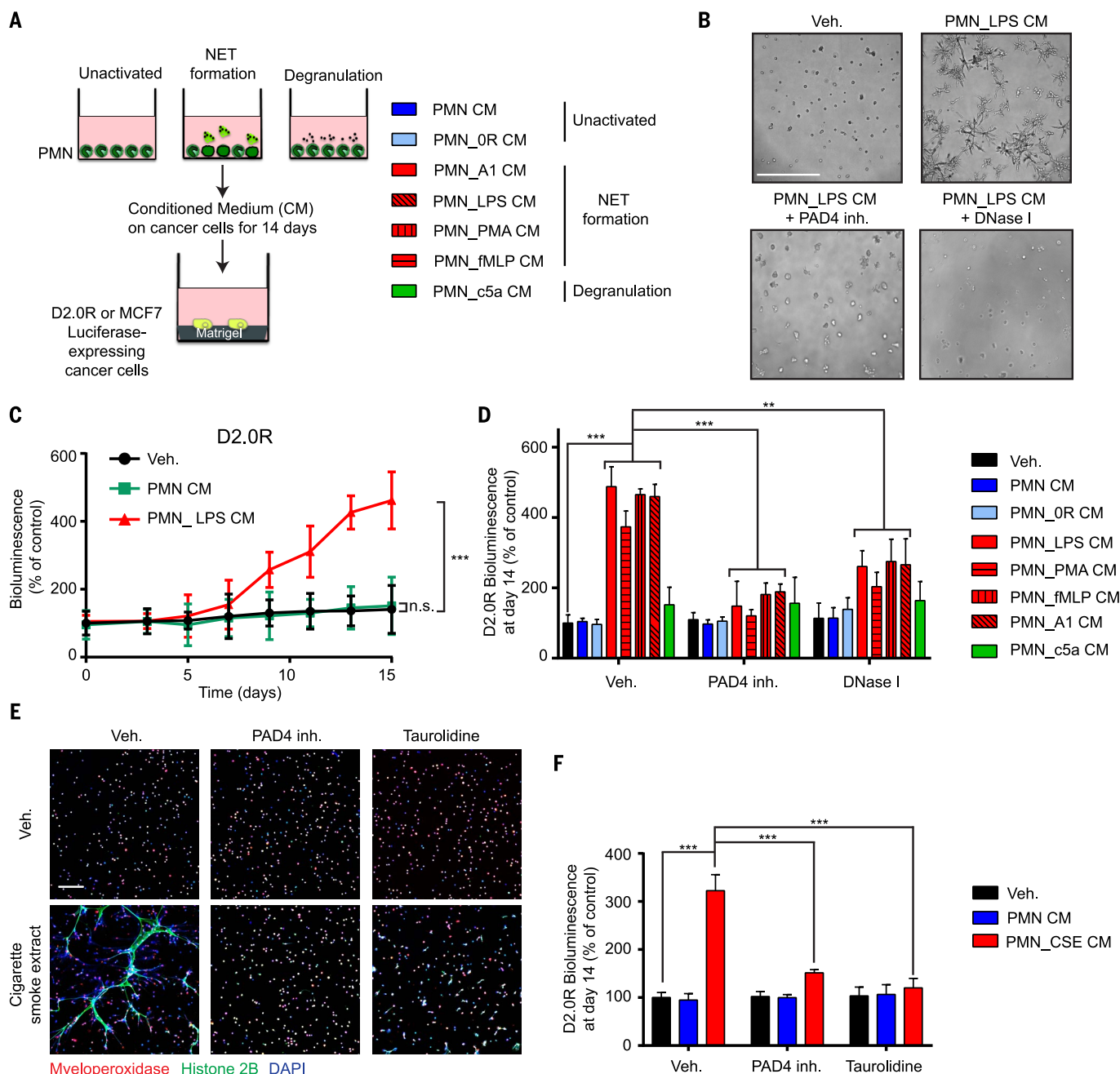


NET-associated proteases: CG, NE, and MMP9. Inhibiting NE or MMP9, but not CG activity, in the NET-containing CM prevented awakening of the slow-cycling cancer cells (Fig. 3A and fig. S8, A to D). We also isolated neutrophils from NE and MMP9 knockout mice and stimulated

them to form NETs. CM from these cultures did not induce awakening (fig. S8, E and F). We found that neutrophils, but not cancer cells, were the main source of these proteases. NET-containing CM did not induce expression of the proteases in cancer cells (fig. S9, A to C), and

short hairpin RNA (shRNA)-mediated knock-down of NE and MMP9 in D2.0R cells did not affect the capacity of these cells to be awakened by NET-containing CM in vitro (fig. S9, D to F).

In vivo, inhibiting NE and MMP9 activity also prevented LPS-induced cancer recurrence



**Fig. 2. NETs induce awakening in vitro in the absence of other host cells.** (A) Experimental design for inducing NETs or degranulation neutrophils (PMNs). CM was collected 20 hours after activation and added to cultures of luciferase-expressing cancer cells plated on top of matrigel. (B to D) NET-containing CM induced awakening of D2.0R cells on matrigel. (B) Representative images of 3D cultures at day 15, treated as indicated. Scale bar, 250  $\mu$ m. (C) BLI quantification over 15 days with indicated treatments ( $n = 3$ ; means  $\pm$  SD). (D) BLI signal 14 days after indicated treatments. PAD4 inhibitor and DNase I were used during

neutrophil culture to block NET formation and digest NET formation, respectively ( $n = 3$ ; means  $\pm$  SD). (E and F) Cigarette smoke extract (CSE) induced the formation of NETs and subsequent awakening in vitro. (E) Images shown immunostaining of mouse neutrophils cultured as indicated. DAPI (blue), anti-MPO (red), and anti-histone H2B (green) staining were used to assess NET formation. Scale bar, 100  $\mu$ m. (F) BLI signal 14 days after indicated treatments. Taurolidine and PAD4 inhibitor were used during neutrophil culture to block NET formation ( $n = 3$ ; means  $\pm$  SD). \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; n.s., not significant.

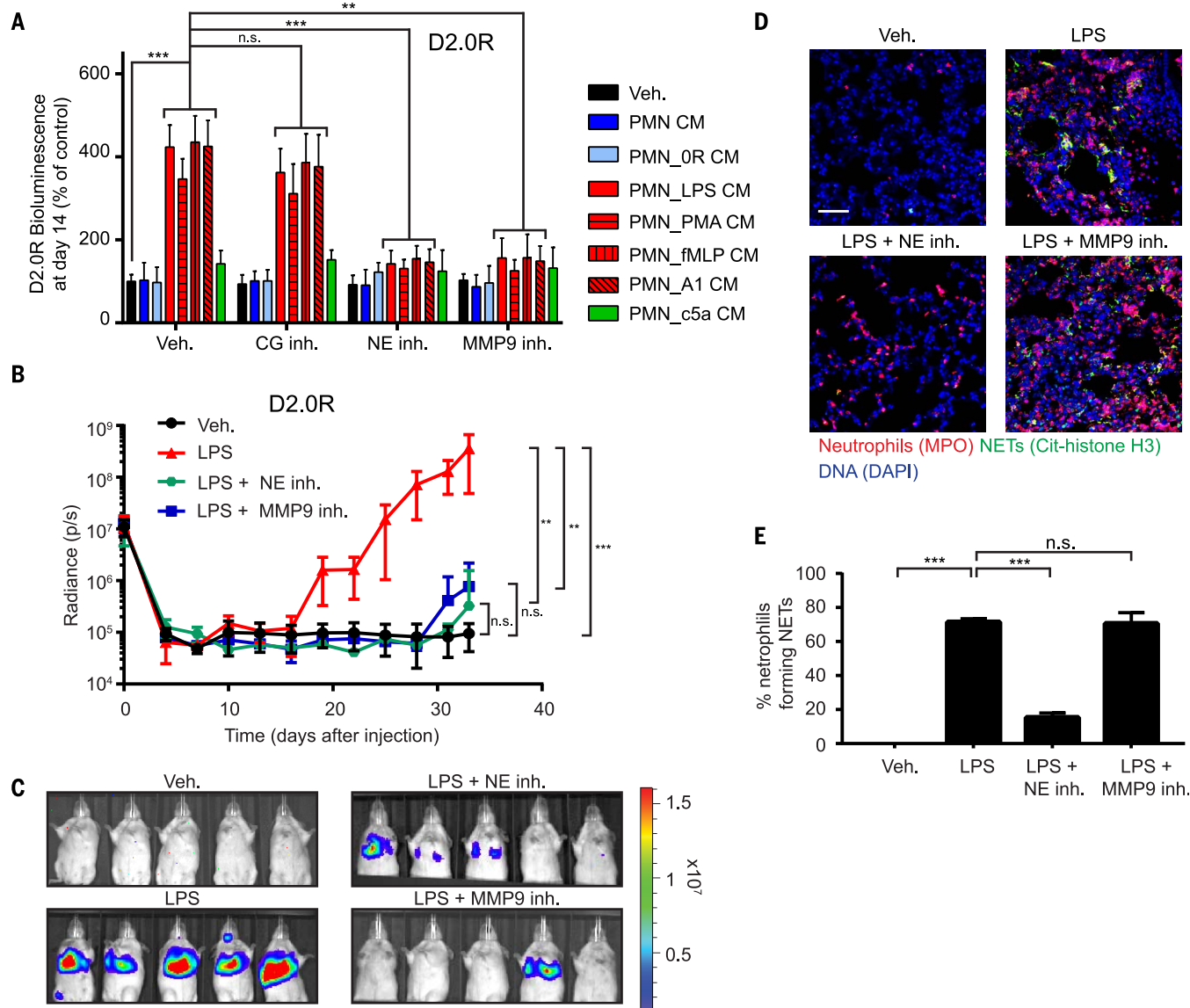
(Fig. 3, B and C, and fig. S9, G to J). NE inhibition prevented NET formation in vivo, consistent with the known requirement for NE activity during NET formation (30, 31). In contrast, NETs still formed in vivo after MMP9 inhibition, but cells remained dormant, suggesting that the activity of this NET protease is critical for awakening (Fig. 3, D and E). Similar to the prevention of NET formation by PAD4 inhibition, inhibition of NE, but not MMP9, also reduced neutrophil recruitment to the lungs after LPS-induced inflammation (fig. S9K).

To determine whether the NET proteases induced awakening by acting directly on the cancer cells or by altering their microenvironment, we preincubated matrigel with NET-containing CM before cancer cell plating (fig. S9L). This was sufficient to induce awakening but required both NE and MMP9 activity (fig. S9M, red bars). If the proteases were inhibited during cancer cell culture, after NET-mediated matrigel remodeling had already occurred, then awakening was not inhibited (fig. S9M, peach-colored bars). These data suggest that NET-associated NE and MMP9

induce awakening of dormant cancer cells through ECM remodeling.

### NETs awaken cancer cells by laminin remodeling

We next sought to identify the ECM substrate of NE and MMP9 during NET-mediated awakening of dormant cancer cells. According to the manufacturer, matrigel is composed mainly of laminin-111 (60%), collagen type IV (30%), and entactin (8%). Proteolysis of laminin-111, which consists of the  $\alpha$ -1,  $\beta$ -1, and  $\gamma$ -1 laminin chains,



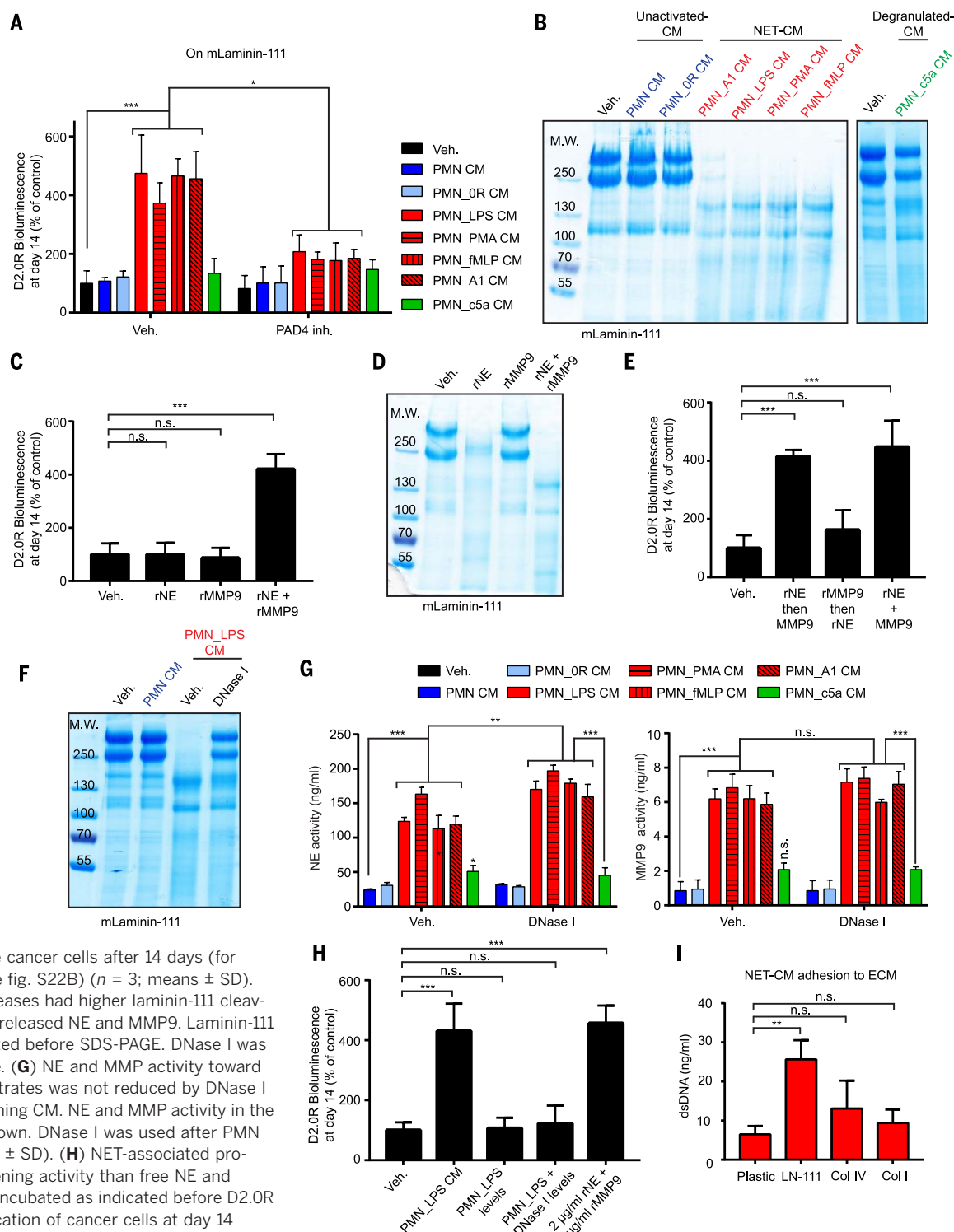
**Fig. 3. NET-associated NE and MMP9 induce awakening from dormancy through ECM remodeling.** (A) NE and MMP9 were required for NET-induced awakening of D2.0R cells in vitro. BLI signal of luciferase-expressing cells was measured 14 days after indicated treatments. CG, NE, and MMP9 inhibitors were used during cancer cell culture ( $n = 3$ ; means  $\pm$  SD). (B and C) NE and MMP9 activity were required for LPS-induced awakening in vivo. (B) Mice with D2.0R cells were treated as indicated and monitored by BLI ( $n = 10$  mice for vehicle and LPS

groups;  $n = 5$  mice for NE inhibitor and MMP9 inhibitor groups; means  $\pm$  SD). (C) Representative BLI images at day 33. (D and E) Inhibition of NE, but not MMP9, prevented formation of NETs after LPS-induced lung inflammation. (D) DAPI (blue), anti-MPO (red), and anti-citrullinated histone H3 (green) staining were used to assess NET formation in the lungs of mice treated as indicated. Scale bar, 50  $\mu$ m. (E) Quantification of NET-forming neutrophils ( $n = 3$  mice; means  $\pm$  SD). \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; n.s., not significant.

# Fig. 4. NETs facilitate laminin-111 remodeling.

**(A)** Culturing cancer cells on laminin-111 allowed NET-dependent control of dormancy and awakening. BLI signal of luciferase-expressing D2.OR cells was measured after 14 days under indicated conditions (PAD4 inhibitor was used during PMN culture,  $n = 3$ ; means  $\pm$  SD). **(B)** NET-containing CM cleaved laminin-111. Cleavage of laminin-111 after incubation with indicated CM was detected by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), under reducing and denaturing conditions, and coomassie blue staining. M.W., molecular weight (in kilodaltons). **(C to E)** Both NE and MMP9 were required for awakening and laminin-111 cleavage. Laminin-111 was incubated with recombinant proteases alone (C) or added sequentially (E), and D2.OR cells were cultured on the re-modeled matrix [(C) and (E)] or laminin-111 was analyzed by SDS-PAGE (D). BLI

was used to quantify the cancer cells after 14 days (for experimental design, see fig. S22B) ( $n = 3$ ; means  $\pm$  SD). **(F)** NET-associated proteases had higher laminin-111 cleavage activity than DNase-released NE and MMP9. Laminin-111 was incubated as indicated before SDS-PAGE. DNase I was added after PMN culture. **(G)** NE and MMP activity toward soluble fluorescent substrates was not reduced by DNase I treatment of NET-containing CM. NE and MMP activity in the indicated PMN CM is shown. DNase I was used after PMN activation ( $n = 3$ ; means  $\pm$  SD). **(H)** NET-associated proteases had higher awakening activity than free NE and MMP9. Laminin-111 was incubated as indicated before D2.OR cell culture. BLI quantification of cancer cells at day 14 ( $n = 3$ ; means  $\pm$  SD) is shown. **(I)** NETs bound preferentially to laminin-111. CM from neutrophils stimulated with LPS to form NETs was incubated on plastic or ECM-coated plates as indicated. dsDNA bound to ECM proteins was quantified ( $n = 3$ ; means  $\pm$  SD). \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$ ; n.s., not significant.



has previously been linked with regulation of proliferation and cellular architecture in 3D cultures (32). Both D2.OR and MCF-7 cells remained slow cycling when cultured on purified laminin-111 in vitro (fig. S10, A and B) but were awakened by NET-containing CM (Fig. 4A, and fig. S10, C to E).

To test whether NET-associated proteases could cleave laminin, we used commercially available purified laminin-111. Whereas NET-containing CM led to laminin-111 cleavage, CM from degranulating or nonactivating conditions did not (Fig. 4B and fig. S10F). Using recombinant (r) NE and MMP9, we found that both proteases

in combination were necessary to induce cancer cell awakening (Fig. 4C and fig. S10, G to I). Laminin-111 was cleaved by rNE alone but not by rMMP9; however, when both proteases were present, the cleavage fragments were different (Fig. 4D and fig. S10J). Similarly, when laminin-111 was treated with NET-containing CM, cleavage



was completely blocked after NE inhibition but was only partially blocked after MMP9 inhibition (fig. S10K). Therefore, we hypothesized that laminin-111 is cleaved sequentially—first by NE and subsequently by MMP9—to trigger awakening. Indeed, proteolysis of laminin-111 first by rNE and then by rMMP9, but not in the reverse order, led to cancer cell awakening (Fig. 4E and fig. S10G).

Our *in vitro* experiments suggest that laminin-111 is the primary substrate of NET proteases in 3D culture. However, laminin-111 has not consistently been detected in adult lungs. Therefore, we assessed the expression of laminin in lung tissue. We confirmed the presence of mRNA coding for the three laminin-111 chains by quantitative polymerase chain reaction (qPCR) and *in situ* hybridization, and we used multiple antibodies for immunofluorescence and Western blot analyses to detect laminin-111 protein (fig. S11, A to D). Laminin-111 mRNA expression was not affected by nasal LPS instillation, but cleaved laminin-111 was readily detectable in lung tissue lysate after inducing inflammation with LPS (fig. S11, A and D). A polyclonal anti-laminin-111 antibody was used for Western blot and immunofluorescence analysis; thus, we could not exclude the possibility that this reagent recognized not only laminin-111 but also other laminin isoforms containing  $\alpha$ -1,  $\beta$ -1, or  $\gamma$ -1. Hence, we tested whether NET-containing CM also could awaken cancer cells cultured on other laminin isoforms. We found that laminin-211, -411, and -511 were all sufficient to support NET-induced awakening (fig. S11E), which suggests that awakening can occur in any tissue containing one of these laminins—e.g., at the perivascular niche.

### NET DNA acts as a proteolysis scaffold

Our data showed that proteolytic remodeling is required for cancer cell awakening; yet, digesting NET DNA with DNase I also prevented awakening *in vivo* and *in vitro* (Figs. 1, E and F; and 2, B and D; and figs. S4, G to I, L, and M; and S7B). Consistently, laminin-111 was readily cleaved by incubation with NET-containing CM, and this cleavage was prevented by DNase I (Fig. 4F). Nevertheless, DNase I digestion did not reduce the protease activity against soluble fluorescent substrates (Fig. 4G). However, adding recombinant proteases at the activities measured in the NET-containing CM (120 ng/ml rNE, 6 ng/ml MMP9) or in NET-containing CM after DNase I treatment (160 ng/ml rNE, 6 ng/ml MMP9) had no effect on awakening. Instead, much higher concentrations of the recombinant proteases (2  $\mu$ g/ml each, the concentrations used in Fig. 4, C to E, and fig. S10, H to K, in accordance with the literature) were necessary to induce cancer cell awakening (Fig. 4H). We found areas of NE colocalizing with MMP9 on the DNA fibers of NETs (fig. S11F). Furthermore, NET DNA preferentially bound to laminin-111 over other tested ECM proteins (Fig. 4I and fig. S11G). These findings suggested that the NET DNA scaffold allowed the NET proteases to cleave their substrate more efficiently than when the proteases

freely diffused because the DNA binds to laminin and because the two proteases colocalize to the same DNA scaffold.

Thrombospondin-1 (TSP-1), a large glycoprotein present in the basement membrane surrounding mature blood vessels, regulates cancer cell dormancy (6). TSP-1 secretion by bone marrow-derived Gr1<sup>+</sup> cells (which include neutrophils) generates a metastasis-resistant microenvironment, which can be overcome through neutrophil-mediated proteolysis of TSP-1 (33, 34). We found that TSP-1 was also a substrate for NET-associated NE and MMP9 *in vitro* (fig. S11H) and that it was degraded *in vivo* after LPS-induced lung inflammation (fig. S11I). TSP-1 degradation alone did not convert slow-cycling cells to proliferating cells *in vitro* (fig. S11J). However, intact TSP-1 decreased proliferation caused by cleaved laminin-111 (fig. S11, K and L). Our data suggest that TSP-1 modulates cancer cell awakening caused by proteolytic remodeling of laminin-111.

### NET-remodeled laminin activates integrin $\alpha$ 3 $\beta$ 1 signaling

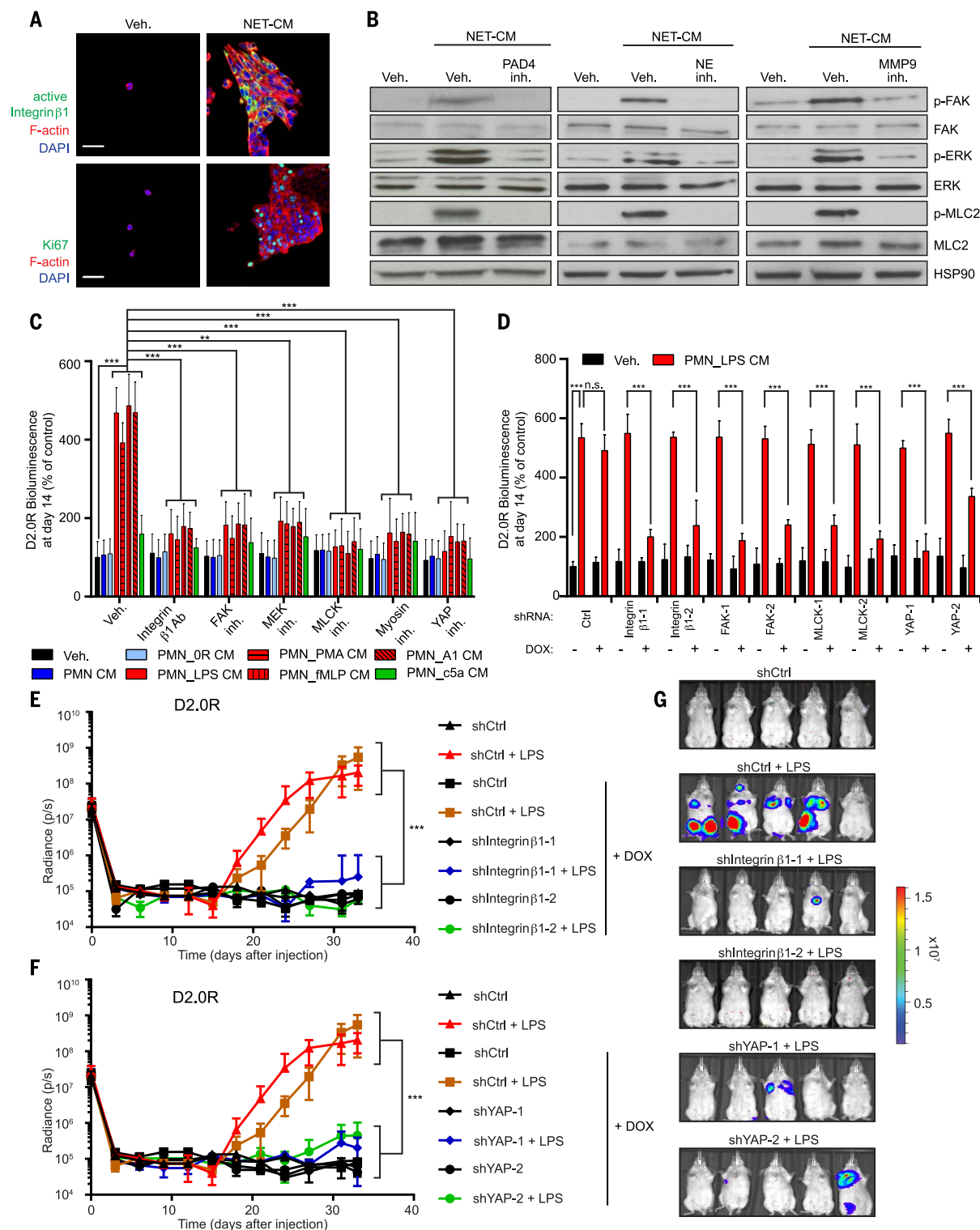
We next focused on how NET remodeling of laminin-111 led to cancer cell awakening. Integrins are cell surface ECM receptors that recognize conformational changes in the ECM, and signaling through integrin  $\beta$ 1 has previously been shown to induce awakening (7, 35–37). Consistently, a marker for cell proliferation, Ki67, was associated with integrin  $\beta$ 1 activity and reorganization of the actin cytoskeleton after incubating D2.0R or MCF-7 cells with NET-containing CM or rNE and rMMP9 (Fig. 5A and fig. S12, A and B). The integrin  $\beta$ 1 outside-in signaling pathway can activate the downstream mediators focal adhesion kinase (FAK), MAP kinase ERK kinase (MEK), extracellular signal-regulated kinase (ERK), myosin light chain kinase (MLCK), myosin light chain 2 (MLC2), and yes-associated protein (YAP). Accordingly, NET-induced awakening led to activation of FAK, ERK, and MLC2, which required NE and MMP9 activity (Fig. 5B). Moreover, targeting the integrin  $\beta$ 1 outside-in signaling pathway at different signaling points using chemical inhibitors or RNA interference inhibited NET-induced awakening *in vitro* (Fig. 5, C and D, and fig. S12, C and D). Inhibition of integrin  $\beta$ 1, FAK, MEK, MLCK, myosin, and YAP activities all led to the loss of FAK, ERK, and MLC2 phosphorylation upon stimulation with NET-containing CM (fig. S12, E to K), in accordance with the positive-feedback loop within this signaling pathway (38). To identify which  $\beta$ 1-containing integrin pair was involved in cancer cell awakening, we tested D2.0R cells for expression of  $\alpha$  integrins known to associate with integrin  $\beta$ 1 (fig. S13A) (39). Through shRNA mediated knockdown (fig. S13B), we found that awakening of dormant cancer cells by NETs required integrin  $\alpha$ 3, integrin  $\beta$ 1, and YAP in the cancer cells (Fig. 5, E to G, and fig. S13, C and D).

LPS-induced lung inflammation resulted in laminin cleavage that was detectable in lung lysate, and this cleavage depended on neutrophil recruitment, NET formation, and NE and

MMP9 activities (fig. S14A). Laminin cleavage was blocked after NE inhibition but was only partially blocked after MMP9 inhibition, consistent with sequential laminin cleavage by NE and MMP9 *in vivo*. However, no discernable changes in laminin immunofluorescent staining patterns were observed in lungs after LPS-induced inflammation (fig. S10H). Additionally, under nonreducing and nondenaturing conditions, we did not observe many laminin-111 fragments after rNE and rMMP9 cleavage but rather a reduction in its apparent molecular weight (Fig. 6A). These results suggest that NET-mediated proteolysis induced a change in the 3D structure of laminin. We hypothesized that a new laminin-111 epitope was revealed after NET-mediated proteolysis and that the dormant cancer cells sensed this new epitope through integrin  $\alpha$ 3 $\beta$ 1, leading to cancer cell proliferation. To test this idea, we generated monoclonal antibodies (mAbs) from rats immunized against purified rNE and rMMP9 cleaved laminin-111 (fig. S14B). Three laminin-111-recognizing antibody clones (Ab19, Ab25, and Ab28) strongly blocked NET-induced integrin  $\beta$ 1 activation and subsequent cancer cell awakening *in vitro* (Fig. 6B and fig. S14C). Of these antibodies, Ab28 specifically recognized NET-remodeled laminin-111 but not intact laminin, whereas Ab19 and Ab25 recognized both laminin-111 forms (fig. S14, D and E). We cannot exclude the possibility that other laminin isoforms share the antibody epitopes after NET-mediated remodeling and also are recognized by the antibodies.

Using Ab28, we detected the NET-remodeled laminin epitope in inflamed lungs. It was not present after neutrophil depletion or after inhibition of NET formation or NE or MMP9 activities (Fig. 6C and fig. S14F). The epitope was also present in lung tissue from tobacco smoke-exposed mice but only at an exposure level that induced NETs (Fig. 6D). Moreover, the NET-remodeled laminin epitope was only detectable in lungs after three LPS instillations, the number of doses required to awaken cancer cells (fig. S15A).

The NET-remodeled laminin epitope was detectable in the same lung regions as the NETs (fig. S15A). Furthermore, using D2.0R cells expressing the FUCCI cell cycle reporter (fig. S2D), we observed that all green cells (indicating S, G<sub>2</sub>, or M cell cycle phase) or cells that were part of a cluster were located adjacent to remodeled laminin, whereas cells close to intact laminin all remained red (indicating G<sub>0</sub> or G<sub>1</sub> cell cycle phase) (fig. S15, B and C). This colocalization pattern suggested that the NET-remodeled laminin epitope was driving the awakening of cancer cells *in vivo*. To test this, we engineered rat monoclonal antibodies as chimeric mouse immunoglobulin G 2a (IgG2a) antibodies (chiAbs), which can be used in mice without eliciting an immune response. ChiAb28 inhibited cancer cell awakening *in vitro* (fig. S15D) and in both *in vivo* systems (LPS- and tobacco smoke-induced inflammation) (Fig. 6, E to I, and fig. S15, E and F). Together, our data identify a mechanism by which sustained lung inflammation, induced by either tobacco smoke or LPS exposure, can lead to



**Fig. 5. Integrin  $\beta 1$  signaling pathway regulates NET-induced awakening.**

(A) NETs activated integrin  $\beta 1$  and proliferation in dormant cancer cells. D2.0R cells were cultured on matrigel-coated 0.2-kPa hydrogel stained for active integrin  $\beta 1$  (green, top) or Ki67 (green, bottom), F-actin (red), and DAPI (blue). Scale bar, 50  $\mu\text{m}$ . (B) NET-containing CM activated the integrin  $\beta 1$  signaling pathway. D2.0R cells were cultured on matrigel for 10 days under the indicated conditions and analyzed for phospho-FAK, phospho-ERK1/2, and phospho-MLC2 by Western blot. Controls: total FAK, ERK1/2, MLC2, and heat shock protein 90 (HSP90). (C and D) Activation of the integrin  $\beta 1$

pathway is required for NET-induced awakening. BLI of luciferase-expressing D2.0R cells (C) cultured on matrigel for 14 days under indicated conditions ( $n = 3$ ; means  $\pm$  SD) and (D) cultured on matrigel using doxycycline to induce expression of indicated shRNAs ( $n = 3$ ; means  $\pm$  SD). (E to G) Inhibition of integrin-mediated YAP activation prevents LPS-induced awakening. (E and F) Mice with inducible shRNA-expressing D2.0R cells were monitored by BLI. Doxycycline treatment was initiated on day 5 ( $n = 5$  mice per group; means  $\pm$  SD). (G) Representative BLI images at day 33. \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; n.s., not significant.

NET-induced proteolytic laminin remodeling that can drive cancer recurrence in the lungs of mice.

## Discussion

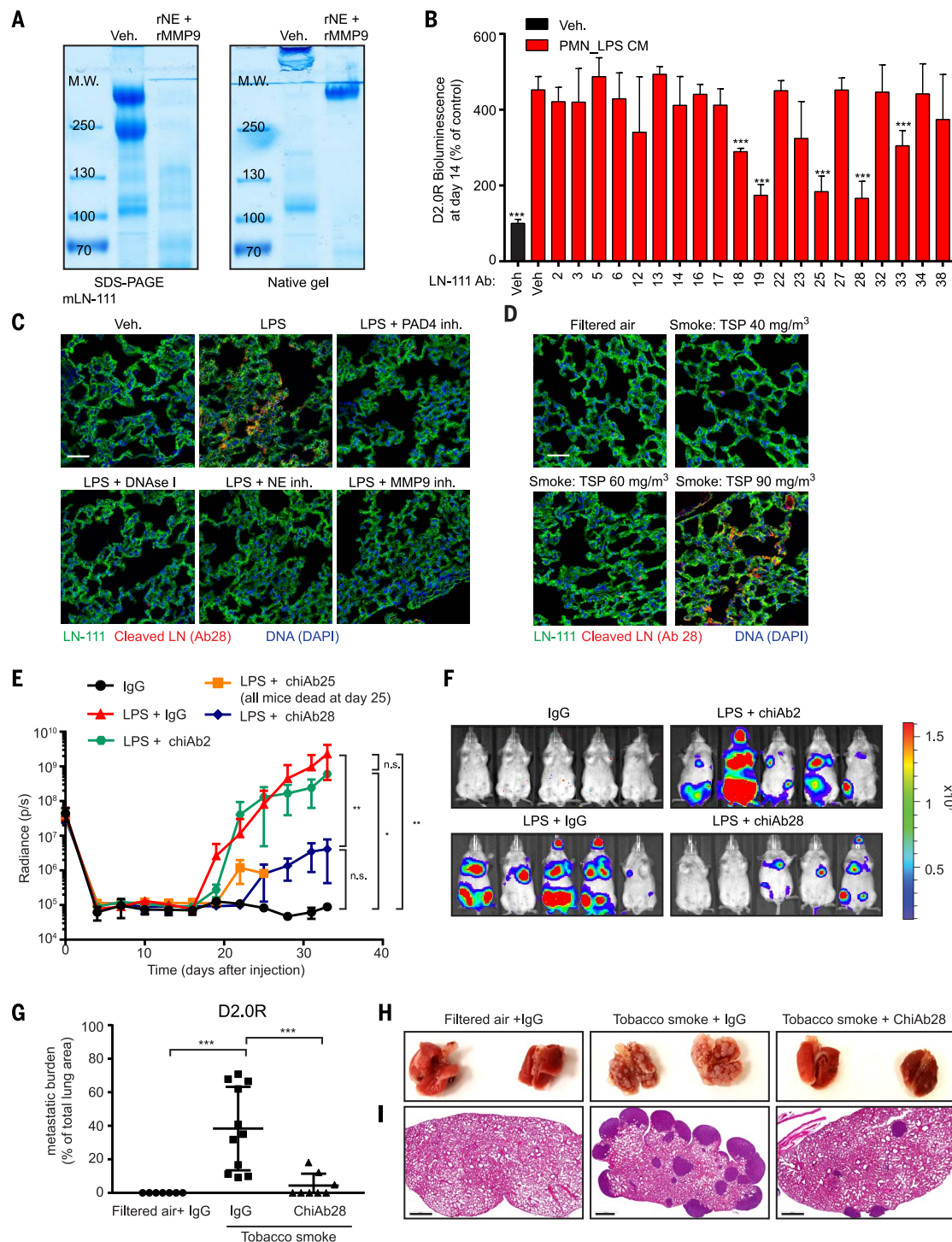
Neutrophils recruited during lung inflammation can initiate awakening of dormant cancer cells, as described here and in previous work (12). Now, we report that NETs formed by neutrophils during LPS- or tobacco smoke-induced lung

inflammation are required to awaken dormant cancer cells and cause metastasis in mice. We propose that NETs concentrate neutrophil proteases, NE and MMP9, at their substrate, laminin, allowing sequential cleavage and generating an epitope that triggers cancer cell awakening.

Previous studies pointed to a central role for integrin  $\beta 1$  signaling in the awakening of dormant cancer cells (7, 21, 35, 40), but it was un-

clear what types of stimuli could activate integrin signaling in dormant cancer cells. We have found that NET-remodeled laminin formed during lung inflammation serves as an integrin  $\alpha 3 \beta 1$ -activating epitope, leading to FAK/ERK/MLCK/YAP signaling in cancer cells and to awakening. On the basis of these results, we developed an inhibitory antibody against NET-remodeled laminin that prevented dormant cancer cells from awakening in

**Fig. 6. A NET-generated laminin epitope promotes awakening.** (A) Laminin-111 was not fully degraded by NET-associated proteases. Laminin-111 was incubated with rNE and rMMP9, and cleavage was assessed by SDS-PAGE and coomassie blue staining under reducing and denaturing or non-reducing and nondenaturing conditions. (B) Antibodies against NE- and MMP9-cleaved laminin-111 prevented NET-mediated awakening. BLI was performed on luciferase-expressing D2.0R cells cultured on laminin-111, 14 days after treatment with the indicated antibody clone ( $n = 3$ ; means  $\pm$  SD). (C and D) Lung inflammation and NETs induce production of cleaved laminin. Immunostaining of lungs for DAPI (blue), full-length laminin-111 (ab11575 from Abcam, green), and cleaved laminin antibody 28 (red), treated as indicated. The number of total suspended particles (TSP) in cigarette smoke exposure conditions is indicated. Scale bars, 50  $\mu$ m. (E to I) Antibodies against NET-remodeled laminin prevented inflammation-induced awakening in vivo. (E) Mice with D2.0R cells were treated as indicated and monitored by BLI ( $n = 5$  mice per group; means  $\pm$  SD). (F) Representative BLI images at day 33. (G) Metastatic burden in lungs from mice exposed to tobacco smoke quantified with H&E staining. (H) Representative photos of lungs. (I) Representative H&E images of lungs from animals described in (H) at day 30 ( $n = 7$  mice for filtered air + IgG,  $n = 8$  mice for tobacco smoke + IgG,  $n = 11$  mice for tobacco smoke + ChiAb28 inhibitor,  $n = 11$  mice for tobacco smoke + IgG; means  $\pm$  SD). Scale bars, 700  $\mu$ m. \*\*\* $P < 0.001$ ; \*\* $P < 0.01$ ; \* $P < 0.05$ ; n.s., not significant.





response to LPS- or tobacco smoke-induced lung inflammation. Given that our antibody preserves normal integrin  $\beta 1$  signaling, it may be a potent strategy to prevent cancer recurrence and more broadly serve as a treatment for other NET-associated pathologies, such as transfusion-related acute lung injury (41).

Our *in vitro* experiments identified laminin-III, -211, -411, and -511 as key ECM proteins that cause NET-induced cancer cell awakening. It is particularly intriguing that NETs target laminin-411 and -511, present in the perivascular niche, because this niche has been shown to regulate breast cancer dormancy (6). In previous studies, TSP-1, which is also present in this niche, was identified as a regulator of tumor dormancy and metastasis (6, 33, 34). *In vitro*, NET-mediated degradation of TSP-1 did not induce awakening. Rather, intact TSP-1 inhibited NET-induced awakening, and NET-mediated degradation of TSP-1 overcame the inhibition. Thus, TSP-1 degradation is likely required in parallel with laminin remodeling in the intact tissue to create a permissive niche for the awakening of dormant cancer cells.

Stiffness of the environment can lead to cell tension-mediated YAP activity, and we have demonstrated that YAP activity is critical for inflammation-induced cancer cell awakening. Our results therefore suggest that changes in cell tension occur after binding to NET-remodeled laminin and play a role in the awakening of disseminated dormant cancer cells. Consistently, experimental lung fibrosis, which is associated with increased ECM stiffness, induces cancer cell awakening from dormancy (7). Furthermore, D2.0R and MCF-7 cells are dormant in soft environments *in vitro*—such as on matrigel, laminin-III gel, and 0.2-kPa hydrogel (which mimics the stiffness of the lungs)—but these cells proliferate on stiff 2D cell culture dishes. *In vivo*, dormant, disseminated tumor cells are preferentially located within the soft niches of stiff tissues such as bones—e.g., in the bone marrow (42).

NETs drive dormant cancer cells to initiate proliferation, but T cells and natural killer cells may recognize these disseminated cells as they start to proliferate (2, 3, 43). The role of the adaptive immune system in our mouse models is unclear, but LPS increases glucocorticoid levels (44), inhibiting adaptive immune cells. Therefore, LPS and smoking may spur the growth of dormant cells by providing signals to both initiate proliferation, via NETs, and to overcome immune control, through glucocorticoids.

Chronic inflammation and smoking are well-known risk factors in metastatic recurrence (8, 9). Obesity is also associated with chronic, low-grade inflammation and with elevated NET release (45), and neutrophils contribute to obesity-associated metastasis in mice (46). Our findings set the stage for epidemiological studies examining whether there are any correlations between inflammation, smoking, and obesity; NETs; and cancer recurrence after a long period of dormancy in patients. If such correlations exist, we envision that NETs and their downstream effectors could be targeted to reduce the risk of cancer recurrence.

## Materials and methods summary

### Proliferation assay

Ninety-six-well culture-plates were coated with 50  $\mu$ l of growth factor reduced matrigel (#356231, Corning) or mLN-III (#3446-005-01, R&D Systems) and incubated at 37°C for 30 min. The murine D2.0R cell line was used on matrigel or mLN-III as indicated in the figures, and the human MCF-7 cell line was used on matrigel. mCherry-luciferase cells ( $2 \times 10^3$ ) were resuspended in 100  $\mu$ l of DMEM supplemented with 1% FCS and 2% matrigel or 2% mLN-III and grown on the coated wells. The next day, media were replaced with 100  $\mu$ l of CM from neutrophils cultured at indicated conditions. The CM was changed every four days. After 14 days, 100  $\mu$ l of medium containing 5  $\mu$ g/ml of luciferin (#Luck-1G, Goldbio) were added to the wells and proliferation was measured by bioluminescence imaging (BLI) using a plate reader (SpectraMax i3, Molecular Devices). To measure the proliferation of D2.0R cells expressing integrin, NE, and MMP9 shRNA, CellTiter 96 AQueous One Solution Cell Proliferation Assay (#G3580, Promega) was used, following the manufacturer's instructions.

### Animals

Female BALB/c, nude, and C57BL/6J mice were purchased from Charles River Laboratories. Female NE knockout (KO) (#6112) and MMP9 KO (#7084) mice were purchased from The Jackson Laboratory. RapidCaP mice consisted of *Pten*<sup>loxP/loxP</sup>; *Trp53*<sup>loxP/loxP</sup>; TdTomato<sup>loxP/+</sup> transgenic mice into which a Luc-Cre lentiviral plasmid (Tyler Jacks, #2090, Addgene, purchased from The University of Iowa Viral Vector Core) was injected into the prostate, as previously described (24). All procedures were approved by the Cold Spring Harbor Laboratory or University of California, Davis Institutional Animal Care and Use Committee and were conducted in accordance with the NIH's *Guide for the Care and Use of Laboratory Animals*.

### Activation of neutrophils and NET formation assays

Isolated neutrophils (250,000) were cultured in 24-well plates containing 500  $\mu$ l of serum-free DMEM and activated overnight with LPS, PMA, fMLP, or CSE to induce NETs. Cancer cells were also cocultured with neutrophils to induce NETs, as previously described (22), using Transwell with the two chambers separated by a 0.4- $\mu$ m porous membrane (#353495, Corning). Neutrophils were placed on the bottom of the lower chamber, whereas the cancer cells (100,000 D2.0R or D2.A1 cells) were placed on the membrane of the upper chamber. To induce degranulation, recombinant complement 5a was used. The next day, the neutrophil CM were collected. GSK484 and DNase I were added 30 min before neutrophil activation to inhibit PAD4 or to digest NET DNA, respectively. To assess NET formation, neutrophils grown on poly-L-lysine-coated coverslips (#354085, Corning) were fixed with 4% paraformaldehyde (PFA) for 20 min at room

temperature, rinsed twice in PBS, incubated in 50 mM of NH<sub>4</sub>Cl for 10 min and permeabilized with 0.5% Triton X-100 (#BB151-500, Thermo Fisher Scientific) for 5 min. Cells were next blocked in PBS containing 1% bovine serum albumin (BSA, #A3294, Sigma) for 30 min and incubated with anti-H2B (1:200) and anti-myeloperoxidase (1:400) antibodies in blocking buffer overnight at 4°C. After two washes in PBS, cells were incubated in the presence of fluorochrome-conjugated secondary antibodies (1:250, Invitrogen) for 40 min, rinsed twice in PBS, stained with 4',6-diamidino-2-phenylindole (DAPI, #D1306, Thermo Fisher Scientific) for 5 min, rinsed in water, and the coverslips were mounted onto glass slides using mounting media (#17985-16, Electron Microscopy Sciences).

### Analysis of recombinant laminin-III degradation

mLN-III (90  $\mu$ g) or hLN-III (1  $\mu$ g) was incubated with neutrophil CM or recombinant proteases at 37°C for 6 hours before adding lysis buffer (25 mM Tris [pH 6.8], 2% SDS, 5% glycerol, 1%  $\beta$ -mercaptoethanol, 0.01% bromophenol blue). For mLN-III, samples were then loaded on SDS-polyacrylamide gel electrophoresis, and the gel was stained with coomassie blue in a small plastic box. For hLN-III, samples were processed as described in the "Western blot" section of the supplementary materials and methods. For analysis on native gel, SDS and  $\beta$ -mercaptoethanol were omitted from the lysis and running buffers.

### Effects of experimental inflammation on dormant cancer cells *in vivo*

To generate dormant disease, we intravenously injected mCherry-luciferase-D2.0R or -MCF-7 cells (500,000 per mouse) in 100  $\mu$ l of PBS into 8-week-old female BALB/c or nude mice, respectively. Nude mice were implanted with 17 $\beta$ -estradiol 0.36 mg/pellet (#SE-121, Innovative Research of America) 48 hours before MCF-7 cell injection. To determine how lung inflammation reactivates dormant tumor cells in the lung, we induced inflammation through intranasal instillation of LPS derived from *Escherichia coli* strain 0111:B4 (#L4391, Sigma). LPS (50  $\mu$ l at a concentration of 0.25 mg/ml, or PBS as control) was administered intranasally on days 7, 10, and 13 or on days 30, 33, and 36 with a P200 pipette under anesthesia (2.5% isoflurane) into mice that had been injected with cancer cells at day 0. Neutrophils were depleted by administering anti-Ly6G antibodies intraperitoneally (clone 1A8, #BP0075-1, Bio X Cell, 200  $\mu$ g/mouse). Anti-Ly6G antibodies were first given one day before the first LPS administration, and the treatment continued three times weekly until the end of the experiment on day 33. Rat IgG2a isotype was used as a control (clone 2A3, #BP0089, Bio X Cell, 200  $\mu$ g/mouse). Inhibitors were intraperitoneally injected, starting on the first day of LPS treatment. On days when the mice received LPS administration, they were treated three times with inhibitors (30 min before LPS administration, three hours after administration, and 6 hours after administration). On all other

days, one daily treatment was given, continuing until the end of the experiment on day 33. The doses of the inhibitors were: 50 mg/kg sivelestat (NE inhibitor, #3535, Tocris); 15,000 units/kg DNase I (#04536282001, Roche); 20 mg/kg GSK484 (PAD4 inhibitor, #17488, Cayman Chemical), 50 mg/kg SB3-CT (MMP9 inhibitor, #HY-12354, Medchem Express). Vehicle consisted of 10% dimethyl sulfoxide (DMSO) in PBS for the mice treated with sivelestat, GSK484, and SB3-CT; and PBS for mice treated with DNase I. Immobilization of an enzyme on the surface of nanoparticles can increase enzyme stability, therefore, we also generated and used DNase I-coated nanoparticles, as previously described (22). The nanoparticles were suspended such that each microliter contained one unit of DNase I; 75 units of DNase I-coated nanoparticles were administered intraperitoneally per mouse. Uncoated nanoparticles were used as control. To follow metastatic disease, mice were injected intraperitoneally with luciferin (5 mg per mouse, #Luck-IG, Goldbio) and imaged using the IVIS Spectrum in vivo imaging system (#128201, PerkinElmer) every 3 days.

For experiments analyzing the effect of inhibitors on lung inflammation, cancer cells were not injected, and mice were treated with LPS at days 0, 3, and 6. On days when the mice received LPS administration, they were treated three times with inhibitors (30 min before LPS administration, three hours after administration, and six hours after administration). On all other days, one daily treatment was given. Mice were sacrificed at day 7 for tissue processing or at the times indicated in the figure legends.

### Effects of tobacco smoke exposure on dormant cancer cells in vivo

To generate dormant disease, mCherry-luciferase-D2.0R (500,000 per mouse) in 100  $\mu$ l of PBS was intravenously injected into 8-week-old female BALB/c mice. To determine how tobacco smoke (TS) exposure reactivates dormant tumor cells in the lung, the mice were exposed to tobacco smoke or filtered air for three weeks starting one week after injection using a smoke exposure system (Pinkerton laboratory, University of California, Davis). For the TS exposed group, mice were exposed five days/week for three weeks to an average concentration of  $75 \pm 11$  mg/m<sup>3</sup> of tobacco smoke particulate for 6 hours/day using 3R4F research cigarettes (Tobacco Research Institute, University of Kentucky) that were burned at a rate of four cigarettes every 10 min with a puff volume of 35 ml over a duration of 2 s, once per minute. Both side-stream and mainstream cigarette smoke were collected via a chimney and passed to a dilution and aging chamber to allow the animals to acclimate to TS exposure. The TS exposure started at an initial exposure concentration of 60 mg/m<sup>3</sup> and was increased, ultimately achieving a final target concentration of 90 mg/m<sup>3</sup>. After each 6-hour exposure to TS, the mice were then kept in the same chamber but in filtered air. For the control group, mice were handled in the same way, but were instead

exposed to filtered air for 24 hours/day, 7 days/week for the duration of the study. The concentration of carbon monoxide in the exposure chambers was monitored and averaged  $215 \pm 36$  ppm over the 3 weeks of exposure. Mice were treated with vehicle (10% DMSO in PBS) or PAD4 inhibitor (GSK484, #17488, Cayman Chemical) intraperitoneally at a dose of 20 mg/kg/day to inhibit NET formation, starting on the first day of TS exposure and every day until the end of the experiment. ChiAb28 was administered intravenously at a dose of 200  $\mu$ g/mouse to target NET-remodeled laminin in the lungs, starting on the first day of environmental TS exposure. ChiAb28 was given every 3 days until the end of the experiment. Mouse IgG2a isotype was used as a control (Clone C1.18.4, #BE0085, Bio X Cell, 200  $\mu$ g/mouse).

### Confocal intravital lung imaging

We used a modified version of our previously described method for lung imaging (22). Mice were anesthetized by intraperitoneal injection of ketamine/dexdomitor (100 mg/kg; 0.5 mg/kg). Saline (500  $\mu$ l) was injected peritoneally prior to surgery, then tracheotomies were performed, and mechanical ventilation was initiated (MiniVent ventilator for mice, Model 845, Harvard Apparatus) with the mice on a heated stage. From this point forward, the mice were kept anesthetized at 0.8 to 1.0% isoflurane delivered through the ventilator at 80 to 100 breaths per minute, with 25 ml of water maintaining positive end expiratory pressure. A thoracotomy was then performed by making an incision between the fourth and fifth rib, and a 3D printed thoracic suction window with an 8-mm coverslip was inserted (47). The window was secured to the stage with single-axis translator micromanipulators (#MSH1.5, Thor Labs), and the left lung was held in place by vacuum pressure, which was adjusted to 15 to 20 mmHg. Mice were imaged with our custom-built spinning disk confocal microscope modified to upright configuration, as described previously (22). The microscope was controlled by MicroManager software, and data analysis and time-lapse movies were generated using Imaris software.

To visualize neutrophil recruitment and neutrophil elastase activity, LysM-EGFP mice [on BALB/c background (48)] were administered intranasally with LPS (50  $\mu$ l, at a concentration of 0.25 mg/ml) or PBS as control as indicated and intravenously with neutrophil elastase 680 FAST probe (4 nmol in 100  $\mu$ l, #NEV11169, PerkinElmer, 3 to 6 hours before imaging). At the time of imaging, DAPI was administered (5  $\mu$ g/ml) intravenously, and images were captured for 2 to 8 hours. To monitor cancer cells exiting dormancy and undergoing cell cycle progression, D2.0R cells labeled with FUCCI reporter (500,000) were injected intravenously into BALB/c mice at day 0, and the mice were treated with LPS on days 7, 10, and 13, as described above. On days of imaging, the neutrophil elastase 680 FAST probe was administered intravenously 3 to 6 hours before imaging.

### Enzyme-linked immunosorbent assay (ELISA) for NETs

96-well Enzyme ImmunoAssay/Radio ImmunoAssay (EIA/RIA) plates (#3590, Costar) were coated overnight at 4°C with an anti-elastase antibody (1:250, #sc-9521, Santa Cruz Biotechnology) in 15 mM of Na<sub>2</sub>CO<sub>3</sub>, 35 mM of NaHCO<sub>3</sub>, at pH 9.6. The next day, the wells were washed three times with PBS, blocked in 5% BSA for two hours at room temperature, and washed three times with PBS. Then, 50  $\mu$ l of plasma samples were added to the wells, incubated for two hours at room temperature on a shaker, and plates were washed three times with wash buffer (1% BSA, 0.05% Tween 20 in PBS). Next, anti-DNA-peroxidase conjugated antibody (1:50, #11774425001, Roche) in 1% BSA in PBS was added to the wells for 2 hours at room temperature, and the wells were washed five times with wash buffer before the addition of 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid (ABTS, #37615, Thermo Fisher Scientific). Optical density was read 40 min later at 405 nm using a plate reader (SpectraMax i3, Molecular Devices).

### Generation of anti-NET-remodeled laminin-111 antibodies

mLN-111 (6 mg) was incubated with recombinant NE and MMP9 (2  $\mu$ g/ml each) at 37°C for 6 hours. To remove the recombinant proteases, the cleaved laminin-111 solution was purified using a 100-kDa molecular weight cut-off (MWCO) column (#VS0641, Vivaspin 6, Sartorius), following the manufacturer's instructions. The presence of laminin-111 and the absence of NE and MMP9 in the upper fraction (>100 kDa) was then assessed using Western blot analysis. The upper fraction containing the cleaved laminin-111 was next dialyzed in PBS using 20,000 MWCO dialysis cassettes (#87735, Thermo Fisher Scientific). Three 6-week-old Sprague Dawley rats (Taconics) were immunized with 1.5 mg of cleaved laminin-111 (100  $\mu$ g per animal per boost, for five weekly boosts). Immune response was monitored by ELISA to measure the serum anti-NET-remodeled laminin-111 IgG titer from blood samples. After a 60-day immunization course, the rat with the strongest anti-cleaved laminin-111 immune response was terminated, and 10<sup>8</sup> splenocytes were collected for making hybridomas by fusing with the rat myeloma cell line YB2/0, following the standard method (49). All procedures were approved by the Cold Spring Harbor Laboratory Institutional Animal Care and Use Committee. The antibodies generated were tested and endotoxin free.

### Generation of chimeric antibodies

To make antibodies for in vivo experiments, we engineered the rat mAbs into mouse chimera antibodies (chiAbs), as this reduces the likelihood of the mice developing neutralization antibodies. In brief, the rat mAb V-domains from the heavy and light chains were cloned into the mice IgG2a framework. The resulting chimeric IgG2a antibody therefore comprised rat V-domains in frame with the constant regions of mice IgG2a.



The recombinant chimeric antibodies were then produced from HEK293 cells transfected with the chimeric IgG2a expression constructs, followed by standard purification procedures from the culture supernatants.

### Statistical analysis

Data from in vivo BLI quantification were analyzed using one-way ANOVA followed by Tukey's procedure, except for fig. S7F, where a two-sided *t* test was used. Data from in vitro BLI were analyzed using two-way ANOVA followed by Tukey's procedure, except for Figs. 4, C and E, and 6B, and figs. S7J and S15D. The data represented in these figures were analyzed using one-way ANOVA followed by Dunnett's procedure, with the reference group indicated on the figure.

Data from NE and MMP9 activity in vitro were analyzed using two-way ANOVA followed by Tukey's procedure, while data from NE and MMP9 activity in vivo were analyzed using one-way ANOVA followed by Dunnett's procedure, where the reference group was indicated on the figure. Assays on dsDNA, plasma NET and LPS, number of DTCs/mm<sup>2</sup>, metastatic burden, neutrophils' number per field, percentage of neutrophils forming NETs, and percentage of Ki67-positive cancer cells in vitro—except for fig. S7G—were analyzed using one-way ANOVA followed by Dunnett's procedure, with the reference group indicated on the figure. Fig. S7G was analyzed using a two-sided *t* test.

Data from qPCR quantification were analyzed using one-way ANOVA followed by Dunnett's procedure, where the reference group was indicated on the figure, except for fig. S10F, where a two-sided *t* test was used. To analyze lung metastasis-free survival (fig. S5C), Kaplan-Meier curves of RapidCaP mice were analyzed using a log-rank (Mantel-Cox) test. For fig. S5C, an exact *P* value was computed using Fisher's exact test to evaluate the likelihood that the metastatic types (macrometastasis, micrometastasis, single DTC) differed among groups. The analyses on figs. S2E and S15B had the cancer cell as the experimental unit, with each cell characterized on the two variables of interest for the figure. For fig. S2E, cells were characterized based on experimental group (vehicle, day 11, day 12, day 13) and by cell feature (G<sub>0</sub>/G<sub>1</sub> single cells, S/G<sub>2</sub>/M single cells, clusters of cells). A total of 182 cells were characterized and Fisher's exact test was used to assess whether the cell feature distribution differed among the different experimental groups. The exact test was significant (*P* < 0.0001): Over time, G<sub>0</sub>/G<sub>1</sub> single cells and S/G<sub>2</sub>/M single cells decreased and clusters of cells increased. For fig. S15B, cells were characterized based on cell type (PBS and G<sub>0</sub>/G<sub>1</sub> single cells, LPS and G<sub>0</sub>/G<sub>1</sub> single cells, LPS and S/G<sub>2</sub>/M single cells, LPS and clusters of cells) and next to Ab28 (yes, no). A total of 182 cells were characterized. Fisher's exact test was used to assess whether the percentage of cells that were “next to Ab28” in each cell type group were different among the groups. The exact test was significant (*P* < 0.0001). All PBS cells

were G<sub>0</sub>/G<sub>1</sub> single cells and were not “next to Ab28.” LPS and G<sub>0</sub>/G<sub>1</sub> single cells were equally likely to be “next to Ab28” or “not next to Ab28.” All LPS and S/G<sub>2</sub>/M single cells and all LPS and clusters of cells were “next to Ab28.” Analysis for figs. S5C, S2E, and S15B were generated using Proc Freq in the SAS/STAT software, version 9.4 of the SAS system for Windows.

A *P* value less than 0.05 was considered significant, and *P* values are indicated in figures as \*\*\**P* < 0.001, \*\**P* < 0.01, and \**P* < 0.05. All statistical analyses were performed using GraphPad Prism software version 7 unless otherwise stated.

### REFERENCES AND NOTES

- 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](https://doi.org/10.1038/nrc3793); pmid: [25118602](https://pubmed.ncbi.nlm.nih.gov/25118602/)
- 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](https://doi.org/10.1016/j.cell.2016.02.025); pmid: [27015306](https://pubmed.ncbi.nlm.nih.gov/27015306/)
- I. Romero, F. Garrido, A. M. Garcia-Lora, Metastases in immune-mediated dormancy: A new opportunity for targeting cancer. *Cancer Res.* **74**, 6750–6757 (2014). doi: [10.1158/0008-5472.CAN-14-2406](https://doi.org/10.1158/0008-5472.CAN-14-2406); pmid: [25411345](https://pubmed.ncbi.nlm.nih.gov/25411345/)
- J. D. Farrar et al., Cancer dormancy. VII. A regulatory role for CD8+ T cells and IFN-γ in establishing and maintaining the tumor-dormant state. *J. Immunol.* **162**, 2842–2849 (1999). pmid: [10072532](https://pubmed.ncbi.nlm.nih.gov/10072532/)
- 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](https://pubmed.ncbi.nlm.nih.gov/9850077/)
- C. M. Ghajar et al., The perivascular niche regulates breast tumour dormancy. *Nat. Cell Biol.* **15**, 807–817 (2013). doi: [10.1038/ncb2767](https://doi.org/10.1038/ncb2767); pmid: [23728425](https://pubmed.ncbi.nlm.nih.gov/23728425/)
- D. Barkan et al., Metastatic growth from dormant cells induced by a col-I-enriched fibrotic environment. *Cancer Res.* **70**, 5706–5716 (2010). doi: [10.1158/0008-5472.CAN-09-2356](https://doi.org/10.1158/0008-5472.CAN-09-2356); pmid: [20570886](https://pubmed.ncbi.nlm.nih.gov/20570886/)
- B. L. Pierce et al., Elevated biomarkers of inflammation are associated with reduced survival among breast cancer patients. *J. Clin. Oncol.* **27**, 3437–3444 (2009). doi: [10.1200/JCO.2008.18.9068](https://doi.org/10.1200/JCO.2008.18.9068); pmid: [19470939](https://pubmed.ncbi.nlm.nih.gov/19470939/)
- J. P. Pierce et al., Lifetime cigarette smoking and breast cancer prognosis in the After Breast Cancer Pooling Project. *J. Natl. Cancer Inst.* **106**, djt359 (2014). doi: [10.1093/jnci/djt359](https://doi.org/10.1093/jnci/djt359); pmid: [24317179](https://pubmed.ncbi.nlm.nih.gov/24317179/)
- A. H. Wu et al., The California Breast Cancer Survivorship Consortium (CBCSC): Prognostic factors associated with racial/ethnic differences in breast cancer survival. *Cancer Causes Control* **24**, 1821–1836 (2013). doi: [10.1007/s10552-013-0260-7](https://doi.org/10.1007/s10552-013-0260-7); pmid: [23864487](https://pubmed.ncbi.nlm.nih.gov/23864487/)
- S. Murin, K. E. Pinkerton, N. E. Hubbard, K. Erickson, The effect of cigarette smoke exposure on pulmonary metastatic disease in a murine model of metastatic breast cancer. *Chest* **125**, 1467–1471 (2004). doi: [10.1378/chest.125.4.1467](https://doi.org/10.1378/chest.125.4.1467); pmid: [15078760](https://pubmed.ncbi.nlm.nih.gov/15078760/)
- J. M. De Cock et al., Inflammation Triggers Zeb1-Dependent Escape from Tumor Latency. *Cancer Res.* **76**, 6778–6784 (2016). doi: [10.1158/0008-5472.CAN-16-0608](https://doi.org/10.1158/0008-5472.CAN-16-0608); pmid: [27530323](https://pubmed.ncbi.nlm.nih.gov/27530323/)
- S. K. Jorch, P. Kubes, An emerging role for neutrophil extracellular traps in noninfectious disease. *Nat. Med.* **23**, 279–287 (2017). doi: [10.1038/nm.4294](https://doi.org/10.1038/nm.4294); pmid: [28267716](https://pubmed.ncbi.nlm.nih.gov/28267716/)
- N. Branzk, V. Papayannopoulos, Molecular mechanisms regulating NETosis in infection and disease. *Semin. Immunopathol.* **35**, 513–530 (2013). doi: [10.1007/s00281-013-0384-6](https://doi.org/10.1007/s00281-013-0384-6); pmid: [23732507](https://pubmed.ncbi.nlm.nih.gov/23732507/)
- C. Carmona-Rivera, W. Zhao, S. Yalavarthi, M. J. Kaplan, Neutrophil extracellular traps induce endothelial dysfunction in systemic lupus erythematosus through the activation of matrix metalloproteinase-2. *Ann. Rheum. Dis.* **74**, 1417–1424 (2015). doi: [10.1136/annrheumdis-2013-204837](https://doi.org/10.1136/annrheumdis-2013-204837); pmid: [24570026](https://pubmed.ncbi.nlm.nih.gov/24570026/)
- T. A. Fuchs et al., Extracellular DNA traps promote thrombosis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15880–15885 (2010). doi: [10.1073/pnas.1005743107](https://doi.org/10.1073/pnas.1005743107); pmid: [20798043](https://pubmed.ncbi.nlm.nih.gov/20798043/)
- A. Brill et al., Neutrophil extracellular traps promote deep vein thrombosis in mice. *J. Thromb. Haemost.* **10**, 136–144 (2012). doi: [10.1111/j.1538-7836.2011.04544.x](https://doi.org/10.1111/j.1538-7836.2011.04544.x); pmid: [22044575](https://pubmed.ncbi.nlm.nih.gov/22044575/)
- S. L. Wong et al., Diabetes primes neutrophils to undergo NETosis, which impairs wound healing. *Nat. Med.* **21**, 815–819 (2015). doi: [10.1038/nm.3887](https://doi.org/10.1038/nm.3887); pmid: [26076037](https://pubmed.ncbi.nlm.nih.gov/26076037/)
- J. Cools-Lartigue et al., Neutrophil extracellular traps sequester circulating tumor cells and promote metastasis. *J. Clin. Invest.* **123**, 3446–3458 (2013). doi: [10.1172/JCI67484](https://doi.org/10.1172/JCI67484); pmid: [23863628](https://pubmed.ncbi.nlm.nih.gov/23863628/)
- S. Tohme et al., Neutrophil extracellular traps promote the development and progression of liver metastases after surgical stress. *Cancer Res.* **76**, 1367–1380 (2016). doi: [10.1158/0008-5472.CAN-15-1591](https://doi.org/10.1158/0008-5472.CAN-15-1591); pmid: [26759232](https://pubmed.ncbi.nlm.nih.gov/26759232/)
- D. Barkan et al., Inhibition of metastatic outgrowth from single dormant tumor cells by targeting the cytoskeleton. *Cancer Res.* **68**, 6241–6250 (2008). doi: [10.1158/0008-5472.CAN-07-6849](https://doi.org/10.1158/0008-5472.CAN-07-6849); pmid: [18676848](https://pubmed.ncbi.nlm.nih.gov/18676848/)
- J. Park et al., Cancer cells induce metastasis-supporting neutrophil extracellular DNA traps. *Sci. Transl. Med.* **8**, 361ra138 (2016). doi: [10.1126/scitranslmed.aag1711](https://doi.org/10.1126/scitranslmed.aag1711); pmid: [27798263](https://pubmed.ncbi.nlm.nih.gov/27798263/)
- S. B. Koh et al., A quantitative FastFUCI assay defines cell cycle dynamics at a single-cell level. *J. Cell Sci.* **130**, 512–520 (2017). doi: [10.1242/jcs.195164](https://doi.org/10.1242/jcs.195164); pmid: [27888217](https://pubmed.ncbi.nlm.nih.gov/27888217/)
- H. Cho et al., RapidCaP, a novel GEM model for metastatic prostate cancer analysis and therapy, reveals myc as a driver of Pten-mutant metastasis. *Cancer Discov.* **4**, 318–333 (2014). doi: [10.1158/2159-8290.CD-13-0346](https://doi.org/10.1158/2159-8290.CD-13-0346); pmid: [24444712](https://pubmed.ncbi.nlm.nih.gov/24444712/)
- D. Barkan, J. E. Green, An in vitro system to study tumor dormancy and the switch to metastatic growth. *J. Vis. Exp.* **2011**, e2914 (2011). pmid: [21860375](https://pubmed.ncbi.nlm.nih.gov/21860375/)
- A. Hosseinzadeh, P. R. Thompson, B. H. Segal, C. F. Urban, Nicotine induces neutrophil extracellular traps. *J. Leukoc. Biol.* **100**, 1105–1112 (2016). doi: [10.1189/jlb.3A80815-379RR](https://doi.org/10.1189/jlb.3A80815-379RR); pmid: [27312847](https://pubmed.ncbi.nlm.nih.gov/27312847/)
- R. William, G. Watson, H. P. Redmond, J. M. Carthy, D. Bouchier-Hayes, Tauridine, an antilipopolysaccharide agent, has immunoregulatory properties that are mediated by the amino acid taurine. *J. Leukoc. Biol.* **58**, 299–306 (1995). doi: [10.1002/jlb.58.3.299](https://doi.org/10.1002/jlb.58.3.299); pmid: [7665985](https://pubmed.ncbi.nlm.nih.gov/7665985/)
- M. Egeblad, Z. Werb, New functions for the matrix metalloproteinases in cancer progression. *Nat. Rev. Cancer* **2**, 161–174 (2002). doi: [10.1038/nrc745](https://doi.org/10.1038/nrc745); pmid: [11990853](https://pubmed.ncbi.nlm.nih.gov/11990853/)
- K. Kessenbrock, V. Plaks, Z. Werb, Matrix metalloproteinases: Regulators of the tumor microenvironment. *Cell* **141**, 52–67 (2010). doi: [10.1016/j.cell.2010.03.015](https://doi.org/10.1016/j.cell.2010.03.015); pmid: [20371345](https://pubmed.ncbi.nlm.nih.gov/20371345/)
- V. Papayannopoulos, K. D. Metzler, A. Hakkim, A. Zychlinsky, Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps. *J. Cell Biol.* **191**, 677–691 (2010). doi: [10.1083/jcb.201006052](https://doi.org/10.1083/jcb.201006052); pmid: [20974816](https://pubmed.ncbi.nlm.nih.gov/20974816/)
- K. D. Metzler, C. Goosmann, A. Lubojemska, A. Zychlinsky, V. Papayannopoulos, A myeloperoxidase-containing complex regulates neutrophil elastase release and actin dynamics during NETosis. *Cell Reports* **8**, 883–896 (2014). doi: [10.1016/j.celrep.2014.06.044](https://doi.org/10.1016/j.celrep.2014.06.044); pmid: [25066128](https://pubmed.ncbi.nlm.nih.gov/25066128/)
- A. Beliveau et al., Raf-induced MMP9 disrupts tissue architecture of human breast cells in three-dimensional culture and is necessary for tumor growth in vivo. *Genes Dev.* **24**, 2800–2811 (2010). doi: [10.1101/gad.1990410](https://doi.org/10.1101/gad.1990410); pmid: [21159820](https://pubmed.ncbi.nlm.nih.gov/21159820/)
- R. Catena et al., Bone marrow-derived Gr1+ cells can generate a metastasis-resistant microenvironment via induced secretion of thrombospondin-1. *Cancer Discov.* **3**, 578–589 (2013). doi: [10.1158/2159-8290.CD-12-0476](https://doi.org/10.1158/2159-8290.CD-12-0476); pmid: [23633432](https://pubmed.ncbi.nlm.nih.gov/23633432/)
- T. El Rayes et al., Lung inflammation promotes metastasis through neutrophil protease-mediated degradation of Tsp-1. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 16000–16005 (2015). doi: [10.1073/pnas.1507294112](https://doi.org/10.1073/pnas.1507294112); pmid: [26668367](https://pubmed.ncbi.nlm.nih.gov/26668367/)
- J. A. Aguirre Ghiso, Inhibition of FAK signaling activated by urokinase receptor induces dormancy in human carcinoma cells in vivo. *Oncogene* **21**, 2513–2524 (2002). doi: [10.1038/sj.onc.1205342](https://doi.org/10.1038/sj.onc.1205342); pmid: [11971186](https://pubmed.ncbi.nlm.nih.gov/11971186/)
- J. A. Aguirre Ghiso, K. Kovalski, L. Ossowski, Tumor dormancy induced by downregulation of urokinase receptor in human carcinoma involves integrin and MAPK signaling. *J. Cell Biol.* **147**, 89–104 (1999). doi: [10.1083/jcb.147.1.89](https://doi.org/10.1083/jcb.147.1.89); pmid: [10508858](https://pubmed.ncbi.nlm.nih.gov/10508858/)
- D. Liu, J. Aguirre Ghiso, Y. Estrada, L. Ossowski, EGFR is a transducer of the urokinase receptor initiated signal that is required for in vivo growth of a human carcinoma. *Cancer Cell* **1**, 445–457 (2002). doi: [10.1016/S1535-6108\(02\)00072-7](https://doi.org/10.1016/S1535-6108(02)00072-7); pmid: [12124174](https://pubmed.ncbi.nlm.nih.gov/12124174/)
- F. Calvo et al., Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nat. Cell Biol.* **15**, 637–646 (2013). doi: [10.1038/ncb2756](https://doi.org/10.1038/ncb2756); pmid: [23708000](https://pubmed.ncbi.nlm.nih.gov/23708000/)



39. R. O. Hynes, Integrins: Bidirectional, allosteric signaling machines. *Cell* **110**, 673–687 (2002). doi: [10.1016/S0092-8674\(02\)00971-6](https://doi.org/10.1016/S0092-8674(02)00971-6); pmid: [12297042](https://pubmed.ncbi.nlm.nih.gov/12297042/)
40. T. Shibue, M. W. Brooks, R. A. Weinberg, An integrin-linked machinery of cytoskeletal regulation that enables experimental tumor initiation and metastatic colonization. *Cancer Cell* **24**, 481–498 (2013). doi: [10.1016/j.ccr.2013.08.012](https://doi.org/10.1016/j.ccr.2013.08.012); pmid: [24035453](https://pubmed.ncbi.nlm.nih.gov/24035453/)
41. A. Caudrillier *et al.*, Platelets induce neutrophil extracellular traps in transfusion-related acute lung injury. *J. Clin. Invest.* **122**, 2661–2671 (2012). doi: [10.1172/JCI61303](https://doi.org/10.1172/JCI61303); pmid: [22684106](https://pubmed.ncbi.nlm.nih.gov/22684106/)
42. K. Pantel, C. Alix-Panabières, Bone marrow as a reservoir for disseminated tumor cells: A special source for liquid biopsy in cancer patients. *Bonekey Rep.* **3**, 584 (2014). doi: [10.1038/bonekey.2014.79](https://doi.org/10.1038/bonekey.2014.79); pmid: [25419458](https://pubmed.ncbi.nlm.nih.gov/25419458/)
43. A. Pommier *et al.*, Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases. *Science* **360**, eaao4908 (2018). doi: [10.1126/science.aao4908](https://doi.org/10.1126/science.aao4908); pmid: [29773669](https://pubmed.ncbi.nlm.nih.gov/29773669/)
44. K. Nakano, S. Suzuki, C. Oh, Significance of increased secretion of glucocorticoids in mice and rats injected with bacterial endotoxin. *Brain Behav. Immun.* **1**, 159–172 (1987). doi: [10.1016/0889-1591\(87\)90018-3](https://doi.org/10.1016/0889-1591(87)90018-3); pmid: [3330674](https://pubmed.ncbi.nlm.nih.gov/3330674/)
45. V. Delgado-Rizo *et al.*, Neutrophil extracellular traps and its implications in inflammation: An overview. *Front. Immunol.* **8**, 81 (2017). doi: [10.3389/fimmu.2017.00081](https://doi.org/10.3389/fimmu.2017.00081); pmid: [28220120](https://pubmed.ncbi.nlm.nih.gov/28220120/)
46. D. F. Quail *et al.*, Obesity alters the lung myeloid cell landscape to enhance breast cancer metastasis through IL5 and GM-CSF. *Nat. Cell Biol.* **19**, 974–987 (2017). doi: [10.1038/ncb3578](https://doi.org/10.1038/ncb3578); pmid: [28737771](https://pubmed.ncbi.nlm.nih.gov/28737771/)
47. M. B. Headley *et al.*, Visualization of immediate immune responses to pioneer metastatic cells in the lung. *Nature* **531**, 513–517 (2016). doi: [10.1038/nature16985](https://doi.org/10.1038/nature16985); pmid: [26982733](https://pubmed.ncbi.nlm.nih.gov/26982733/)
48. N. Faust, F. Varas, L. M. Kelly, S. Heck, T. Graf, Insertion of enhanced green fluorescent protein into the lysozyme gene creates mice with green fluorescent granulocytes and macrophages. *Blood* **96**, 719–726 (2000). pmid: [10887140](https://pubmed.ncbi.nlm.nih.gov/10887140/)
49. E. A. Greenfield, *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, ed. 2, 2014).

#### ACKNOWLEDGMENTS

We thank J. Green (NIH) for the D2.0R and D2.A1 cells and Z. Werb (University California, San Francisco) for the MMP9 antibody. **Funding:** This work was supported by the CSHL Cancer Center Support P30-CA045508 (to J.T.-H.Y., C.B., K.C., S.K.L., and P.A.G.); the Department of Defense (W81XWH-14-1-0078) and the Pershing Square Sohn Cancer Research Alliance Prize (to M.E.); the Association pour la Recherche sur le Cancer (AE20141202131), the Terri Brodeur Breast Cancer Foundation, the European Molecular Biology Organization (ALTF 1425-2015), and the Susan G. Komen Foundation (PDF16376754) (to J.A.); the Lustgarten Foundation, the National Cancer Institute, and the Cedar Hill Foundation (to D.T.F.); R01 NIH Research Grant (5R01CA137050) (to L.C.T. and A.A.); CSHL and Northwell Health (to M.A.S., E.M.C., C.G.P., M.S.G., and M.E.); a Starr Centennial Scholarship and a George A. & Marjorie H. Anderson Scholarship from the Watson School of Biological Sciences (to E.B. and L.M.); a Boehringer Ingelheim Fonds Ph.D. fellowship (to L.M.); a Deutsche Forschungsgemeinschaft research fellowship (KU 3264/1-1) (to V.K.); UC Davis NIEHS Environmental Health Science Center (P30 ES023513) (to K.E.P., M.E.P., P.U., and D.L.U.); and an AACR-Bayer Healthcare Basic Cancer Research Fellowship (to M.A.S.). **Author contributions:** J.A. and M.E. designed and directed the project. J.A. and D.N. performed all experiments,

except as follows: M.E.P., P.U., D.L.U., and K.E.P. performed tobacco smoke exposure experiments on mice; A.A. and L.C.T. designed and performed experiments on RapidCaP mice; C.B. and J.T.-H.Y. generated antibodies against cleaved murine laminin-111; C.G.P., E.M.C., and M.S.G. generated DNase I and free coated nanoparticles; A.P. and D.T.F. designed and performed flow cytometry experiments; K.C. generated the viral particles for shRNA-mediated inducible knockdown of integrin  $\beta$ 1, FAK, MLCK, and YAP; and M.A.S. performed confocal intravital lung imaging. E.B., L.M., and V.K. generated reagents and/or helped analyze data. S.K.L. provided the FUCCI cell cycle reporter plasmid. P.A.G., J.A., and M.E. performed statistical analysis. J.A. and M.E. wrote the manuscript. **Competing interests:** D.T.F. is a cofounder of Myosotis LLC (a company developing cancer immunotherapies) and is on the scientific advisory boards of iTEOS Therapeutics (a company developing immuno-oncology drugs), IFM Therapeutics, LLC (a company developing therapies targeting the innate immune system), and Kymab (a company developing therapeutic antibodies). The authors declare no other competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The Ab28 monoclonal antibody against cleaved laminin is available upon signing a material transfer agreement.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6409/eaao4227/suppl/DC1](http://www.sciencemag.org/content/361/6409/eaao4227/suppl/DC1)  
Materials and Methods  
Figs. S1 to S15  
Reference (50)  
Movies S1 to S8

21 July 2017; accepted 3 August 2018  
[10.1126/science.aao4227](https://doi.org/10.1126/science.aao4227)

## RESEARCH ARTICLE SUMMARY

## EPIGENOMICS

# Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci

Vitor Onuchic\*, Eugene Lurie\*, Ivenise Carrero, Piotr Pawliczek, Ronak Y. Patel, Joel Rozowsky, Timur Galeev, Zhuoyi Huang, Robert C. Altshuler, Zhizhuo Zhang, R. Alan Harris, Cristian Coarfa, Lillian Ashmore, Jessica W. Bertol, Walid D. Fakhouri, Fuli Yu, Manolis Kellis, Mark Gerstein, Aleksandar Milosavljevic†

**INTRODUCTION:** A majority of imbalances in DNA methylation between homologous chromosomes in humans are sequence-dependent; the DNA sequence differences between the two chromosomes cause differences in the methylation state of neighboring cytosines on the same chromosome. The analyses of this sequence-dependent allele-specific methylation (SD-ASM) traditionally involved measurement of average methylation levels across many cells. Detailed understanding of SD-ASM at the single-cell and single-chromosome levels is lacking. This gap in understanding may hide the connection

between SD-ASM, ubiquitous stochastic cell-to-cell and chromosome-to-chromosome variation in DNA methylation, and the puzzling and evolutionarily conserved patterns of intermediate methylation at gene regulatory loci.

**RATIONALE:** Whole-genome bisulfite sequencing (WGBS) provides the ultimate single-chromosome level of resolution and comprehensive whole-genome coverage required to explore SD-ASM. However, the exploration of the link between SD-ASM, stochastic variation in DNA methylation, and gene regulation re-

quires deep coverage by WGBS across tissues and individuals and the context of other epigenomic marks and gene transcription.

**RESULTS:** We constructed maps of allelic imbalances in DNA methylation, histone marks, and gene transcription in 71 epigenomes from 36 distinct cell and tissue types from 13 donors. Deep (1691-fold) combined WGBS read coverage across 49 methylomes revealed CpG methylation imbalances exceeding 30% differences at 5% of the loci, which is more conservative than previous estimates in the 8 to 10% range; a similar value (8%) is observed in our dataset when we lowered our threshold for detecting

## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aar3146>

allelic imbalance to 20% methylation difference between the two alleles.

Extensive sequence-dependent CpG methylation imbalances were observed at thousands of heterozygous regulatory loci. Stochastic switching, defined as random transitions between fully methylated and unmethylated states of DNA, occurred at thousands of regulatory loci bound by transcription factors (TFs). Our results explain the conservation of intermediate methylation states at regulatory loci by showing that the intermediate methylation reflects the relative frequencies of fully methylated and fully unmethylated epialleles. SD-ASM is explainable by different relative frequencies of methylated and unmethylated epialleles for the two alleles. The differences in epiallele frequency spectra of the alleles at thousands of TF-bound regulatory loci correlated with the differences in alleles' affinities for TF binding, which suggests a mechanistic explanation for SD-ASM.

We observed an excess of rare variants among those showing SD-ASM, which suggests that an average human genome harbors at least ~200 detrimental rare variants that also show SD-ASM. The methylome's sensitivity to genetic variation is unevenly distributed across the genome, which is consistent with buffering of housekeeping genes against the effects of random mutations. By contrast, less essential genes with tissue-specific expression patterns show sensitivity, thus providing opportunity for evolutionary innovation through changes in gene regulation.

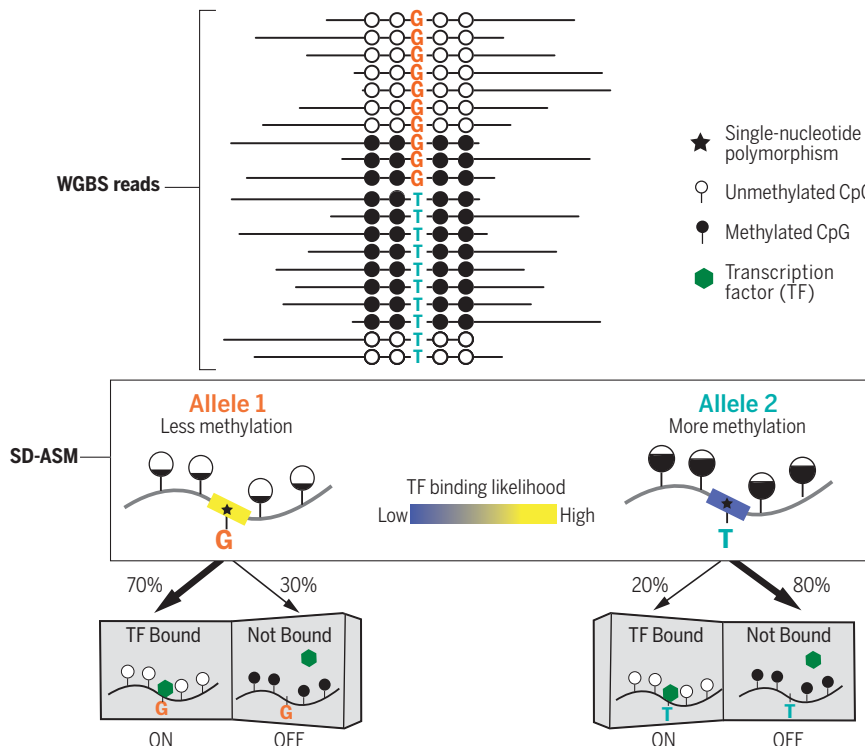
**CONCLUSION:** Analysis of allelic epigenome maps provides a unifying model that links sequence-dependent allelic imbalances of the epigenome, stochastic switching at gene regulatory loci, selective buffering of the regulatory circuitry against the effects of random mutations, and disease-associated genetic variation. ■

The list of author affiliations is available in the full article online.

\*These authors contributed equally to this work.

†Corresponding author. Email: [amilosav@bcm.edu](mailto:amilosav@bcm.edu)

Cite this article as V. Onuchic et al., *Science* 361, eaar3146 (2018). DOI: 10.1126/science.aar3146



**SD-ASM is explainable by different frequencies of epialleles.** Genetic variants affect the methylation state of neighboring cytosines on the same chromosome. Every WGBS read provides a readout of both a genetic variant (G or T) at a heterozygous locus and the methylation state of neighboring cytosines ("epiallele"). Epiallele frequencies correlate with the affinity of TF binding at regulatory sites.

## RESEARCH ARTICLE

## EPIGENOMICS

# Allele-specific epigenome maps reveal sequence-dependent stochastic switching at regulatory loci

Vitor Onuchic<sup>1,2,3,4,\*†</sup>, Eugene Lurie<sup>1,3,4,\*</sup>, Ivenise Carrero<sup>1,3</sup>, Piotr Pawliczek<sup>1,3</sup>, Ronak Y. Patel<sup>1,3</sup>, Joel Rozowsky<sup>5,6</sup>, Timur Galeev<sup>5,6</sup>, Zhuoyi Huang<sup>1,7</sup>, Robert C. Altshuler<sup>4,8,9</sup>, Zhizhuo Zhang<sup>8,9</sup>, R. Alan Harris<sup>1,3,4</sup>, Cristian Coarfa<sup>1,3,4</sup>, Lillian Ashmore<sup>1,2,3</sup>, Jessica W. Bertol<sup>10</sup>, Walid D. Fakhouri<sup>10</sup>, Fuli Yu<sup>1,2,7</sup>, Manolis Kellis<sup>4,8,9</sup>, Mark Gerstein<sup>5,6</sup>, Aleksandar Milosavljevic<sup>1,2,3,4,†</sup>

To assess the impact of genetic variation in regulatory loci on human health, we constructed a high-resolution map of allelic imbalances in DNA methylation, histone marks, and gene transcription in 71 epigenomes from 36 distinct cell and tissue types from 13 donors. Deep whole-genome bisulfite sequencing of 49 methylomes revealed sequence-dependent CpG methylation imbalances at thousands of heterozygous regulatory loci. Such loci are enriched for stochastic switching, which is defined as random transitions between fully methylated and unmethylated states of DNA. The methylation imbalances at thousands of loci are explainable by different relative frequencies of the methylated and unmethylated states for the two alleles. Further analyses provided a unifying model that links sequence-dependent allelic imbalances of the epigenome, stochastic switching at gene regulatory loci, and disease-associated genetic variation.

A majority of imbalances in DNA methylation between homologous chromosomes in humans are associated with genetic variation in cis, in which the genetic variants affect the methylation state of neighboring cytosines on the same chromosome (1, 2). Such sequence-dependent allele-specific methylation (SD-ASM) affects at least 8 to 10% of the autosomal genome (3–5). SD-ASM is an ideally “controlled” natural experiment that provides information about consequences of genetic variation in cis because both “case” and “control” loci can be found within an individual on homologous chromosomes within the same cellular and nuclear environment.

In contrast to association-based methods—such as expression quantitative trait loci (eQTLs) and methylation quantitative trait loci (mQTLs), used to establish the functional effects of common (>5% minor allele frequency) noncoding variants (6)—allelic imbalances (AIs) can be established by profiling a single sample, revealing the functional effects in cis of both relatively low-risk common variants and highly penetrant disease-causing rare and de novo variants (1, 2, 5, 7–10).

Although the analyses of SD-ASM traditionally involved measurement of average methylation levels across many cells, the epigenome is known to exhibit stochastic cell-to-cell variation and even variation between the two chromosomes within the same nucleus (11). Whole-genome bisulfite sequencing (WGBS) has the potential to provide insights into stochastic variation at the ultimate single-chromosome level of resolution because it provides information about the genetic variant and methylation of neighboring cytosines within the same sequencing read that comes from a single chromosome. Sequencing-based studies of DNA methylation have revealed pervasive stochastic epigenetic polymorphisms within autosomal loci (12, 13).

Methylation patterns evolve along predictable trajectories during normal development (14), being highly stochastic in early metastable stages and stabilizing within differentiated tissues, which results in mosaicism (15). The methylation patterns are maintained in stem cells through a dynamic stochastic epigenetic switching equilib-

rium (an ergodic process) that provides both epigenetic buffering of environmental noise and responsiveness to specific transcription factors (TFs) (12). By contrast, more differentiated cells (12) and tumor cells (13, 16) appear to use the more error-prone mechanism of direct replication of CpG methylation patterns, which is associated with clonality and mosaicism. Despite the stochastic nature of DNA methylation changes during development, in response to environmental input and in human diseases, the effects of genetic variation on stochasticity and mosaicism of the epigenome remain unexplored.

## Results

### Patterns of AIs across epigenomic marks

To explore the effects of genetic variation on the epigenome, the National Institutes of Health (NIH) Roadmap Epigenomics Project (17) has now completed whole-genome sequencing (WGS) on genomes of 13 donors and published NIH Roadmap reference epigenomes from 71 combined samples that collectively represent 27 distinct tissue types and nine cell types (fig. S1). For accurate identification of heterozygous genomic loci, we sequenced the donor genomes (18). Eight assays were included in most of the samples and used for AI detection: WGBS, RNA sequencing (RNA-seq), and chromatin immunoprecipitation sequencing (ChIP-seq) for six different histone marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3, and H3K27ac) (fig. S1). We performed allele-specific methylation (ASM) analysis at heterozygous single-nucleotide polymorphism (SNP) loci within the 49 WGBS methylomes using a threshold of absolute methylation difference of >30% between alleles and by estimating significance by means of Fisher's exact test on the counts of methylated and unmethylated cytosines observed on the same sequencing read with each of the two SNP alleles (fig. S2A) (18). We performed the identification of AIs for histone marks and transcription using the AlleleSeq pipeline (fig. S2B) (18, 19).

Considering the AIs in all the marks, the imbalances in DNA methylation were by far the most abundant (table S1 and fig. S3), largely because of the genome-wide distribution of DNA methylation, in contrast to the uneven genomic distribution of other marks. Among the histone marks, H3K27ac had more imbalance calls than others (table S1), in part owing to deeper ChIP-seq coverage for H3K27ac (table S1 and fig. S3). At promoters, H3K27ac and H3K4me3 marks were more abundant on the allele with less DNA methylation (Fig. 1A). Conversely, H3K9me3 signal was more abundant on the allele with more methylation in promoters (Fig. 1A). At enhancers, H3K27ac tended to occur more often on the allele with less DNA methylation (Fig. 1A). We also detected, at high specificity, enrichment of AIs in methylation and coordinated changes in transcription and histone marks within a majority of those imprinted loci that included a heterozygous SNP (figs. S4 and S5) (18).

We next evaluated the extent of reported SD-ASM. Consistent with genetic effects in cis

<sup>1</sup>Molecular and Human Genetics Department, Baylor College of Medicine, Houston, TX, USA. <sup>2</sup>Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX, USA. <sup>3</sup>Epigenome Center, Baylor College of Medicine, Houston, TX, USA. <sup>4</sup>NIH Roadmap Epigenomics Project. <sup>5</sup>Program in Computational Biology and Bioinformatics, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA. <sup>6</sup>Department of Computer Science, Yale University, New Haven, CT, USA. <sup>7</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. <sup>8</sup>Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>9</sup>Broad Institute of Harvard University and Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>10</sup>Center for Craniofacial Research, Department of Diagnostic and Biomedical Sciences, School of Dentistry, University of Texas Health Science Center at Houston, Houston, TX, USA.

\*These authors contributed equally to this work.

†Present address: Illumina, San Diego, CA, USA.

‡Corresponding author. Email: amilosav@bcm.edu



(6, 20–22), co-occurrence of ASM at the same heterozygous locus across different samples was higher than expected by chance under a permutation-based null model (fig. S6A). The degree of co-occurrence of ASM tended to be higher for pairs of samples across tissues of the same individual than between pairs from the same tissue across different individuals, which was higher than for samples without matching tissue or individual (fig. S6B). Low concordance in ASM calls between individuals may be due to local haplotype context, epigenetic drift, or other nongenetic factors (3, 4, 6, 20, 22, 23). Gaussian mixture modeling (18) showed that allelic differences in methylation (above the 30% threshold) at heterozygous SNPs had a tendency to occur in the same direction (the same allele showing higher methylation than the other) across pairs of samples (fig. S6, C to E).

In order to increase the power to detect SD-ASM at high sensitivity, we pooled the reads across all 49 methylomes and applied the same detection method as for individual samples (fig. S7A) (18). The deep coverage of the combined set (1691-fold total coverage in bisulfite sequencing reads in the combined set of 49 methylomes) increased our power to detect those sequence-associated AIs that were detectable across different tissues and donors (fig. S7, B to D), whereas our power to detect tissue-dependent and donor-dependent SD-ASM was reduced (fig. S8, A and B). The number of accessible heterozygous loci (those having at least six counts per allele), for SD-ASM determination, after pooling rose to 4,913,361, increasing our SD-ASM mapping resolution—measured as an average distance between “index hets”—to 600 base pairs (bp). At the 30% methylation difference default threshold, AIs were detected at 5% of index hets; lowering the threshold to 20%, a total of ~8% index hets showed AIs.

### Sensitivity of the methylome to genetic variation varies across classes of genomic elements

We next explored whether SD-ASM had the tendency to occur within any particular type of genomic element. Using the reads pooled across the 49 methylomes, we observed depletion of SD-ASM within promoters containing CpG islands (Fig. 1B), as well as within CpG islands in general (Fig. 1C), which is consistent with observations that ASM is depleted in CpG islands (4, 21) and that mQTLs are depleted within promoters of genes within CpG islands (24, 25) and that expression quantitative trait methylation is enriched within CpG island shores and not in CpG islands themselves (26). By contrast, and mirroring previous mQTL patterns (25), promoters of genes not in CpG islands showed high levels of SD-ASM (Fig. 1D). We also observed enrichment of SD-ASM downstream from the promoter and into the gene body (Fig. 1, B and D) and positive association between allele-specific expression (ASE) and ASM over exons (Fig. 1A), which is consistent with higher methylation of actively transcribed

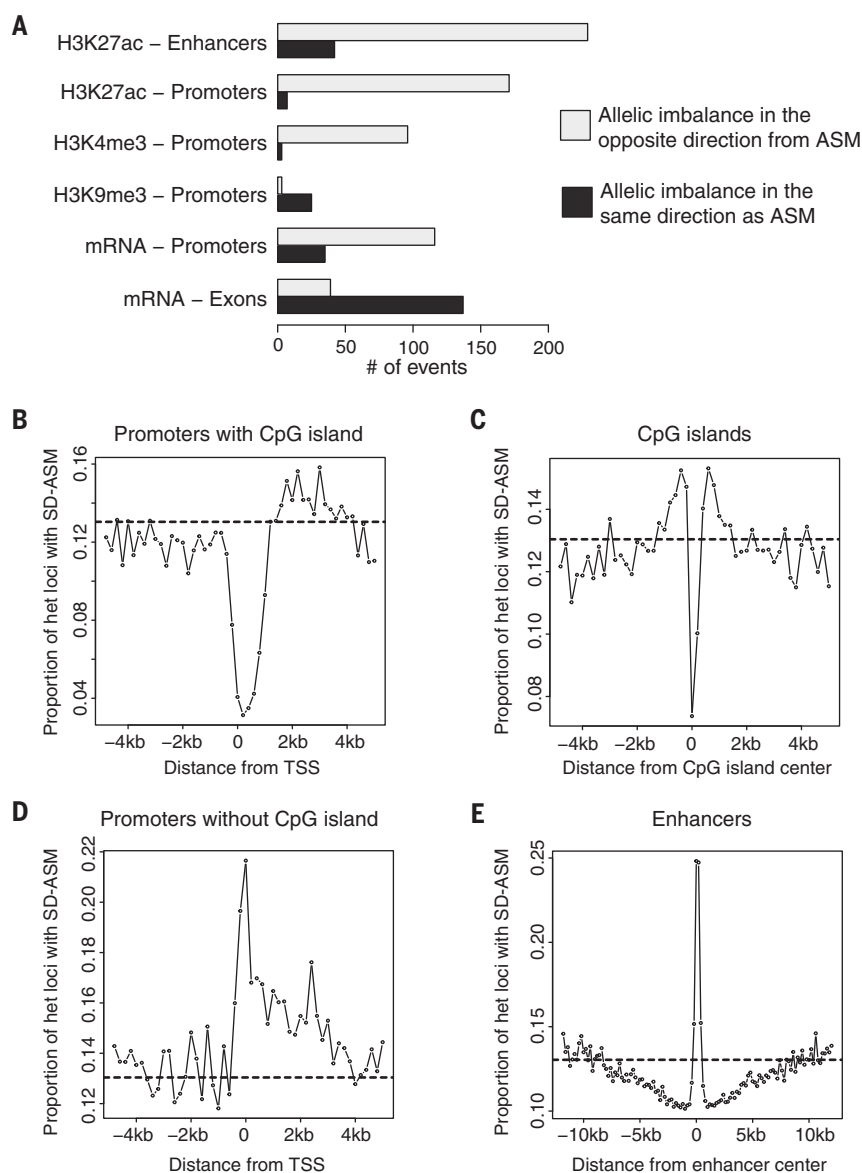
regions, including those on the X chromosome (27) and with the enrichment of mQTLs in regions flanking the transcription start site (TSS) (23, 28). One factor contributing to the ASM, particularly near the transcription start sites (Fig. 1D), may be the presence of transcriptional regulatory signals (29).

SD-ASM was also highly enriched within enhancers (Fig. 1E), which is consistent with previous reports (24, 28). The abundance of TF binding sites within enhancers suggests that SD-ASM may result from disruption of TF binding (23). Under that assumption, our data suggest that TF binding at CpG islands and CpG-rich promoters is buffered against genetic perturbations, whereas the TF binding to non-CpG promoters and enhancers is most sensitive. We also observed a somewhat puzzling mild depletion

of SD-ASM in the flanking regions of enhancers (Fig. 1E), which also suggests buffering in those regions.

### SD-ASM is attributable to differences between allele-specific epiallele frequency spectra

We next asked whether the lack of buffering at SD-ASM loci may result in excess stochasticity and metastability, which is defined by the presence of more than one stable state, each stable state corresponding to an epiallele (single-chromosome methylation pattern). To answer this question, we made use of the deep combined WGBS read coverage across 49 methylomes (table S2) and that each read relates a single variant to a single epiallele. We assessed epialleles by scoring the methylation status of four



**Fig. 1. AIs vary depending on genomic region.** (A) Number of AIs in histone marks and transcription, overlapping ASM loci, over classes of genomic elements. (B to E) Proportions of SD-ASM loci over total heterozygous loci in 200-bp bins near promoters, CpG islands, and enhancers.

homozygous CpG sites ( $4^2 = 16$  possible epialleles) that were the closest to each index het in individual WGBS reads (13, 14) (Fig. 2A). [Our use of the term “epiallele” follows the most recent usage (12, 13) and does not comply with the original definition (30), which implies intergenerational inheritance. Our use of the term “metastability” is consistent with its use in dynamical systems theory and does not imply inheritance of an epiallele during cell division.]

To quantify the amount of stochasticity at index het loci, we used Shannon entropy (18). The entropy values ranged from 0 to 4: An even distribution of frequencies across the 16 possible epiallele patterns produces a maximum entropy score of 4 bits, whereas a complete absence of stochasticity because of maximal “buffering” implies just one epiallele with nonzero frequency and an entropy score of 0 bits. To assess quantitatively any differences in buffering (lack of sensitivity to genetic variation) between SD-ASM and control loci, we identified SD-ASM loci that had sufficient coverage and a close index het without ASM and compared entropies. A total of 6619 (2.7%) of 241,360 loci with SD-ASM met the two criteria (18). We observed a striking difference in entropy, providing a quantitative assessment of the higher stochasticity at the SD-ASM versus control loci (Fig. 2B).

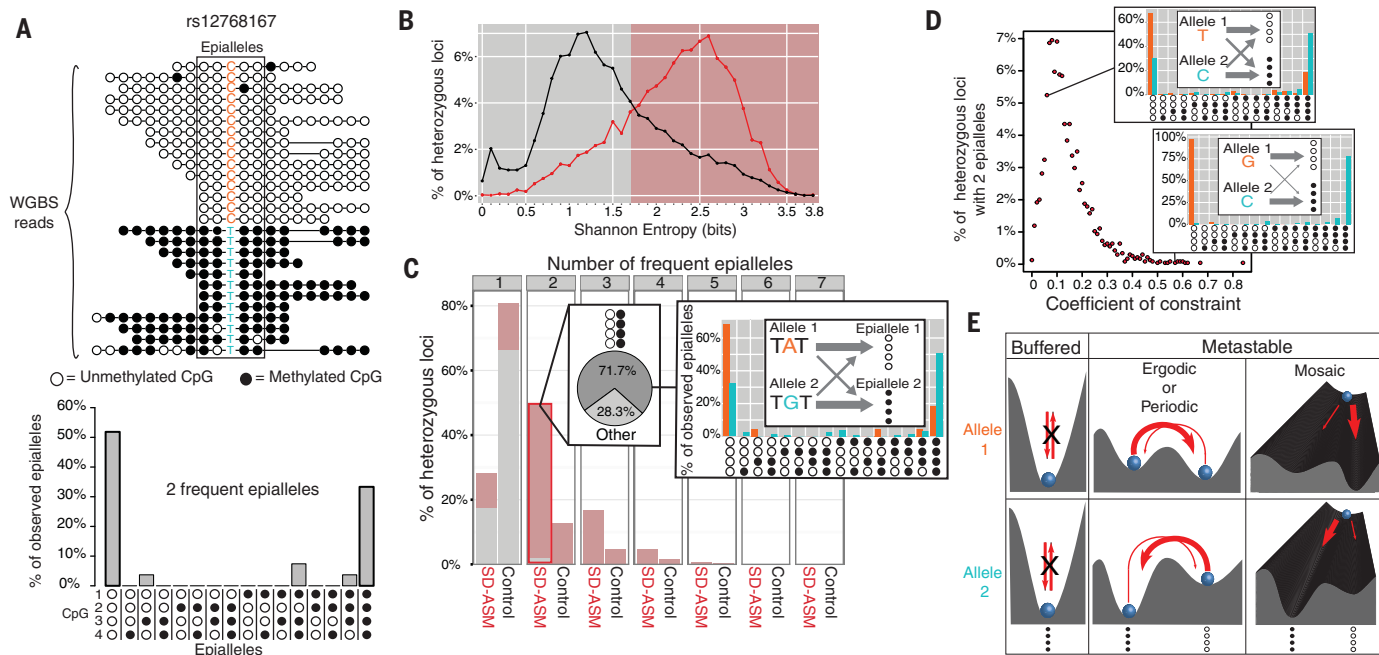
We next examined enrichment for epigenetic polymorphisms at SD-ASM loci. We estimated the number of frequent epialleles for each locus

by sorting the epialleles from the most to the least frequent and identified the minimal-size “top-list” of epialleles that accounted for at least 60% of all the reads with ascertained epialleles. In contrast to the control loci, which typically had only one high-frequency epiallele on the “top-list” and were therefore not epigenetically polymorphic, SD-ASM loci showed multiple frequent epialleles—in most cases, just two (Fig. 2C). By examining the top pairs of epialleles, we found that 71.7% of the pairs consisted of one that was completely methylated and another completely unmethylated (Fig. 2C). This is concordant with previous reports of biphasic (fully methylated and fully unmethylated) distributions of methylation in amplicons with high interindividual methylation variance and in polymerase chain reaction clones with bimodal methylation patterns (3, 31). AIs at SD-ASM loci could be traced to shifts in epiallele frequency spectra between alleles, typically shifts in relative frequencies of the fully methylated and fully unmethylated epialleles (Fig. 2C). We validated the observed excess of stochasticity and the enrichment for the biphasic pattern at SD-ASM loci using an independent WGBS dataset from the Encyclopedia of DNA Elements (ENCODE) (fig. S9, A to C) (18).

We next quantified the relationship between genetic variation and stochastic epialleles. At each locus, we estimated the probabilities of epialleles for each allele (higher probabilities are indicated by thicker arrows in Fig. 2, C and D).

We then quantified the degree to which genetic alleles determine epiallele frequencies using a coefficient of constraint (18), an information-theoretic measure that is a generalization of the  $R^2$  coefficient of determination that is commonly used in genetics and is more appropriate for quantifying genetic determination of stochastic phenotypes. A larger value for the coefficient of constraint value signifies that epigenetic variation is more constrained and determined by genetic variation in cis. Intuitively, a larger coefficient of constraint indicates a larger difference in the epiallele frequency spectra corresponding to the two alleles, implying a higher degree of determination of epiallele frequency spectra by the genetic alleles (Fig. 2D).

There are two general mechanistic models that could explain the effect of sequence variation in cis on epiallele frequency spectra. The ergodic/periodic model stipulates ongoing switching between metastable states, the transitions being stochastic with a possible component of periodicity, such as circadian oscillations. If a sufficient number of stochastic transitions from one epiallele to another occur, that epiallele frequency spectrum depends largely on the sequence-dependent shape of the current energy landscape (state transition probabilities) and not on the epigenetic memory of past events (Fig. 2E). By contrast, the mosaic model stipulates that epialleles are stably transmitted over time and even during cell division, being “frozen” after a period



**Fig. 2. Differences in epiallele frequency spectra causing SD-ASM.**

(A) Example of an epiallele frequency spectrum (bottom) derived from observed epialleles in WGBS reads (top). (B) Histograms of Shannon entropy, in bits, for the epiallele frequency spectra for thehets showing SD-ASM (red) and the nearest (control) hets without SD-ASM (black). (C) Most heterozygous loci with two frequent epialleles show SD-ASM and have entropy larger than 1.7 bits (red portion of the bar), the two epialleles being biphasic (fully methylated or fully unmethylated) 71.7% of the

time. The callout on the right provides an example of a het in which the difference between epiallele frequency spectra of allele 1 (A, orange) and allele 2 (G, blue) explains SD-ASM. (D) Histogram of coefficients of constraint for SD-ASM loci with two frequent epialleles. The callouts illustrate an example het (T/C, top right callout) with a low coefficient of constraint, and another (G/C, bottom right callout) with a high coefficient of constraint. (E) Illustration of buffering in contrast to ergodic/periodic and mosaic metastability.

of initial metastability into one of the stable states. Both models entail a period of metastability, whether past (mosaicism) or current (ergodic/periodic model).

### CTCF binding loci show sequence-dependent stochastic switching and looping

Because of its association with DNA methylation at a large number of binding sites, we next examined the role of CTCF-binding factor (CTCF) in creating the metastable states that correspond to epialleles. Metastability is known to be created by positive (including double-negative) feedback loops (32) that in our case also include interactions in cis, such as the protection against DNA methylation by CTCF binding and reciprocal preference of CTCF for unmethylated DNA (33). The first indication of the role of CTCF binding in metastability came from the observation that the heterozygote with the larger coefficient of constraint (G/C het) also showed larger differences in predicted CTCF binding affinity between the two alleles than the other (T/C het) (Fig. 2D). Considering that the coefficient of constraint is proportional to the differences in epiallele frequency spectra for the two alleles (identical epiallele frequency spectra resulting in coefficient of constraint value of 0), this observation suggested a positive correlation between the coefficient of constraint and the differences in CTCF binding affinity for the two genetic alleles, which was indeed observed (Fig. 3A). In terms of the epigenetic landscape distortion due to genetic variation, we see that sequence variants that show larger differences in CTCF binding affinity also show greater differences in their epigenetic (energy) landscapes, as reflected in the more prominent shifts between alleles in their occupancy of metastable states (as measured by higher values of coefficient of constraint) (Fig. 3A, top). Because CTCF binding and demethylation of its binding site are mutually reinforcing (forming a positive-feedback loop and a metastable state), the model also predicts that the variants associated with higher CTCF binding affinities will show lower methylation, which is indeed the case (Fig. 3A, bottom) as previously observed (23, 24). Taken together, these results suggest sequence-dependent stochastic epigenetic switching between metastable states that is mediated by CTCF binding.

Because the CTCF TF establishes chromatin loops (34), we asked whether the allelic state of methylation also coincided with allelic looping. Toward this goal, we used a study (35) that reports heterozygous SNP loci that associate both with allelic CTCF binding and allelic chromatin looping, as determined by means of chromatin interaction analysis by paired-end tag sequencing (ChIA-PET). Indeed, a total of 44 of those SNP loci were also present in our dataset. Comparing our signals for the methylation state of CTCF binding sites with the predicted CTCF motif disruption scores suggested that SD-ASM is a more accurate indicator of allelic CTCF binding and looping than the motif disruption score (Fig. 3B) (18).

### TF binding sites show sequence-dependent shifts in epiallele frequency spectra and AIs

Analyses of ASM at regulatory elements and eQTLs revealed associations between ASM and allele-specific histone marks with downstream allele-specific transcription (fig. S10, A to F) (18). These results complemented previous studies (22, 23) and suggested involvement of allele-specific TF binding and cofactors in ASM. To examine the role of allele-specific TF binding, we focused on the set of 377 TFs assessed for binding affinity using the high-throughput systematic evolution of ligands by exponential enrichment (SELEX) method (36). As for CTCF, we identified the subset of binding motif loci in a heterozygous state with two frequent epialleles and examined the correlation between coefficient of constraint and difference in predicted allelic binding affinities across these loci for each TF (table S3). Because of the relatively small number of such loci per TF, only 13 showed significant individual  $P$  values (Student's  $t$  test,  $P < 0.05$ ), with only CTCF surviving Bonferroni correction (for testing 377 TFs) (table S3). However, a majority (11 of 13) of the TFs that showed individually significant correlation also showed positive correlation ( $P = 0.01$ , binomial test), which is consistent with the pattern observed for CTCF where larger differences in TF binding affinities correspond to larger distortions in the configuration of metastable states within the landscape (table S3).

Likewise, we next examined for all 377 TFs whether disruptions of their predicted binding sites associated with methylation imbalances. A majority (241) showed SD-ASM enrichment within their binding motifs compared with flanking loci (500 bp on each side) (Fig. 3C and table S4), suggesting that TF binding associates with allelic DNA methylation. The SD-ASM outside of the examined motifs may be attributable to sequence variation within undiscovered binding loci, within motifs of noncoding RNAs, or within loci in physical proximity or contact with regions of perturbed TF activity.

We then examined the relation between allelic differences in motif strengths and methylation levels at SD-ASM loci (18). We observed that for more than half of the TFs tested (207), there was an association between motif strength and level of methylation (Fig. 3C). Most TFs (159) showed gain in methylation on the allele with the disrupted motif, which is consistent with the TF binding either protecting a region from passive methylation (37) or causing active demethylation (Fig. 3C) (38). By contrast, a smaller number of TFs (48), including members of TF families that recruit methyltransferases such as the ETS-domain TF family members (39, 40), showed loss of methylation on the allele with the disrupted motif (Fig. 3, C and D). About a quarter of TFs that show enrichment for SD-ASM show no bias in directionality (table S4), the lack of bias being explainable by contextual behavior at different binding loci, such as for nuclear factor of activated T cells 1 (41) or because of competing TFs at overlapping motifs. Our results support that

TF motif sequences are predictive of proximal CpG methylation levels (23, 42, 43).

We sought to validate the downstream functional consequences of SD-ASM variants, with predicted allelic differences in TF binding, using a luciferase assay. We prioritized cis-overlapping motifs (CisOMs), including those of c-MYC proto-oncogene (cMYC) and tumor suppressor p53 (TP53) that show competitive binding at many loci (44), because CisOMs provide one of the mechanisms of metastability (Fig. 3E), and also those that may have consequences for human disease (table S5). All four SNP validations showed allelic effects on luciferase expression, including two SNPs within CisOMs for cMYC and TP53 and some falling within disease-associated loci (fig. S11) (18), which suggests that SD-ASM helps identify those disease-associated variants that also have functional consequences.

### SD-ASM is enriched near disease-associated loci

We observed that heterozygous variants with SD-ASM were enriched in the neighborhood of variants previously reported as significant in genome-wide association studies (GWASs) of common disease (Fig. 4A) (22, 23, 45). The enrichment was stronger around GWAS variants that have been replicated in multiple studies versus those that have not. To explore more specifically the role of enhancers, we performed a similar enrichment analysis focusing only on GWAS and SD-ASM variants overlapping enhancer elements. Enhancers that contain replicated GWAS variants were significantly ( $P < 0.0001$ ,  $\chi^2$  test) more likely to also contain a variant with SD-ASM than enhancers that did not contain replicated GWAS variants (Fig. 4B). Taken together, these results indicate that AIs provide information about the role of specific loci in common diseases, pointing to the loci that are sensitive to the effects of genetic variation and have functional effects. The enrichment of both GWAS loci and AIs at enhancers, and sensitivity of TF binding to genetic variation discussed in previous sections, provide a mechanistic link between AIs and GWAS associations.

### Variants showing SD-ASM are under purifying selection

Because the variants with large effects are under purifying selection, they tend to be rare, with frequencies below the detection threshold of association studies such as GWAS, mQTL, and eQTL. By contrast, AIs may provide evidence for functional effects even for rare variants that may be detected in only one individual. On the basis of previous studies that have used signatures of purifying selection such as shifts toward smaller derived allele frequency (DAF) to identify functional variants (46, 47), we would expect that ASM variants would also tend to have a lower DAF than those without ASM. Therefore, we obtained DAF estimates from the 1000 Genomes Project (48), ignoring variants that overlapped regions with low accessibility to variant calling. We observed that in nearly every sample in our



The lower bounds from individual samples may underestimate the extent of purifying selection

	Allele 1 [CCGG]	Allele 2 [TTAA]
CTCF	Bound	Not Bound
Methylation	CpG Methylation	CpG Methylation
Competitive Binding	Competitive Binding	Competitive Binding

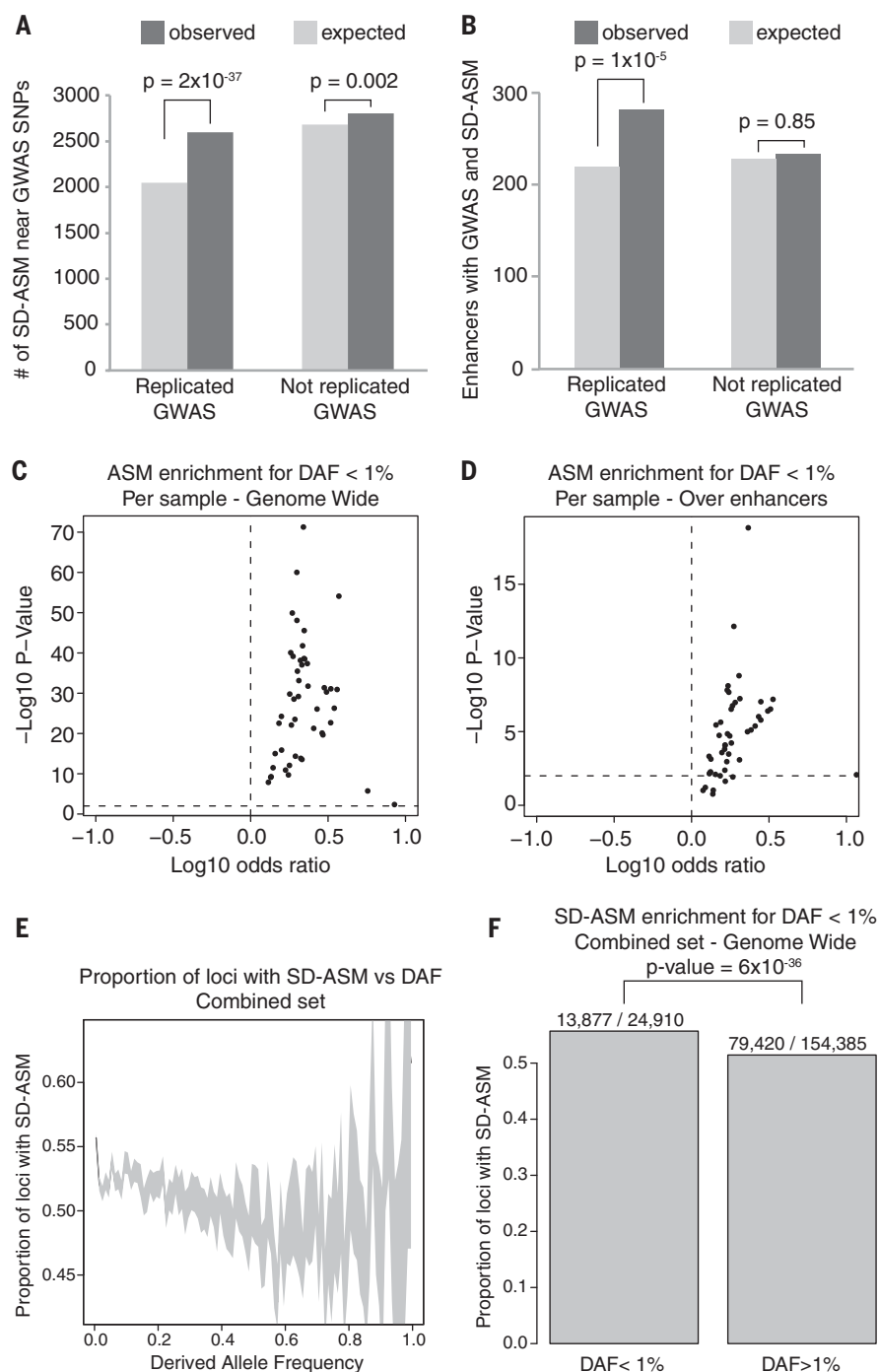
because of underdetection of SD-ASM. We therefore investigated whether an enrichment for rare variants could also be seen for those variants associated with SD-ASM from the combined dataset, using neighboring variants as controls (18). We observed that the chance of a locus having SD-ASM decreased as the derived allele frequency increased (Fig. 4E; there were very few variants with DAF > 50%, causing high variance and large confidence intervals). We further tested whether there was a significant enrichment for variants with DAF < 1% among those with SD-ASM and found that such enrichment was indeed significant (odds ratio 1.18;  $P < 0.0001$ ,  $\chi^2$ ) (Fig. 4F). That enrichment represents an excess of 2184 rare variants among those with SD-ASM compared with controls. Considering that this observed excess represents a set of 11 genomes (nine individuals and two cell lines), we estimate at least ~200 variants with SD-ASM under purifying selection per individual donor.

## Discussion

Taken together, our findings suggest a mechanistic link between sequence-dependent AIs of the epigenome, stochastic switching at gene regulatory loci, and disease-associated genetic variation. Our allelic epigenome map reveals CpG methylation imbalances exceeding 30% differences at 5% of the loci, which is more conservative than previous estimates in the 8 to 10% range (3, 4); a similar value (8%) is observed in our dataset when we lowered our threshold for detecting AI to 20% methylation difference between the two alleles. We observed an excess of rare variants among those showing ASM, suggesting that an average human genome harbors at least ~200 detrimental rare variants that also show ASM.

The methylome's sensitivity to genetic variation is unevenly distributed across the genome. The higher buffering of CpG islands, and associated promoters, may be due to their use by housekeeping genes. Conversely, other promoters and enhancers may show lower buffering because of their presence in tissue-specific genes that tend not to be associated with CpG islands (49). Our findings are consistent with the evolutionary advantages that may stem from the buffering of housekeeping genes against the effects of random mutations—while still retaining the potential for evolutionary innovation through changes in the regulation of less essential genes with tissue-specific expression patterns—and refine reports that suggest that all promoters show epigenetic buffering (50). Highest sensitivity to genetic variation at enhancers, and potentially perturbations in general, is consistent with observations of high cell-to-cell variability of their methylation (28, 51). Validated GWAS loci, and the enhancers within those loci, show enrichment for AIs, suggesting that sensitivity of those loci to genetic variation, and potentially also to environmental influences, may at least in part explain their role in common diseases.

Overall, our results suggest an explanation for the conservation of “intermediate methylation”



**Fig. 4. Association of ASM with disease loci and purifying selection.** (A and B) Enrichment of ASM in the proximity of GWAS loci. ASM hets within 1 kb of GWAS loci are compared with colocalized hets without ASM. (C to F) Evidence of purifying selection acting on rare variants with ASM. [(C) and (D)] Proportion of variants associated with ASM compared with those without ASM among the rare (DAF < 1%) variants across individual methylomes. [(E) and (F)] Proportion of loci with ASM over total heterozygous loci over windows of increasing DAF in the combined set of methylomes. (F) This bar chart summary of the data in (E) shows the excess of SD-ASM variants among those with DAF < 1%.  $\chi^2$  tests were used for significance of enrichments.

states at regulatory loci (52). These intermediate methylation states reflect the relative frequencies of fully methylated and fully unmethylated epialleles corresponding to biphasic on/off switching patterns (31) at regulatory loci marked with SD-

ASM. Moreover, our analyses reveal that the SD-ASM is explainable by allele-specific switching patterns at thousands of heterozygous loci.

Waddington's epigenetic landscape has served as a guiding metaphor for the emergence of cellular

identities during development. The landscape has until now been an abstract construct (53), disconnected from the mechanistic function of gene regulation. Our energy landscapes (Figs. 2D and 3E) may be interpreted as a special type of epigenetic landscape model of gene regulatory interactions in cis in which epialleles correspond to metastable states (attractors) within the landscape (30). As cells transition into a more differentiated state, the cellular epigenome enters a buffered “valley” at a bifurcation point in the landscape (Fig. 2E). Because our current dataset is static, it precludes us from being able to distinguish between the mosaic and stochastic/periodic models that characterize the more and less differentiated states, respectively.

Consistent with proposed theoretical models that define the landscape in terms of potential energy functions (53) and postulate local attractors created by positive-feedback loops (32), our model involves interactions between one or more TFs, DNA methylation, and likely other epigenomic marks. Competitive binding of TFs at CisOMs may be one of the mechanisms of metastability. By putting the metastable states within the landscape in correspondence with epialleles and TF binding, we bring Waddington’s landscapes into correspondence with assayable and quantifiable epiallele frequency spectra and with specific mechanisms of gene regulation.

Our findings are consistent with the role for bistable switching and stochasticity in bacterial gene regulation (54) and extend stochasticity to eukaryotic cells in vivo within their natural tissue context across a diversity of human tissues. One obvious question is the possible purpose of stochasticity at gene regulatory loci across both domains life. The sharp contrast between the “digital” nature of regulatory elements and the “analog” nature of concentrations of upstream TFs, and downstream gene products, suggests that we may be observing a naturally evolved system similar to von Neumann’s “stochastic computer,” an early hybrid analog-digital computer design that can implement regulatory circuits within control systems in which analog quantities are not encoded by using the usual binary or decimal system but are encoded as fractions of “on” states in a stochastic series of on and off states (55). In contrast to a stand-alone stochastic computer, there is a multitude of cells within a tissue, raising the question about the role of intercellular communication within tissue microenvironment in the establishment of a dynamic equilibrium that results in observed methylation averages.

To promote further data analyses and experimental work by the community, we provide an Allelic Epigenome Atlas, a collection of annotations, including AI scores, for all of the ~4.9 million heterozygous loci analyzed here (18). The breadth of the genome-wide AI scores available in the Allelic Epigenome Atlas, which includes the multitude of different individual tissue and cell types profiled here, may serve as additional layers of functional evidence for prioritization of disease-causal candidate variants, including

subthreshold GWAS variants in implicated regions and noncoding variants detected with clinical whole-genome sequencing.

## Materials and methods

We identified heterozygous SNP loci using a joint variant-calling pipeline on WGS datasets from 13 donors from the NIH Roadmap Epigenomics Project. We used ChIP-seq and RNA-seq datasets from a total of 71 various tissues of these donors to detect AIs in histone marks and transcription as described (19). We detected allele-specific methylation using an in-house script on 49 WGBS datasets to compare counts of methylated and unmethylated cytosines proximal to each allele. We tested different regulatory regions and GWAS and eQTL loci for enrichment of AIs using nearby heterozygous SNPs without AI as controls. We compared differences in methylation between alleles with the alleles’ predicted TF motif strengths. We analyzed epialleles by quantifying the methylation patterns of the four closest CpGs to the alleles in single WGBS reads and comparing these patterns between alleles. We calculated Shannon entropy at different loci using epiallele patterns to quantify stochasticity at SD-ASM loci. We calculated coefficient of constraints at SD-ASM loci to determine the constraint of epiallele polymorphisms by genetic variants.

## REFERENCES AND NOTES

- K. Kerker et al., Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.* **40**, 904–908 (2008). doi: [10.1038/ng.174](#); pmid: [18568024](#)
- L. C. Schalkwyk et al., Allelic skewing of DNA methylation is widespread across the genome. *Am. J. Hum. Genet.* **86**, 196–212 (2010). doi: [10.1016/j.ajhg.2010.01.014](#); pmid: [20159110](#)
- Y. Zhang, C. Rohde, R. Reinhardt, C. Voelcker-Rehage, A. Jeltsch, Non-imprinted allele-specific DNA methylation on human autosomes. *Genome Biol.* **10**, R138 (2009). doi: [10.1186/gb-2009-10-12-r138](#); pmid: [19958531](#)
- J. Gertz et al., Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLOS Genet.* **7**, e1002228 (2011). doi: [10.1371/journal.pgen.1002228](#); pmid: [21852959](#)
- A. Hellman, A. Chess, Extensive sequence-influenced DNA methylation polymorphism in the human genome. *Epigenetics Chromatin* **3**, 11 (2010). doi: [10.1186/1756-8935-3-11](#); pmid: [20497546](#)
- C. G. Bell et al., Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLOS ONE* **5**, e14040 (2010). doi: [10.1371/journal.pone.0014040](#); pmid: [21124985](#)
- B. Tycko, Allele-specific DNA methylation: Beyond imprinting. *Hum. Mol. Genet.* **19**, R210–R220 (2010). doi: [10.1093/hmg/ddq376](#); pmid: [20855472](#)
- S. M. Waszak et al., Population variation and genetic control of modular chromatin architecture in humans. *Cell* **162**, 1039–1050 (2015). doi: [10.1016/j.cell.2015.08.001](#); pmid: [26300124](#)
- G. McVicker et al., Identification of genetic variants that affect histone modifications in human cells. *Science* **342**, 747–749 (2013). doi: [10.1126/science.1242429](#); pmid: [24136359](#)
- W. Sun et al., Histone acetylation-wide association study of autism spectrum disorder. *Cell* **167**, 1385–1397.e11 (2016). doi: [10.1016/j.cell.2016.10.031](#); pmid: [27863250](#)
- M. F. Lyon, Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372–373 (1961). doi: [10.1038/190372a0](#); pmid: [13764598](#)
- Z. Shipony et al., Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513**, 115–119 (2014). doi: [10.1038/nature13458](#); pmid: [25043040](#)

- G. Landan et al., Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat. Genet.* **44**, 1207–1214 (2012). doi: [10.1038/ng.2442](#); pmid: [23064413](#)
- E. Florio et al., Tracking the evolution of epialleles during neural differentiation and brain development: D-Aspartate oxidase as a model gene. *Epigenetics* **12**, 41–54 (2017). doi: [10.1080/15592294.2016.1260211](#); pmid: [27858532](#)
- P. Ginart et al., Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an H19 loss-of-imprinting mutant. *Genes Dev.* **30**, 567–578 (2016). doi: [10.1101/gad.275958.115](#); pmid: [26944681](#)
- K. D. Siegmund, P. Marjoram, Y. J. Woo, S. Tavaré, D. Shibata, Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4828–4833 (2009). doi: [10.1073/pnas.0810276106](#); pmid: [19261858](#)
- A. Kundaje et al., Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi: [10.1038/nature14248](#); pmid: [25693563](#)
- Materials and methods are available as supplementary materials.
- J. Rozowsky et al., AlleleSeq: Analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.* **7**, 522 (2011). doi: [10.1038/msb.2011.54](#); pmid: [21811232](#)
- M. D. Schultz et al., Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015). doi: [10.1038/nature14465](#); pmid: [26030523](#)
- D. Leung et al., Integrative analysis of haplotype-resolved epigenomes across human tissues. *Nature* **518**, 350–354 (2015). doi: [10.1038/nature14217](#); pmid: [25693566](#)
- J. N. Hutchinson et al., Allele-specific methylation occurs at genetic variants associated with complex disease. *PLOS ONE* **9**, e98464 (2014). doi: [10.1371/journal.pone.0098464](#); pmid: [24911414](#)
- C. Do et al., Mechanisms and disease associations of haplotype-dependent allele-specific DNA methylation. *Am. J. Hum. Genet.* **98**, 934–955 (2016). doi: [10.1016/j.ajhg.2016.03.027](#); pmid: [27153397](#)
- C. G. Bell et al., Obligatory and facilitative allelic variation in the DNA methylome within common disease-associated loci. *Nat. Commun.* **9**, 8 (2018). doi: [10.1038/s41467-017-01586-1](#); pmid: [29295990](#)
- M. Gutierrez-Arcelus et al., Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* **2**, e00523 (2013). pmid: [23755361](#)
- M. Gutierrez-Arcelus et al., Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. *PLOS Genet.* **11**, e1004958 (2015). doi: [10.1371/journal.pgen.1004958](#); pmid: [25634236](#)
- A. Hellman, A. Chess, Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007). doi: [10.1126/science.1136352](#); pmid: [17322062](#)
- W. A. Cheung et al., Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biol.* **18**, 50 (2017). doi: [10.1186/s13059-017-1173-7](#); pmid: [28283040](#)
- S. G. Park, S. Hannenhalli, S. S. Choi, Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. *BMC Genomics* **15**, 526 (2014). doi: [10.1186/1471-2164-15-526](#); pmid: [24964727](#)
- V. K. Rakan, M. E. Blewitt, R. Druker, J. I. Preis, E. Whitelaw, Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002). doi: [10.1016/S0168-9525\(02\)02709-9](#); pmid: [12127774](#)
- S. N. Martos et al., Two approaches reveal a new paradigm of ‘switchable’ or genetics-influenced allele-specific DNA methylation’ with potential in human disease. *Cell Discov.* **3**, 17038 (2017). doi: [10.1038/celldisc.2017.38](#); pmid: [29387450](#)
- J. Davila-Velderrain, J. C. Martinez-Garcia, E. R. Alvarez-Buylla, Modeling the epigenetic attractors landscape: Toward a post-genomic mechanistic understanding of development. *Front. Genet.* **6**, 160 (2015). doi: [10.3389/fgene.2015.00160](#); pmid: [25954305](#)
- M. T. Maurano et al., Role of DNA methylation in modulating transcription factor occupancy. *Cell Reports* **12**, 1184–1195 (2015). doi: [10.1016/j.celrep.2015.07.024](#); pmid: [26257180](#)
- Y. Guo et al., CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015). doi: [10.1016/j.cell.2015.07.038](#); pmid: [26276636](#)



35. Z. Tang *et al.*, CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015). doi: [10.1016/j.cell.2015.11.024](https://doi.org/10.1016/j.cell.2015.11.024); pmid: [26686651](https://pubmed.ncbi.nlm.nih.gov/26686651/)
36. A. Jolma *et al.*, DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013). doi: [10.1016/j.cell.2012.12.009](https://doi.org/10.1016/j.cell.2012.12.009); pmid: [23332764](https://pubmed.ncbi.nlm.nih.gov/23332764/)
37. R. E. Thurman *et al.*, The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012). doi: [10.1038/nature11232](https://doi.org/10.1038/nature11232); pmid: [22955617](https://pubmed.ncbi.nlm.nih.gov/22955617/)
38. A. Feldmann *et al.*, Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLOS Genet.* **9**, e1003994 (2013). doi: [10.1371/journal.pgen.1003994](https://doi.org/10.1371/journal.pgen.1003994); pmid: [24367273](https://pubmed.ncbi.nlm.nih.gov/24367273/)
39. E. Hervouet, F. M. Vallette, P. F. Cartron, Dnmt3/transcription factor interactions: An alternative mechanism of DNA methylation inheritance. *Genes Cancer* **1**, 434–443 (2010). doi: [10.1177/1947601910373794](https://doi.org/10.1177/1947601910373794); pmid: [21779454](https://pubmed.ncbi.nlm.nih.gov/21779454/)
40. E. Hervouet, F. M. Vallette, P. F. Cartron, Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics* **4**, 487–499 (2009). doi: [10.4161/epi.4.7.9883](https://doi.org/10.4161/epi.4.7.9883); pmid: [19786833](https://pubmed.ncbi.nlm.nih.gov/19786833/)
41. A. Nayak *et al.*, Sumoylation of the transcription factor NFATc1 leads to its subnuclear relocalization and interleukin-2 repression by histone deacetylase. *J. Biol. Chem.* **284**, 10935–10946 (2009). doi: [10.1074/jbc.M900465200](https://doi.org/10.1074/jbc.M900465200); pmid: [19218564](https://pubmed.ncbi.nlm.nih.gov/19218564/)
42. H. Zeng, D. K. Gifford, Predicting the impact of non-coding variants on DNA methylation. *Nucleic Acids Res.* **45**, e99 (2017). doi: [10.1093/nar/gkx177](https://doi.org/10.1093/nar/gkx177); pmid: [28334830](https://pubmed.ncbi.nlm.nih.gov/28334830/)
43. N. E. Banovich *et al.*, Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLOS Genet.* **10**, e1004663 (2014). doi: [10.1371/journal.pgen.1004663](https://doi.org/10.1371/journal.pgen.1004663); pmid: [25233095](https://pubmed.ncbi.nlm.nih.gov/25233095/)
44. K. Kin, X. Chen, M. Gonzalez-Garay, W. D. Fakhouri, The effect of non-coding DNA variations on P53 and cMYC competitive inhibition at cis-overlapping motifs. *Hum. Mol. Genet.* **25**, 1517–1527 (2016). doi: [10.1093/hmg/ddw030](https://doi.org/10.1093/hmg/ddw030); pmid: [26908612](https://pubmed.ncbi.nlm.nih.gov/26908612/)
45. D. Welter *et al.*, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014). doi: [10.1093/nar/gkt1229](https://doi.org/10.1093/nar/gkt1229); pmid: [24316577](https://pubmed.ncbi.nlm.nih.gov/24316577/)
46. D. R. De Silva, R. Nichols, G. Elgar, Purifying selection in deeply conserved human enhancers is more consistent than in coding sequences. *PLOS ONE* **9**, e103357 (2014). doi: [10.1371/journal.pone.0103357](https://doi.org/10.1371/journal.pone.0103357); pmid: [25062004](https://pubmed.ncbi.nlm.nih.gov/25062004/)
47. L. D. Ward, M. Kellis, Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* **337**, 1675–1678 (2012). doi: [10.1126/science.1225057](https://doi.org/10.1126/science.1225057); pmid: [22956687](https://pubmed.ncbi.nlm.nih.gov/22956687/)
48. A. Auton *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: [10.1038/nature15393](https://doi.org/10.1038/nature15393); pmid: [26432245](https://pubmed.ncbi.nlm.nih.gov/26432245/)
49. J. Zhu, F. He, S. Hu, J. Yu, On the nature of human housekeeping genes. *Trends Genet.* **24**, 481–484 (2008). doi: [10.1016/j.tig.2008.08.004](https://doi.org/10.1016/j.tig.2008.08.004); pmid: [18786740](https://pubmed.ncbi.nlm.nih.gov/18786740/)
50. M. T. Maurano *et al.*, Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* **47**, 1393–1401 (2015). doi: [10.1038/ng.3432](https://doi.org/10.1038/ng.3432); pmid: [26502339](https://pubmed.ncbi.nlm.nih.gov/26502339/)
51. S. Gravina, X. Dong, B. Yu, J. Vijg, Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol.* **17**, 150 (2016). doi: [10.1186/s13059-016-1011-3](https://doi.org/10.1186/s13059-016-1011-3); pmid: [27380908](https://pubmed.ncbi.nlm.nih.gov/27380908/)
52. G. Elliott *et al.*, Intermediate DNA methylation is a conserved signature of genome regulation. *Nat. Commun.* **6**, 6363 (2015). doi: [10.1038/ncomms7363](https://doi.org/10.1038/ncomms7363); pmid: [25691127](https://pubmed.ncbi.nlm.nih.gov/25691127/)
53. G. Jenkinson, E. Pujadas, J. Goutsias, A. P. Feinberg, Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nat. Genet.* **49**, 719–729 (2017). doi: [10.1038/ng.3811](https://doi.org/10.1038/ng.3811); pmid: [28346445](https://pubmed.ncbi.nlm.nih.gov/28346445/)
54. H. H. McAdams, A. Arkin, It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69 (1999). doi: [10.1016/S0168-9525\(98\)01659-X](https://doi.org/10.1016/S0168-9525(98)01659-X); pmid: [10098409](https://pubmed.ncbi.nlm.nih.gov/10098409/)
55. A. Alaghi, J. P. Hayes, Survey of stochastic computing. *ACM Trans. Embed. Comput. Syst.* **12**, 1 (2013). doi: [10.1145/2465787.2465794](https://doi.org/10.1145/2465787.2465794)

#### ACKNOWLEDGMENTS

We thank I. Golding for helpful comments and reviewing the paper and Y. Ruan for assisting with allelic looping datasets. **Funding:** A.M. acknowledges support from the Common Fund of the National Institutes of Health (Roadmap Epigenomics Program, grant U01 DA025956). M.G. acknowledges support from the

National Human Genome Research Institute (grant 5U24HG009446-02). W.D.F. acknowledges support from the National Institute of General Medical Sciences (grant GM122030-01). M.K. acknowledges support from the National Institute of Mental Health (grant R01 MH109978) and the National Human Genome Research Institute (grants U01 HG007610 and R01 HG008155). **Author contributions:** V.O., E.L., and A.M. conceptualized research goals and designed studies; R.A.H. and C.C. performed collection and organization of sequencing datasets; C.C., P.P., and R.Y.P. performed processing of sequencing datasets; V.O., E.L., P.P., R.Y.P., J.R., T.G., Z.H., R.C.A., Z.Z., and L.A. performed data curation; Z.H. created and performed variant-calling pipelines with supervision from F.U.; J.R., T.G., R.C.A., and Z.Z. provided software and participated in allelic histone calling with supervision from M.K. and M.G.; V.O., E.L., and I.C. performed bioinformatics, statistical analyses, and presentation of data with supervision from A.M.; J.W.B. performed luciferase assay validation experiments with supervision from W.D.F.; E.L. performed validation experiments on external datasets; L.A. developed the online resource; V.O., E.L., and A.M. wrote the original draft; and E.L., M.K., and A.M. revised and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All allelic datasets are available on <https://genboree.org/genboreeKB/projects/allelic-epigenome>. Code used for bioinformatic analyses can be found at [https://github.com/BRL-BCM/allelic\\_epigenome](https://github.com/BRL-BCM/allelic_epigenome). The whole-genome sequencing, whole-genome bisulfite sequencing, and ChIP-seq reads that were used are stored on the Roadmap Epigenomics BioProject page (accession nos. PRJNA34535 and PRJNA259585) and dbGaP (accession nos. phs000791.v1.p1, phs000610.v1.p1, and phs0007000.v1.p1).

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6409/eaar3146/suppl/DC1](http://www.sciencemag.org/content/361/6409/eaar3146/suppl/DC1)  
Materials and Methods  
Figs. S1 to S12  
Tables S1 to S5  
References (56–78)

26 October 2017; resubmitted 7 May 2018  
Accepted 10 August 2018  
Published online 23 August 2018  
[10.1126/science.aar3146](https://doi.org/10.1126/science.aar3146)

## RESEARCH ARTICLE SUMMARY

## ARCHAEOLOGY

# Ancient lowland Maya complexity as revealed by airborne laser scanning of northern Guatemala

Marcello A. Canuto<sup>\*†</sup>, Francisco Estrada-Belli<sup>\*†</sup>, Thomas G. Garrison<sup>\*†</sup>, Stephen D. Houston<sup>‡</sup>, Mary Jane Acuña, Milan Kováč, Damien Marken, Philippe Nondédéo, Luke Auld-Thomas<sup>‡</sup>, Cyril Castanet, David Chatelain, Carlos R. Chiriboga, Tomáš Drápela, Tibor Lieskovský, Alexandre Tokovinine, Antolín Velasquez, Juan C. Fernández-Díaz, Ramesh Shrestha

**INTRODUCTION:** Lowland Maya civilization flourished from 1000 BCE to 1500 CE in and around the Yucatan Peninsula. Known for its sophistication in writing, art, architecture, astronomy, and mathematics, this civilization is still obscured by inaccessible forest, and many questions remain about its makeup. In 2016, the Pacunam Lidar Initiative (PLI) undertook the largest lidar survey to date of the Maya region, mapping 2144 km<sup>2</sup> of the Maya Biosphere Reserve in Guatemala. The PLI data have made it possible to characterize ancient settlement and infrastructure over an extensive, varied, and representative swath of the central Maya Lowlands.

**RATIONALE:** Scholars first applied modern lidar technology to the lowland Maya area in 2009, focusing analysis on the immediate surroundings of individual sites. The PLI covers twice the area of any previous survey and involves a consortium of scholars conducting collaborative and complementary analyses of the entire survey region. This cooperation among

scholars has provided a unique regional perspective revealing substantial ancient population as well as complex previously unrecognized landscape modifications at a grand scale throughout the central lowlands in the Yucatan peninsula.

**RESULTS:** Analysis identified 61,480 ancient structures in the survey region, resulting in a density of 29 structures/km<sup>2</sup>. Controlling for a number of complex variables, we estimate an average density of ~80 to 120 persons/km<sup>2</sup> at the height of the Late Classic period (650 to 800 CE). Extrapolation of this settlement density to the entire 95,000 km<sup>2</sup> of the central lowlands produces a population range of 7 million to 11 million. Settlement distribution is not homogeneous, however; we found evidence of (i) rural areas with low overall density, (ii) periurban zones with small urban centers and dispersed populations, and (iii) urban zones where a single, large city integrated a wider population.

The PLI survey revealed a landscape heavily modified for intensive agriculture, necessary to sustain populations on this scale. Lidar

shows field systems in the low-lying wetlands and terraces in the upland areas. The scale of wetland systems and their association with dense populations suggest centralized planning, whereas upland terraces cluster around residences, implying local management. Analysis identified 362 km<sup>2</sup> of deliberately modified

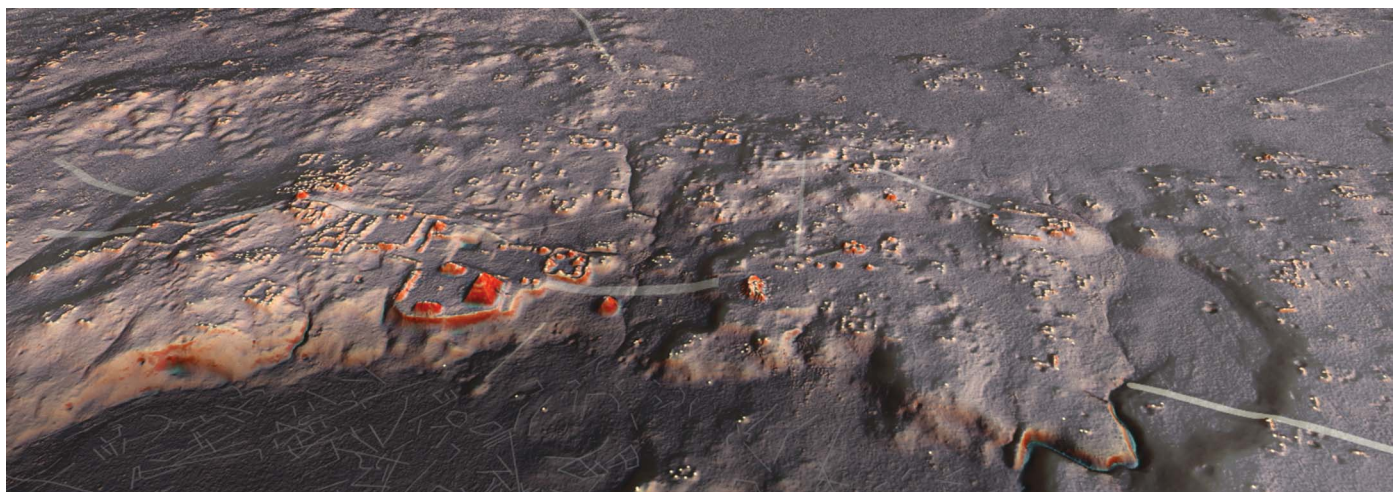
## ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aau0137>

agricultural terrain and another 952 km<sup>2</sup> of unmodified uplands for potential swidden use. Approximately 106 km of causeways within and between sites constitute evidence of inter- and intracommunity connectivity. In contrast, sizable defensive features point to societal disconnection and large-scale conflict.

**CONCLUSION:** The 2144 km<sup>2</sup> of lidar data acquired by the PLI alter interpretations of the ancient Maya at a regional scale. An ancient population in the millions was unevenly distributed across the central lowlands, with varying degrees of urbanization. Agricultural systems found in lidar indicate how these populations were supported, although an irregular distribution suggests the existence of a regional agricultural economy of great complexity. Substantial infrastructural investment in integrative features (causeways) and conflictive features (defensive systems) highlights the interconnectivity of the ancient lowland Maya landscape. These perspectives on the ancient Maya generate new questions, refine targets for fieldwork, elicit regional study across continuous landscapes, and advance Maya archaeology into a bold era of research and exploration. ■

The list of author affiliations is available in the full article online.  
<sup>\*</sup>Corresponding author. Email: [mcanuto@tulane.edu](mailto:mcanuto@tulane.edu) (M.A.C.); [festradi@tulane.edu](mailto:festradi@tulane.edu) (F.E.-B.); [tgarrison1@ithaca.edu](mailto:tgarrison1@ithaca.edu) (T.G.G.)  
<sup>‡</sup>These authors contributed equally to this work.  
<sup>†</sup>These authors contributed equally to this work.  
 Cite this article as M. A. Canuto *et al.*, *Science* **361**, eaau0137 (2018). DOI: 10.1126/science.aau0137



Representation of the archaeological site of Naachtun, Petén, at twilight. Each ancient structure is marked by a yellow dot.

CREDIT: L. AULD-THOMAS AND M. A. CANUTO

## RESEARCH ARTICLE

## ARCHAEOLOGY

# Ancient lowland Maya complexity as revealed by airborne laser scanning of northern Guatemala

Marcello A. Canuto<sup>1\*†</sup>, Francisco Estrada-Belli<sup>1\*†</sup>, Thomas G. Garrison<sup>2\*†</sup>, Stephen D. Houston<sup>3†</sup>, Mary Jane Acuña<sup>4</sup>, Milan Kováč<sup>5</sup>, Damien Marken<sup>6</sup>, Philippe Nondédéo<sup>7</sup>, Luke Auld-Thomson<sup>8†</sup>, Cyril Castanet<sup>9</sup>, David Chatelain<sup>8</sup>, Carlos R. Chiriboga<sup>10</sup>, Tomáš Drápela<sup>5</sup>, Tibor Lieskovský<sup>11</sup>, Alexandre Tokovinine<sup>12</sup>, Antolín Velasquez<sup>13</sup>, Juan C. Fernández-Díaz<sup>14</sup>, Ramesh Shrestha<sup>14</sup>

Lowland Maya civilization flourished in the tropical region of the Yucatan peninsula and environs for more than 2500 years (~1000 BCE to 1500 CE). Known for its sophistication in writing, art, architecture, astronomy, and mathematics, Maya civilization still poses questions about the nature of its cities and surrounding populations because of its location in an inaccessible forest. In 2016, an aerial lidar survey across 2144 square kilometers of northern Guatemala mapped natural terrain and archaeological features over several distinct areas. We present results from these data, revealing interconnected urban settlement and landscapes with extensive infrastructural development. Studied through a joint international effort of interdisciplinary teams sharing protocols, this lidar survey compels a reevaluation of Maya demography, agriculture, and political economy and suggests future avenues of field research.

Lowland Maya civilization flourished for nearly 2500 years (1000 BCE to 1500 CE) in southern Mexico, Guatemala, and Belize (Fig. 1), the central part of which consists of approximately 95,000 km<sup>2</sup> of rolling karst topography interspersed with wetlands (1). Today, this area is largely covered by tropical forest that has severely limited the scale of on-the-ground archaeological research, hindering regional assessments about ancient urbanism, population size, resource management, and sociopolitical complexity.

In 2016, the Pacunam Lidar Initiative (PLI) undertook the largest lidar survey to date of the lowland Maya region, mapping a total of 2144 km<sup>2</sup> of the Maya Biosphere Reserve (MBR) in Petén, Guatemala. This survey mapped 10 non-contiguous areas (polygons) ranging in size from 91 to 454 km<sup>2</sup>, making it possible to characterize ancient settlement, agriculture, and physical infrastructure over an extensive and varied swath of the central Maya Lowlands by using a standardized and consistent dataset. Previous lidar surveys had taken place in Belize, Guatemala, and Mexico (2–8), providing important data on ancient lowland Maya civilization. However,

the PLI sample, undertaken at far larger scale, captured greater variety in topography and ancient settlement for the central Maya Lowlands, the demographic and political heartland (9) of the Classic Maya culture. These data provide an opportunity to address current debates in Maya archaeology and to deepen studies of complex, tropical civilizations in the ancient world.

Some scholars suggest that the Maya Lowlands contained small city-state centers ruled by warring elites (10–13), in settlements supported by a relatively sparse rural population practicing swidden farming (14), with only limited input from intensive agriculture (15–21). In contrast, other views point to a regional network of densely populated cities with complex integrative mechanisms (22–25) that depended on heavy labor investments inside and outside urban cores (15, 26, 27). Even though the latter view has been ascendant in recent years, the absence of regional data has left the debate unresolved. The PLI data cover a sufficiently large area to provide robust support for the latter model and offer further insights about human-environment interactions in the region, including unexpectedly extensive fortifications and road networks.

## Data collection, processing, and field validation

Lidar is an active remote sensing technology that uses laser pulses to map landcover and ground surface in three-dimensional space. The PLI effort collected lidar data with a Teledyne Optech Titan MultiWave multichannel, multi-spectral, narrow-pulse width lidar system, the most advanced aerial lidar system (28) deployed in the region to date (Table 1). The improvements include scanning of the terrain from six different view angles (versus the more usual two views) and higher sensor-range resolution, enabling more precise mapping of topographic signatures under short vegetation. The data collection was designed with a nominal laser density of 15 pulses/m<sup>2</sup> from a flying altitude of 650 m above ground level. The fidelity of this sensor reveals details of local topography, ancient architectural and agricultural features, and damage from looting (Fig. 2).

The data were collected in 10 polygons over 12 designated survey blocks (Table 1). Each survey block was named for the largest known archaeological site within its boundaries; two survey blocks in archaeologically unknown areas were designated Env 1 and Env 2. The current dataset arose from roughly 33.5 billion laser pulses, producing about 60 billion returns, of which 5.2 billion were classified as ground returns. The average ground densities for different polygons range from 1.1 to 5.3 returns/m<sup>2</sup>, the result of varying vegetation densities and canopy coverage across the survey area. Of the total mapped area, 23% has an average ground return density of 1.1 returns/m<sup>2</sup>, 60% has a ground return density range of 1.5 to 2.7 returns/m<sup>2</sup>, and 17% has an average ground return density of more than 5.2 returns/m<sup>2</sup>. These data were interpolated to create bare-earth terrain models with a spatial resolution of 1 m per grid cell.

Identification of archaeological features within the lidar dataset was iterative (29). The PLI study region has been subject to thousands of person-days of pedestrian survey over the course of eight decades of field research. Consequently, there was an extensive “ground-truthed” dataset available for most survey blocks (Fig. 3). Analysts from each participating team made preliminary identifications using existing site maps as training samples. Features of archaeological interest were identified through visual inspection of lidar-derived terrain models. Multiple visualization rasters and manual identification were used to highlight distinct kinds of topographic features (Fig. 4); Table 2 specifies the range of greatest utility in such visualizations. In addition to previously published methods (30–32), we developed a new technique, “prismatic openness” (33), that

<sup>1</sup>Middle American Research Institute, Tulane University, New Orleans, LA, USA. <sup>2</sup>Department of Anthropology, Ithaca College, Ithaca, NY, USA. <sup>3</sup>Department of Anthropology, Brown University, Providence, RI, USA. <sup>4</sup>Department of Anthropology, Washington University, St. Louis, MO, USA. <sup>5</sup>Center for Mesoamerican Studies, Comenius University, Bratislava, Slovakia. <sup>6</sup>Department of Anthropology, Bloomsburg University, Bloomsburg, PA, USA. <sup>7</sup>CNRS-ARCHAM UMR 8096, Université Paris 1-Panthéon-Sorbonne, Paris, France. <sup>8</sup>Department of Anthropology, Tulane University, New Orleans, LA, USA. <sup>9</sup>Laboratoire de Géographie Physique-CNRS UMR 8591, Université Paris 8, Paris, France. <sup>10</sup>Department of Anthropology, Yale University, New Haven, CT, USA. <sup>11</sup>Department of Theoretical Geodesy, Slovak University of Technology, Bratislava, Slovakia. <sup>12</sup>Department of Anthropology, University of Alabama, Tuscaloosa, AL, USA. <sup>13</sup>Universidad de San Carlos de Guatemala, Guatemala City, Guatemala. <sup>14</sup>National Center for Airborne Laser Mapping, University of Houston, Houston, TX, USA.

\*Corresponding author. Email: mcanuto@tulane.edu (M.A.C.); festradi@tulane.edu (F.E.-B.); tgarrison1@ithaca.edu (T.G.G.) †These authors contributed equally to this work. ‡These authors contributed equally to this work.



facilitated detection of small features such as simple house platforms and agricultural terraces.

After preliminary identifications backed by preexisting map and excavation data, ground-truthing and test excavations occurred in 2017 within the Holmul, Corona, Perú, Naachtun, Tintal, Uaxactun, and Zotz survey blocks. This allowed each project to (i) calibrate their feature identifications to ground conditions and (ii) identify patterns in false-positive and false-negative identifications. With this knowledge in hand, a second wave of feature identification took place from August to December 2017.

Overall, approximately 165 km<sup>2</sup> within the study area have either been used as training samples, subsequently “ground-truthed,” or both (Table 3); that is, 7.7% of the entire PLI survey region has been verified through ground survey, although mapped areas were not uniformly distributed (Table 3). Moreover, all field teams within the PLI study region have ground-validated features belonging to every class reported on here: buildings of various types, defensive features, upland and wetland agricultural features, causeways, canals, and reservoirs.

PLI’s preliminary field validation efforts suggest that the feature identifications are accurate if slightly conservative. Most projects report a net increase in field-verified structures of ~15% over those identified through visual analysis of the lidar data. Furthermore, certain features—in particular, small channels or berms as well as large causeways—are clearly visible in the lidar although almost impossible to recognize on the ground without excavation, meaning that in-field visual inspection alone is not always capable of assessing the accuracy of lidar feature identifications. Nonetheless, because ground-truthing efforts are ongoing, and because error rates (false positives and negatives) are likely to vary for different classes of features, all analyses presented here use the 2017 PLI feature dataset with no further adjustments.

### Population estimates

Using the above methodology, 61,480 structures have been identified in the PLI survey region, resulting in an aggregate settlement density of 29 structures/km<sup>2</sup> (Table 4). Because this structure count derives from a much-expanded and regionally representative dataset, it can serve as the basis for a uniquely robust regional population estimate. Even so, producing population estimates from archaeological data requires some assumptions of varying detail and scope from ethnographic analogy and archaeological data; as a consequence, the exercise comes with some degree of uncertainty that only future field research can reduce.

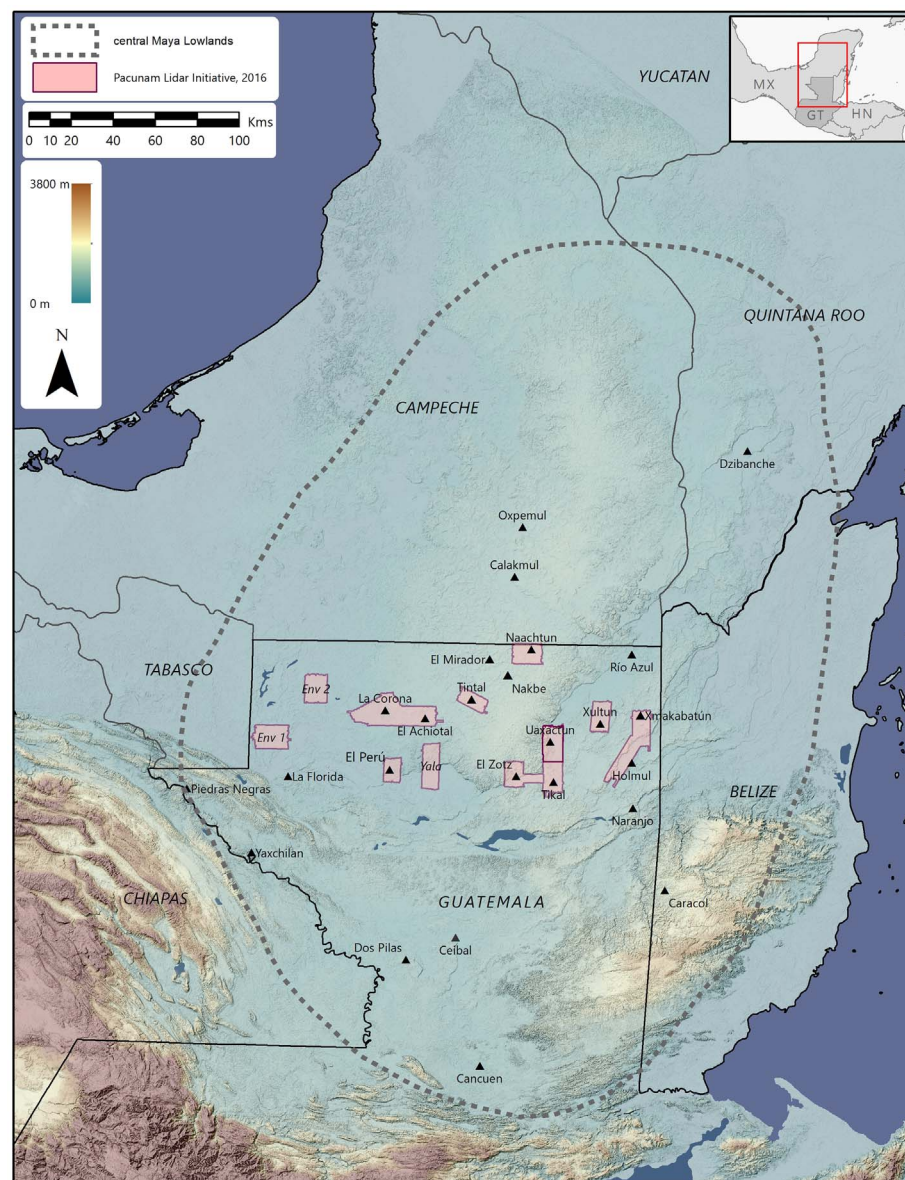
To estimate Late Classic population from structure counts, well-established methodologies for the Maya Lowlands were adjusted by controlling for invisible structures, contemporaneity of use, nonresidential function, non-Late Classic occupation, and numbers of individuals per building (Table 5) (34). The values of these adjustments vary widely across the literature,

given the regional variability of the archaeological record. The geographic scope of the PLI study area, which includes diverse landscapes with distinct settlement histories, necessitates a broader view beyond any single set of local adjustments. As such, values for each adjustment were collated from published scholarship and categorized into five sets according to their overall impact (minimum, low, middle, high, and maximum). All adjustments were then multiplied to derive a composite population index range (*PopIndex*), which, when multiplied by the total number of structures, produced an overall population range (29).

Population estimates were calculated to the nearest half million to avoid false precision and are presented as a range rather than a single

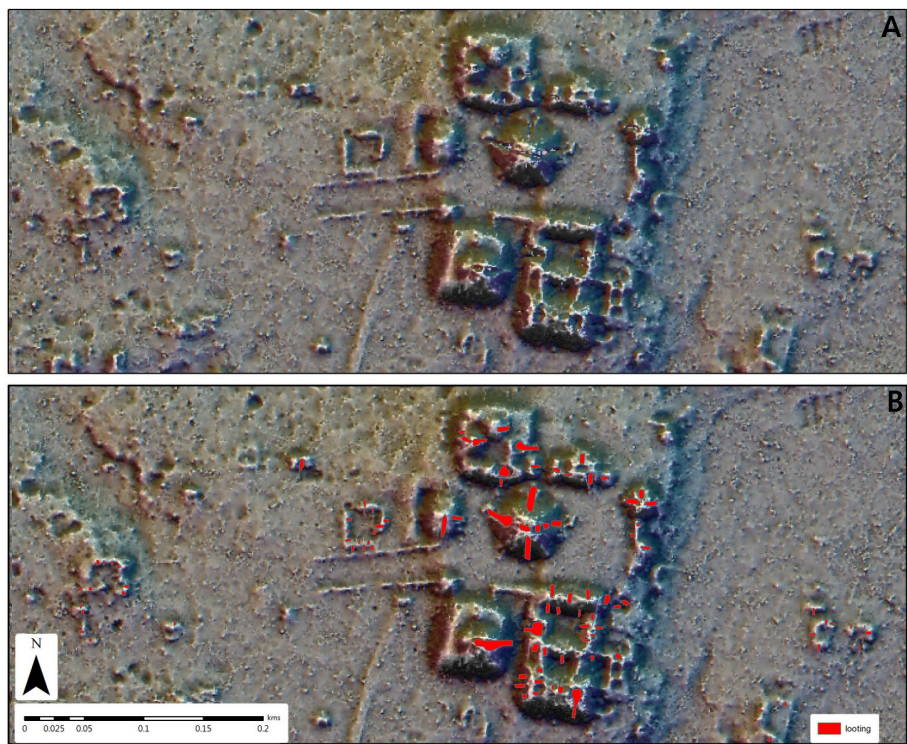
value for the central Maya Lowlands. Given the nature of the available excavation data, this study does not propose a diachronic demographic curve. It only estimates total population during the Late Classic population maximum, 650 to 800 CE, because all PLI survey blocks subjected to archaeological investigation have shown evidence for substantial Late Classic populations.

Following this procedure and guidelines, we estimate a Late Classic population range of 150,000 to 240,000 for the entire PLI survey region. This number amounts to an average density of ~80 to 120 persons/km<sup>2</sup>. This density value accords with the ~100 persons/km<sup>2</sup> suggested by previous nonlidar research (9, 34–37). Inomata *et al.*’s recently published lidar-derived settlement data from the Ceibal region (with



**Fig. 1. Distribution of 12 PLI survey blocks.** Location of 12 PLI survey blocks in relation to the central Maya Lowlands area as well as to the sites mentioned in text; for those survey blocks without a named site, the name of the block is provided (i.e., Env 1, Env 2, and Yala).

Table 1. Lidar ground point averages. Important archaeological sites and lidar ground point densities within each PLI polygon.						
Polygon	Survey block	Main archaeological site(s)	Area (km <sup>2</sup> )	Lidar returns: Total points	Lidar returns: Ground points	Ground points/m <sup>2</sup>
1601	Holmul	Holmul	308.64	8,294,456,212	310,174,174	1.11
1602	Xultun	Xmakabatun	124.02	3,427,027,990	138,067,383	1.13
		San Bartolo				
1603	Tikal	Tikal	146.93	14,614,355,752	926,848,573	2.07
	Uaxactun	Uaxactun	163			
1604	Zotz	El Zotz	143.73	3,666,496,941	197,825,181	1.48
	Naachtun	Naachtun	135.29			
1605	Tintal	Tintal	97.05	3,118,138,624	184,909,324	1.94
1606	Corona	La Corona	431.67	12,388,384,352	1,144,897,043	2.72
		El Achiotal				
1607	Yala	—	172.27	4,495,443,060	331,572,441	1.95
1608	Perú	El Perú-Waka'	91.27	2,302,198,440	97,224,397	1.09
1609	Env 2	—	145.66	3,599,229,811	868,902,279	5.17
1610	Env 1	—	184.28	4,147,724,212	971,717,525	5.32
Total			2,143.81	60,053,455,394	5,172,138,320	2.45



**Fig. 2. Close-up of Xmakabatun.** Image of Xmakabatun that demonstrates the high-fidelity detail of the Teledyne Opech Titan sensor. (A) Looting depressions highlighted by an openness visualization. (B) Looting features as drawn in the field (shown in red) as well as evidence of monumental architecture, causeways, residential structures, ditches, and terraces.

31 observed structures/km<sup>2</sup> across a sample of 470 km<sup>2</sup> (38) are consistent with the settlement data presented here.

There is substantial variability in settlement density across the PLI survey blocks, which reflects a pattern characteristic of the entire cen-

tral Maya Lowlands. That is, some areas will be densely settled (comparable to the Tikal, Xultun, or Naachtun survey blocks) and others will be sparsely occupied (comparable to the Corona, Yala, and Env 1 and 2 survey blocks). However, because average settlement density over large

areas is likely to be similar, the PLI dataset can be extrapolated with confidence over a 95,000-km<sup>2</sup> area known as the “central Maya Lowlands” (Fig. 1). This area excludes the northern Yucatan, the Gulf Coast plains, and coastal Belize and Quintana Roo because they differ physiogeographically from the PLI survey region.

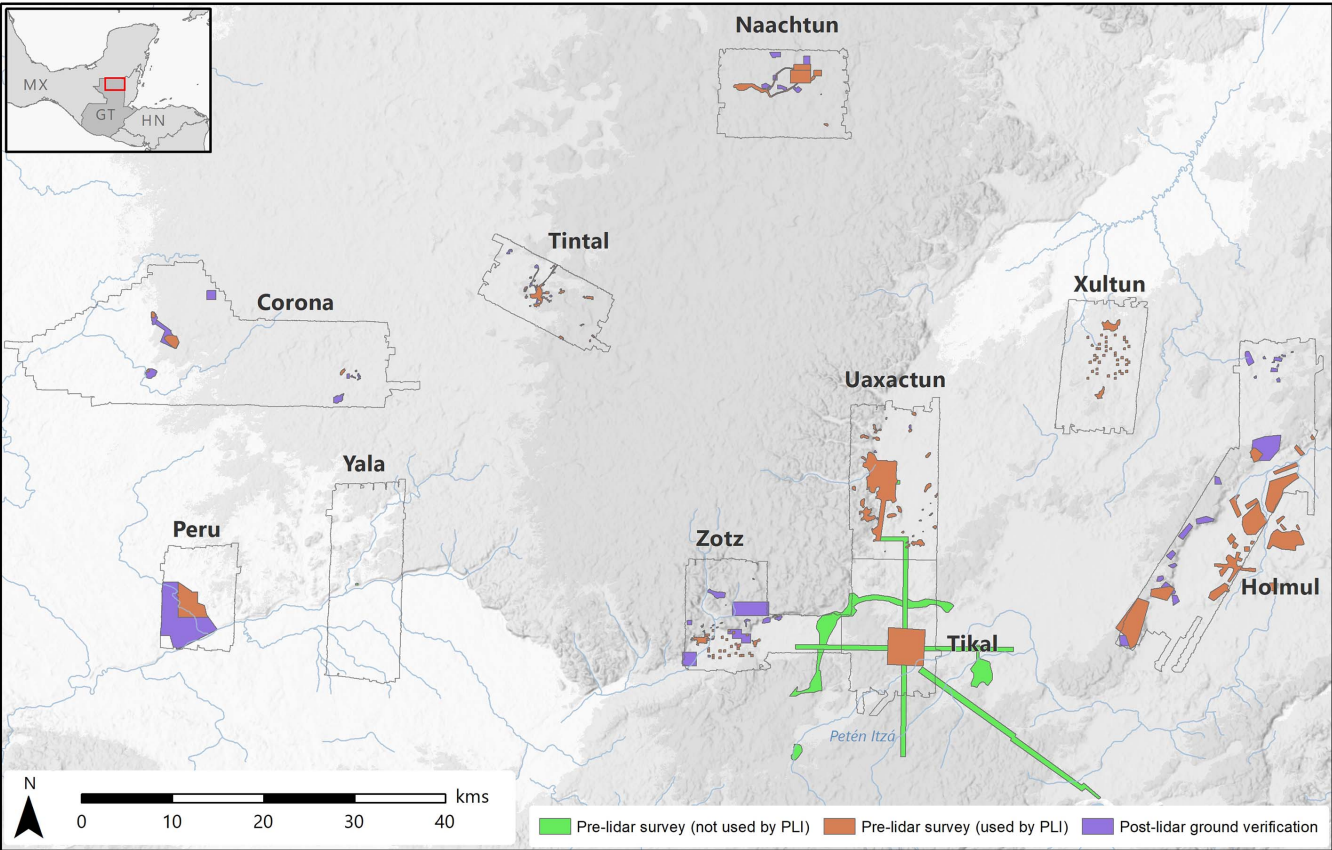
The Late Classic population for the central Maya Lowlands was calculated by multiplying (i) the mean structure density (structures per km<sup>2</sup>) of the sample, (ii) the total area of the central Maya Lowlands (95,000 km<sup>2</sup>), and (iii) the *PopIndex* range (see below):

$$\text{Population} = \frac{\text{Total number of structures}}{\text{Total area surveyed}} \times 95,000 \times \text{PopIndex} \quad (1)$$

This formula suggests that 61,480 identified structures within 2144 km<sup>2</sup> of PLI’s surveyed area can be extrapolated to 2,724,396 structures over an area of 95,000 km<sup>2</sup>. Once multiplied by the *PopIndex* range, these numbers yield a population range of 7 million to 11 million for the central Maya Lowlands during the Late Classic period. Notably, this range aligns with the upper end of prior estimates for the same region, derived from pedestrian surveys (36, 39–41). The incorporation of Inomata *et al.*’s published data from Ceibal (38) to both PLI’s structure and coverage area totals results in the same extrapolated population levels for the central Maya Lowlands—a good indication that the PLI dataset is a representative sample for the region and that the extrapolations are sound.

This estimate is moderate, even conservative, primarily for two reasons: (i) The traditional method of calculating regional population—multiplying area (95,000 km<sup>2</sup>) by a constant population density (80 to 120 persons/km<sup>2</sup>)—has been found to underestimate the resident populations of larger centers (Tikal, Naachtun)





**Fig. 3. Mapped areas within PLI zone.** Map of survey blocks showing (i) areas covered by pedestrian surveys before acquisition of lidar data and (ii) areas where ground verification of the lidar data has been completed to date.

Table 2. Lidar visualizations. Raster visualization combinations created and used for feature identifications.					
Terrain visualization	Raster combo	Color ramp/band combo	Transparency	Utility	Notes
Simple Local Relief Model (SLRM)				Very small buildings, slight depressions, agricultural features	
Layer 1	SLRM	Black to white	0%		Created in Relief Visualization Toolbox (RVT) (30)
Modified Red Relief Image Map (RRIM)				Small buildings, depressions, contour terraces	Based on procedures detailed in (31)
Layer 1	Slope	Custom: white to red	0%		Created in ArcToolbox Slope; visualized in QGIS with blending mode "multiply"
Layer 2	SLRM	Custom: blue-gray-yellow	0%		Created in RVT (30)
Sky-View Factor (SVF)				Depressions, orthogonal buildings	
Layer 1	SVF	Black to white	0%		Created in RVT (30, 32)
Prismatic Openness				Small buildings, agricultural features	"Openness" (33)
Layer 1	Positive openness	White to black	50%		Created in RVT (30)
Layer 2	Negative openness	Black to white	50%		Created in RVT (30)
Layer 3	16-dir. hillshade	RGB: 3,2,1	0%		Created in RVT (30) with vertical exaggeration of 3–5



that become denser with increasing scale (42). Accordingly, the densest urban areas likely would have contained more people than these current approximations. (ii) Although comparison of

lidar-derived structure data and pedestrian survey data shows strong agreement within the PLI survey area, vegetation and taphonomic factors reduce the visibility of small structures in lidar-

based visualizations. Ground validation of the PLI lidar-derived dataset has resulted in a ~15% net increase in archaeological structures. Because both of these factors would result in larger populations, only further fieldwork can render these estimates more accurate.

Regardless of these caveats, the PLI data suggest overall regional populations of such scale that some degree of agricultural intensification (43–45) would have been required to sustain them. Approximately half of the central Lowlands are seasonal wetlands known as *bajos* (15). Because permanent settlements tended to avoid these flood-prone and poorly drained areas, they remained largely uninhabited and could then be available, after added investment, to intensive agriculture (46).

### Agricultural intensification

Early applications of airborne synthetic aperture radar (SAR) suggested the existence of wetland field systems throughout the entire central Maya Lowlands (47–49). However, most of the patterns identified as evidence of intensive agricultural features were later dismissed as “noise” in the imagery or artifacts of the local geology (15, 21, 50–52). Furthermore, given that most of the wetlands in the interior of the peninsula contain clay vertisols, their overall suitability for ancient agriculture was challenged (17, 50, 52–57). Although aerial photography and fieldwork succeeded in identifying both upland and wetland intensive agricultural features in various zones within the lowlands (58–60), the extent of known agricultural improvements remained limited to a few specific areas (18–20, 55, 61–67).

Even though the idea that lowland Maya society only relied on swidden was discarded in the 1980s in favor of a “new orthodoxy” claiming some reliance on a mosaic of more intensive agricultural strategies, the dominant view among specialists was that wetland agriculture was generally limited to areas peripheral to the central Maya Lowlands and that the elevated interior would have relied primarily on swidden agriculture (15, 18, 19, 27, 50, 52, 54, 61–63, 68, 69).

The PLI survey revealed a landscape heavily modified for intensive agriculture: 11 of the 12 survey blocks include agricultural features of some kind. In wetland areas, the survey primarily identified drainage channels. These were designed either to draw water away from natural streams toward infrequently flooded areas or to drain those same areas during major floods. Their dimensions vary from 1 to 2 m in width and 20 to 50 cm in depth. Major channels can measure more than 1 km in length (one extreme case at Tintal is 2.5 km long). Large and small channels intersected at regular intervals, forming nested grids within “channelized” fields (Fig. 5). Overall, these data suggest that wetlands throughout the central Maya Lowlands were modified for agricultural exploitation (16, 43, 61, 62, 70), rather than being used exclusively as sources of water and transportable soils (16, 71). The largest total area of wetland

**Table 3. Extent mapping and ground verification.** Areas in the PLI survey region covered by both pre-lidar pedestrian surveys and post-lidar ground validation efforts (these sometimes overlap).

Site	Area (km <sup>2</sup> )	Pre-lidar pedestrian survey (km <sup>2</sup> )	Used by PLI		
			Pre-lidar pedestrian survey (km <sup>2</sup> )	Post-lidar ground validation (km <sup>2</sup> )	Combined sampled area (km <sup>2</sup> )
Corona	432	2.0	2.0	6.9	6.9
Env 1	184	0.0	0.0	0.0	0.0
Env 2	146	0.0	0.0	0.0	0.0
Holmul	309	46.0	37.0	16.3	50.8
Naachtun	135	7.9	7.9	7.7	15.6
Perú	91	8.2	8.2	21.6	29.8
Tikal	147	86.2	16.3	0.0	16.3
Tintal	97	2.8	2.8	1.0	3.8
Uaxactun	163	21.4	21.4	0.5	21.9
Xultun	124	4.6	4.6	0.0	4.6
Yala	172	0.04	0.0	0.0	0.0
Zotz	144	3.1	3.1	16.0	16.0
Total	2144	182.2	103.2	70.0	165.6

**Table 4. Total structures and settlement density estimates by PLI survey blocks.**

Name	Area (km <sup>2</sup> )	Structures	Structures/km <sup>2</sup>
Corona	432	3,629	8
Env 1	184	307	2
Env 2	146	254	2
Holmul	309	7,207	23
Naachtun	135	11,979	89
Perú	91	4,096	45
Tikal	147	12,341	84
Tintal	97	4,130	43
Uaxactun	163	5,075	31
Xultun	124	5,513	44
Yala	172	740	4
Zotz	144	6,254	44
Total	2,144	61,480	29

**Table 5. Pop/Index calculations.** Set of adjustments used to calculate a “population index” (*PopIndex*), which produces an estimated population when multiplied by structure number. The range in the value of each adjustment indicates the variation within the published record.

Adjustment	Min	Low	Middle	High	Max
1. Invisible or hidden structures	110%	110%	110%	110%	110%
2. Contemporaneity of occupation	75%	80%	83%	87%	90%
3. Residential structures	75%	80%	81.1%	83.5%	85.7%
4. Late Classic	80%	83%	87.5%	92%	95%
5. Persons per structure	4.00	4.37	4.89	5.40	5.60
<i>PopIndex</i> (for Late Classic)	1.98	2.55	3.16	3.97	4.51

agriculture is found in the Naachtun region, with 31.5 km<sup>2</sup> of interspersed channelized zones, followed by Holmul and Tikal. The largest single zones of contiguous wetland fields occur in the Holmul survey block, ranging from 2.3 to 7 km<sup>2</sup>, followed by Tikal and Naachtun. These wetlands may have been selected because of local edaphic conditions, yet they also lie near the most populated areas.

On upland terrain, numerous linear stone features such as terraces and low field walls were identified. These features measure 1 to 2 m in width and are no higher than 50 cm. Some are strictly linear (i.e., contour terraces and check dams); others create enclosures (Fig. 5). Several features form gridded fields with or without enclosure walls, and, in some examples, surround residential structures similar to the *albarradas* (circular walls) of the northern Maya Lowlands (72). Because of their multiple functions, stone features vary widely in length, from only a few meters to >1 km.

By identifying these features, we defined “zones of intensive cultivation” (29) where agricultural improvements were either the only feature in the landscape (e.g., wetland fields) or were interspersed with settlement. The latter case encompasses production systems generally described as infields, orchards, and houselot gardens. A total of 67 km<sup>2</sup> (6659 ha) of wetland fields and 295 km<sup>2</sup> (29,517 ha) of upland fields were classified as zones of intensive cultivation, accounting for ~17% (362 km<sup>2</sup>) of the entire PLI survey area. Investments in upland agriculture are most common in the central and eastern sectors of our sample. The largest area totals occur, by descending order, in the Tikal (64 km<sup>2</sup>), Uaxactun (57 km<sup>2</sup>), Naachtun (47 km<sup>2</sup>), and Xultun (45 km<sup>2</sup>) survey blocks.

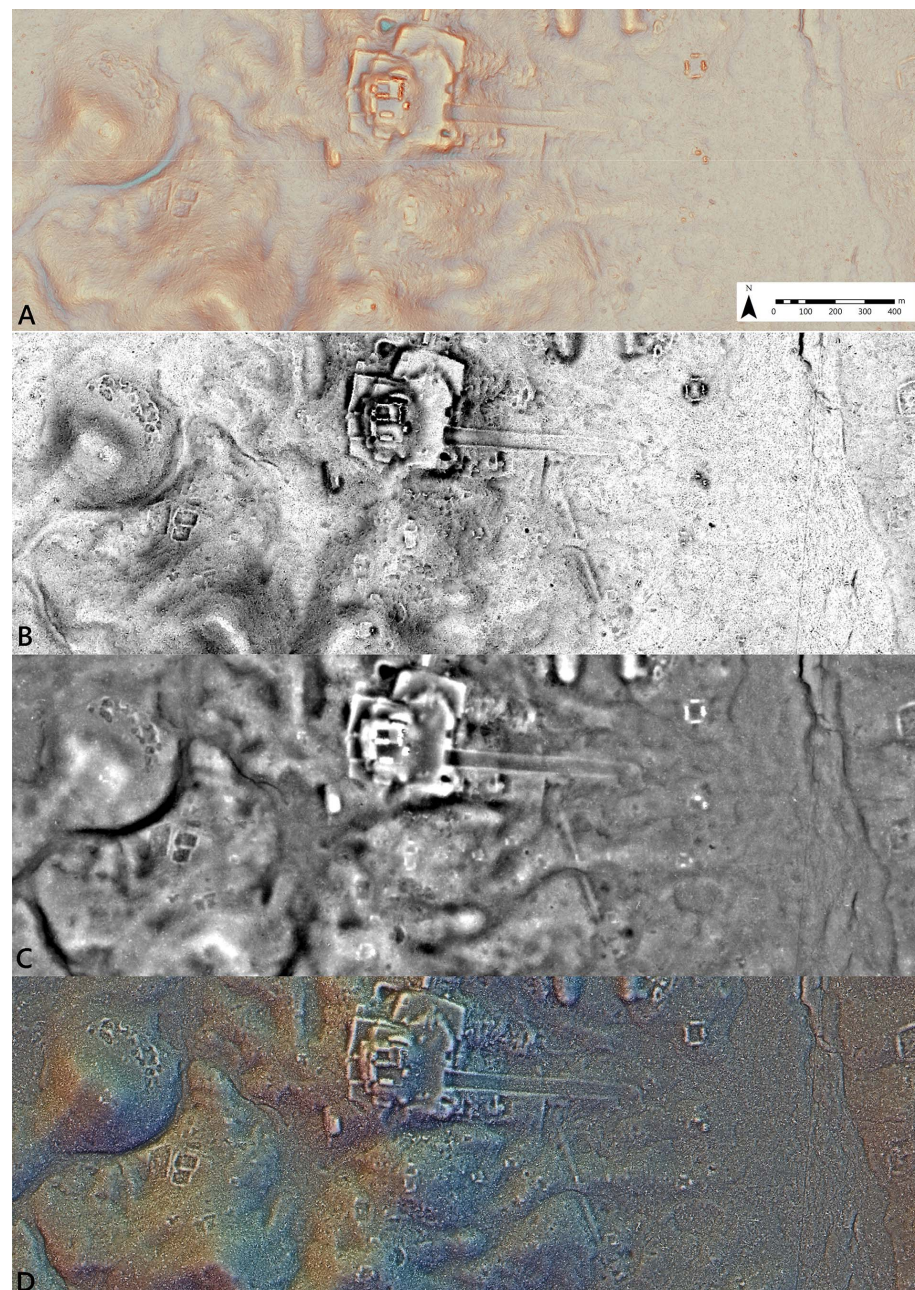
Estimates for the overall productive capacity of this heavily modified agricultural landscape were calculated to determine whether it could have supported the Late Classic populations reported above. Little ethnographic information on the productivity of intensive agriculture in the Maya Lowlands is available, so estimations of overall productive capacity must rely on the productivity of swidden agriculture as the sole point of reference. Because some of the improved agricultural land was unusable (e.g., *bajo*) or prone to rapid degradation (e.g., erodible slopes), many improvements would have succeeded in extending productivity to otherwise unproductive areas (73). However, because some upland features were indeed placed on optimal land (i.e., gentle slopes) and the moisture-retaining properties of terraces improve agricultural productivity (37, 74), it is likely that some upland features did improve overall yields beyond what traditional swidden agriculture could produce. Nonetheless, these calculations assume that all zones of intensive cultivation resulted in the same productive capacity as swidden.

Ethnographic data on the productivity of traditional lowland Maya swidden agriculture were collated and standardized to estimate the number of hectares necessary to support one

person. Data on swidden productivity were drawn from Cowgill (75–77), Griffin (78), Schwartz (79, 80), Nations and Nigh (81), and Ford and Nigh (37). Standardization entailed, among other adjustments, normalizing the annual productivity of different fallow regimes to multiyear averages. These calculations resulted in a mean productivity value of 0.48 ha per person [the mean of the range (0.34 to 0.62 ha per person) resulting from using different adjustments from the above-cited sources] (table S3). This value was then applied to all the land in the PLI sample (1314 km<sup>2</sup>) available for agricultural production (zones of intensive cultivation

and nonwetland rural areas) to derive a maximum supportable population. Finally, by the same logic applied to population estimates, this number was adjusted to account for the possibility that these features of agricultural intensification were not all coeval. These calculations (Table 6) suggest that the total amount of land available for agricultural production in the PLI sample would have been more than sufficient to sustain the population levels reported above (29).

Consequently, the PLI data indicate some combination of the following possibilities: (i) There was capacity for surplus food production; (ii) substantial portions of agricultural land



**Fig. 4. Visualizations used for lidar analysis.** (A) Red Relief Image Map (RRIM); (B) Sky-View Factor (SVF); (C) Simple Local Relief Model (SLRM); (D) Prismatic Openness.



were dedicated to the production of nonfoodstuffs (such as cotton or cacao); (iii) large areas of the uplands remained covered in forest (37, 82, 83); or (iv) populations were higher than our current estimates. Agricultural intensification could also have been achieved using strategies such as

fallow-period shortening, multicropping, and manipulation of crop sociology; these would not have occasioned changes to the landscape visible in lidar. In addition to corn, a variety of other edible plants and fruit trees are known to have been grown in and around improved

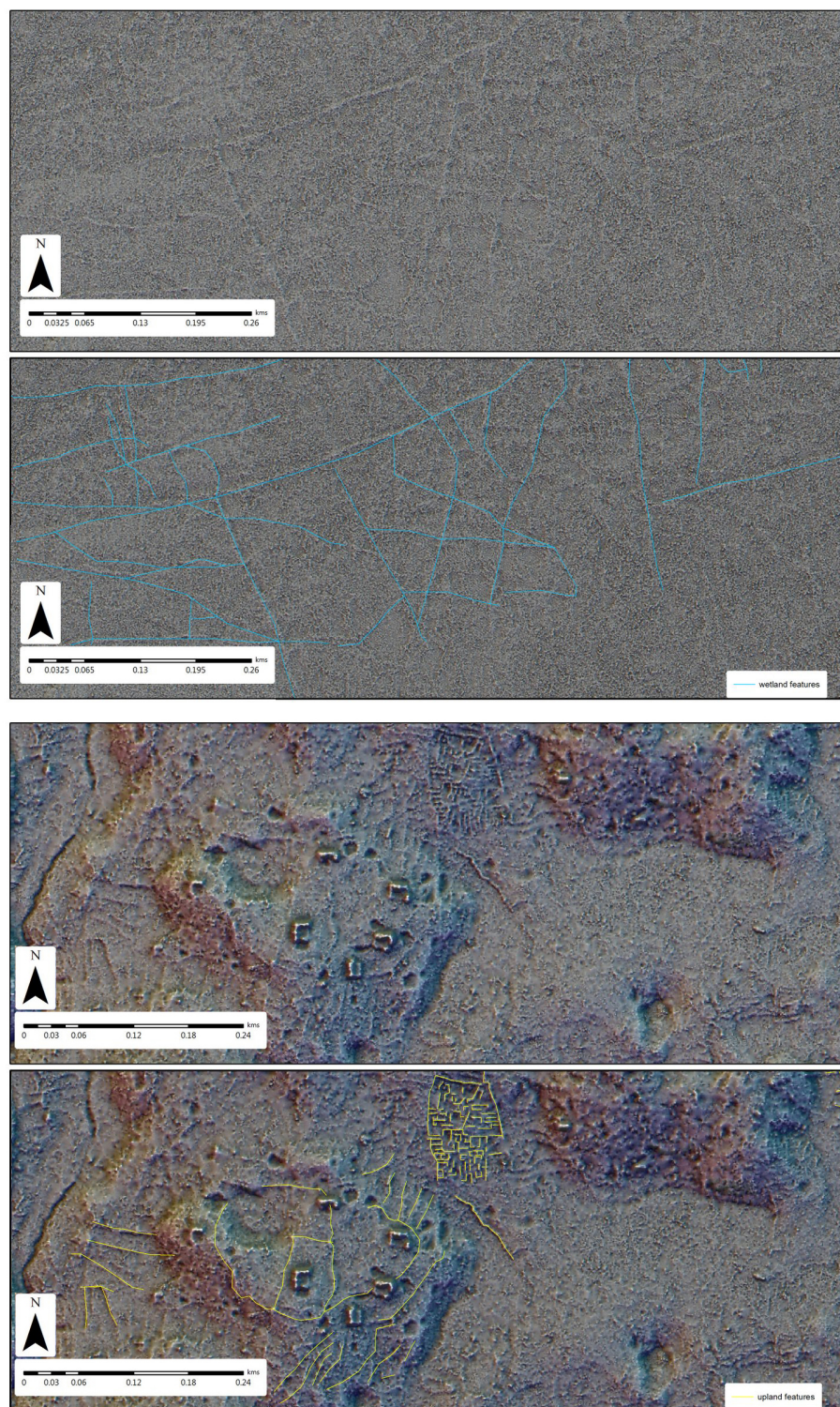
upland fields (14). Therefore, this study's reliance on both physical infrastructure (as an index for intensification) and ethnographic data of modern swidden productivity likely underestimates the total productive capacity of ancient lowland Maya agriculture.

This dataset also demonstrates great variability in agricultural investment across the region (Fig. 6). The Corona survey block (432 km<sup>2</sup>) has <1% of its available land improved for cultivation, whereas the Tikal block (147 km<sup>2</sup>) has as much as 70%. Indeed, the lowest densities of agricultural features occur in the less populated western blocks of the sample, and the highest densities are located in the central and eastern sectors where settlement is densest. Across the PLI sample, areas defined here as periurban and urban (see below) contained the greatest relative density of agricultural features. Urban cores contained no agricultural features, whereas rural areas contained relatively fewer improved fields (Table 7 and Fig. 6). The scale of wetland field systems and their close association with densely populated areas suggest some degree of centralized involvement. Conversely, agricultural modifications of the uplands cluster around residential groups, which suggests that they were likely managed by household or small corporate groups. Upland improvements and terracing are common, likely making a substantial contribution to overall regional production (26).

The co-occurrence of dense settlement with agricultural improvements suggests that investments to maximize agricultural production were directed to areas where population sprawl limited the possibility for extensive farming. Despite these efforts, however, the estimated yield of the agricultural lands within the most heavily modified survey blocks—Tikal, Naachtun, Xultun, and Perú—would not have been sufficient to sustain their populations (Table 6). Even if the zones of intensive cultivation within these survey blocks had been twice as productive as swidden fields, none of these areas would have harvested enough to achieve self-sufficiency. That is, people living in these blocks were partly dependent on foodstuffs imported from surrounding areas. Such data point to a lowland Maya landscape that was a mosaic of interdependent densely populated political centers and extensive periurban hinterlands engaged in a complex set of interactions.

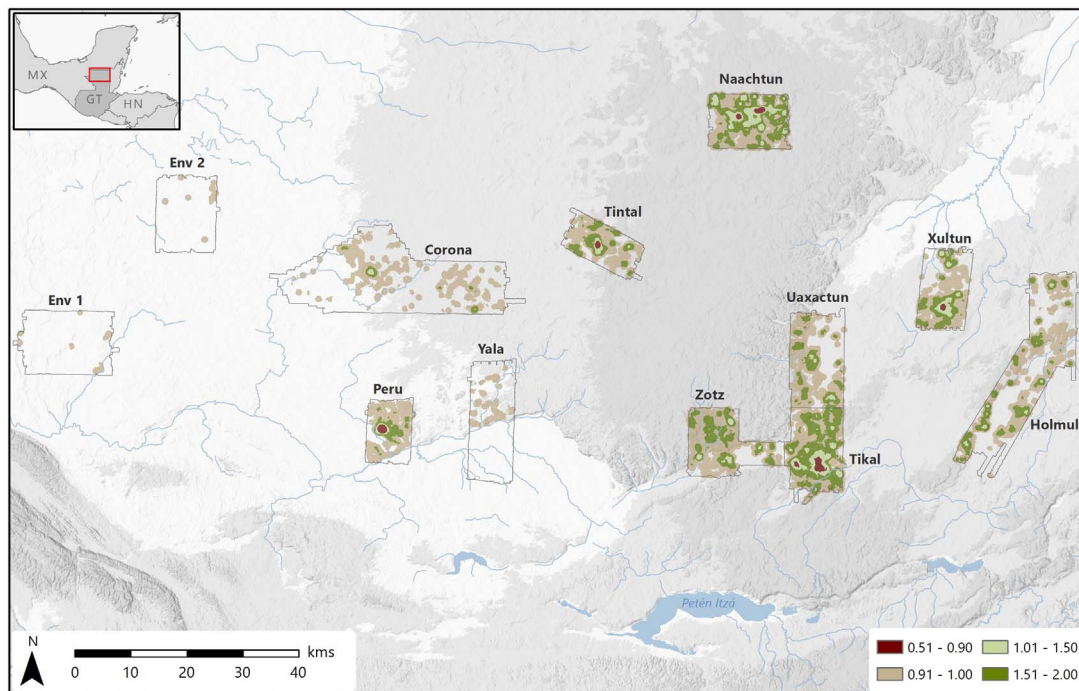
### Urbanization

Settlement studies at lowland Maya sites have often used categories such as “site core,” “periphery,” and “rural” to delimit zones of different settlement density (84–86). These are useful categories at local scales. However, the scope of the PLI dataset allowed for a comparison of settlement density at a regional level. A shared settlement density scale was developed by applying figures compiled by Rice and Culbert (35) (Table 8). Their data, based on ground survey of sites with variably defined limits, provided the justification for four classes; a fifth—“vacant”—was added for a total of five classes. In summary, five structure

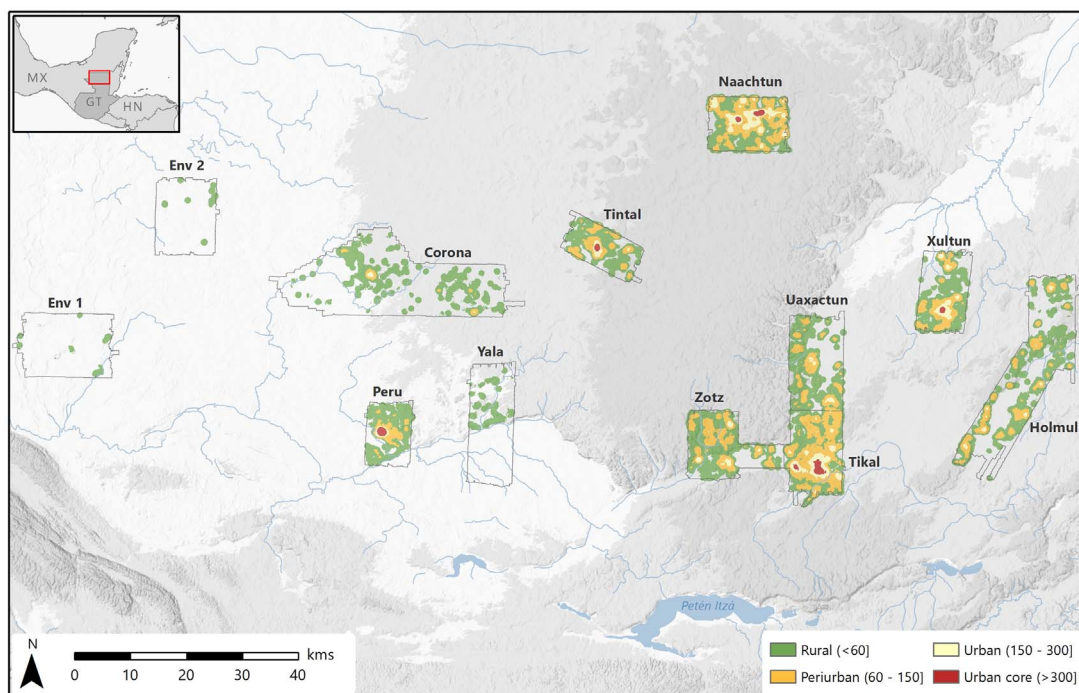


**Fig. 5. Wetland and upland agricultural features.** Examples of intensive wetland fields featuring networks of canals (above) and upland field zones featuring low stone wall features and terraces.





**Fig. 6. Agricultural features relative to population densities.** Map of survey blocks showing deviations from expected density of upland field areas relative to settlement density class area. Expected (i.e., “no difference”) value is 1.0.



**Fig. 7. Settlement density patterns.** Density values are expressed as structures/km<sup>2</sup>.

density categories—vacant, rural, periurban, urban, and urban core—were defined for the entire dataset (29) to discern differences not only in the size and density of cities but also in the relationship between cities and hinterlands (86–89).

The “urban core” class was defined on the basis of structure densities at the heart of the largest Maya centers, such as Tikal (~700 structures/km<sup>2</sup>). “Urban” describes the structure density at the heart of many smaller cities

within the PLI study area (e.g., Uaxactun, Zotz) as well as the densely settled surroundings of major cities’ urban cores (e.g., Xultun, Naachtun, and Tikal). “Periurban” zones (also known in the literature as “urban fringes”) combine characteristics of both urban and rural areas (90). The parameters of this category encompass those areas designated by prior surveys as “peripheral” to the largest Maya sites as well as sites often characterized as “minor centers” whose economic

and political fortunes were closely articulated with larger cities nearby (91–93). Our “rural” class follows existing descriptions of settlement density in the hinterlands of smaller cities and minor centers. Finally, we classified as “vacant” all zones with fewer than 10 structures/km<sup>2</sup>.

At the regional level, sprawling urban and periurban settlement zones covered large areas in the east, whereas the west remained mostly rural (Fig. 7). Urban core densities ran as high

**Table 6. Maximum sustainable population of each survey block.** Calculations designed to determine the maximum sustainable population within each survey block based on (i) its total measured agricultural area (values adjusted for Late Classic contemporaneity using AgroIndex) and (ii) a minimum swidden area of 0.48 ha/person (mean value of range reported in table S3).

Survey block	Uplands (ha)	Wetlands (adj. ha)	Modified uplands (adj. ha)	Pop. capacity (0.48 ha/person)	Population (PopIndex 3.16)	Under/over capacity	% capacity
Naachtun	3,387	2,288	2,574	17,062	37,889	20,827	222%
Tikal	6,069	407	4,305	22,298	38,892	16,594	174%
Xultun	1,992	0	2,847	10,008	17,438	7,430	174%
Perú	3,892	0	351	8,777	12,956	4,179	148%
Tintal	5,409	380	746	13,516	13,063	-453	97%
Uaxactun	4,238	281	3,958	17,532	16,052	-1,480	92%
Zotz	11,485	79	94	24,112	19,781	-4,331	82%
Holmul	12,866	1,350	2,322	34,206	22,796	-11,410	67%
Corona	19,399	0	14	40,150	11,478	-28,671	29%
Yala	7,738	0	82	16,172	2,341	-13,831	14%
Env 1	8,243	0	130	17,319	971	-16,347	6%
Env 2	10,487	52	44	21,887	803	-21,084	4%
PLI survey area	95,205	4,836	17,468	243,039	194,460	-48,579	0%

**Table 7. Correlation of upland agricultural features with settlement density.** Correlation matrix of intensive agriculture upland field zones and settlement density zones in m<sup>2</sup> (upland zones strongly correlate with periurban and urban zones; *N* = 12).

		Vacant	Rural	Periurban	Urban	Urban core	Intensive agr. fields
Vacant	Pearson correlation	1	0.612*	-0.368	-0.381	-0.460	-0.289
	Significance (two-tailed)		0.034	0.239	0.221	0.133	0.361
Rural	Pearson correlation	0.612*	1	0.278	-0.021	-0.152	0.248
	Significance (two-tailed)	0.034		0.382	0.949	0.637	0.438
Periurban	Pearson correlation	-0.368	0.278	1	0.835**	0.694*	0.749**
	Significance (two-tailed)	0.239	0.382		0.001	0.012	0.005
Urban	Pearson correlation	-0.381	-0.021	0.835**	1	0.775**	0.645*
	Significance (two-tailed)	0.221	0.949	0.001		0.003	0.023
Urban core	Pearson correlation	-0.460	-0.152	0.694*	0.775**	1	0.505
	Significance (two-tailed)	0.133	0.637	0.012	0.003		0.094
Upland area	Pearson correlation	-0.289	0.248	0.749**	0.645*	0.505	1
	Significance (two-tailed)	0.361	0.438	0.005	0.023	0.094	

\*Correlation is significant at the 0.05 level (two-tailed). \*\*Correlation is significant at the 0.01 level (two-tailed).

as 1000 to 2000 persons/km<sup>2</sup>, whereas rural areas were as low as 50 persons/km<sup>2</sup> over continuous areas. However, the PLI sample also revealed striking regional differences in urbanization patterns (Table 9).

Cluster analysis of the settlement patterns within the PLI sample revealed several intra-regional patterns. Because the PLI survey blocks varied in shape and in total area, standardized 80-km<sup>2</sup> sample areas were designed for cluster analysis. These areas were delineated to maintain the primary urban center in as central a location as possible within the sample area. Because of the variable shape and dimensions of each PLI survey block, sample areas could not be drawn with identical proportions; nonetheless, areas were 80 km<sup>2</sup> in all cases, with proportions as standardized as possible. The Corona and Holmul survey blocks were large enough to de-

lineate two distinct sample areas for each survey block. These two additional sample areas were named Achiotal and Xmakabatur (for the Corona and Holmul survey blocks, respectively) after the name of the primary center within each area.

Cluster analysis of these 14 areas (Fig. 8) (29) indicates that settlement patterns at the regional level fall into three classes: (i) rural areas with low overall density and limited political integration (Env 1, Env 2, Yala, Corona, Achiotal); (ii) periurban areas containing small urban centers with large periurban and rural populations (Uaxactun, Holmul, Xmakabatur, and Zotz); and (iii) urban areas with periurban and rural populations surrounding and likely integrated with a single large urban center. Within this third class are two subclasses: moderately urbanized areas with more substantial periurban and rural

populations (Tintal and Xultun), and highly urbanized areas in which most of the population lived in urban core, urban, and periurban zones (Naachtun and Tikal).

Such variety cannot be explained by sole reference to topographic, pedological, or ecological conditions; political and economic forces, including conflict, likely exerted commensurate or stronger influence. Furthermore, nearly all the areas identified as urban were also those incapable of sustaining themselves agriculturally, whereas all the rural and periurban areas were capable of producing food surpluses to feed the rest of the sample's population. These data suggest that lowland Maya settlement was characterized by both high variability and high interdependence along a rural-urban continuum rather than by a dispersed, homogeneous, low-density settlement (94).

**Table 8. Late Classic structure density classes, modified from Rice and Culbert (35).**

PLI density class		Site centers			Site peripheries			Rural	
Class	Density	Site	Area (km <sup>2</sup> )	Density	Site	Area (km <sup>2</sup> )	Density	Region	Density
Urban core	300+	Caracol	2.2	300					
		Tikal	9	235					
Urban	150–300	Ceibal	1.6	222					
		Becan	3	171	Tikal (16 km)	7	181		
		Tayasal	2.5	128	Ceibal	13.6	116	Belize Valley	116
Periurban	60–150	Cenote	0.5	128					
		Uaxactun	16	106					
								Tikal/Yaxha	60
		Nohmul	4	58				Lakes Macanche-Salpeten	57
								Lakes Yaxha-Sacnab	47
Rural	<60							Lakes Quexil-Petenxil	36
								Tikal (survey strips)	39
								Uaxactun	30
								Nohmul	12
Vacant	≤10								

**Table 9. Relative frequency of structures in each density class by PLI survey block.**

Name	Rural	Periurban	Urban	Urban core
Corona	81%	14%	5%	0%
Env 1	100%	0%	0%	0%
Env 2	100%	0%	0%	0%
Holmul	43%	48%	8%	0%
Naachtun	13%	41%	36%	10%
Perú	35%	24%	6%	34%
Tikal	12%	49%	25%	14%
Tintal	28%	48%	11%	12%
Uaxactun	46%	47%	7%	0%
Xultun	25%	42%	25%	7%
Yala	100%	0%	0%	0%
Zotz	36%	54%	11%	0%
All	30%	43%	18%	9%

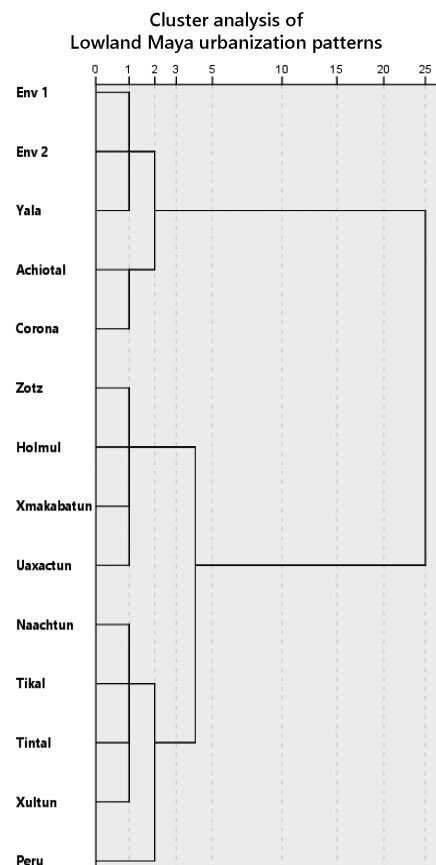
A word of caution is warranted about assuming contemporaneity of all structures in this classification. For instance, three survey blocks—Zotz, Uaxactun, and Holmul—hold substantial centers that were abandoned in the Late Pre-classic period and only partly resettled in the Classic period. However, because these blocks already form a statistically distinct group, we are confident that they would remain clustered even after factoring out all earlier settlement in those three areas.

A few other patterns bear mentioning. First, the city of El Perú-Waka' with a sustained density of 1100 structures/km<sup>2</sup> (95)—nearly double that of Tikal—exhibits a uniquely nucleated urban core that transitions rapidly to a rural hinterland. The singular urban morphology of El Perú-Waka', although it generally fits the urban area pattern, might reflect distinct economic or political influences on its development and organization. Second, although the Corona block was essen-

tially rural, it did include several small pockets of urban density. This region had a strategic role in the expansion of the hegemonic Kaanul polity during the sixth through eighth centuries CE (96–102), so settlement nucleation here was probably conditioned by regional geopolitics in addition to local demographic processes. Finally, the urban and periurban settlement zones of Tikal extend over at least 76 km<sup>2</sup>, representing one of the largest continuous settlement zones in the Maya Lowlands (26).

### Engineering and infrastructure

Lidar data also elucidate the extent of lowland Maya investment in water management, regional communication, and defense. The Maya constructed reservoirs that required considerable labor (103). At the household level, reservoirs, quarries, wells, and underground cisterns were commonly cut into bedrock to collect rainwater (104–110). Within urban centers, the ancient

**Fig. 8. Cluster analysis of density patterns.**

Ward linkage method, squared Euclidean distance.

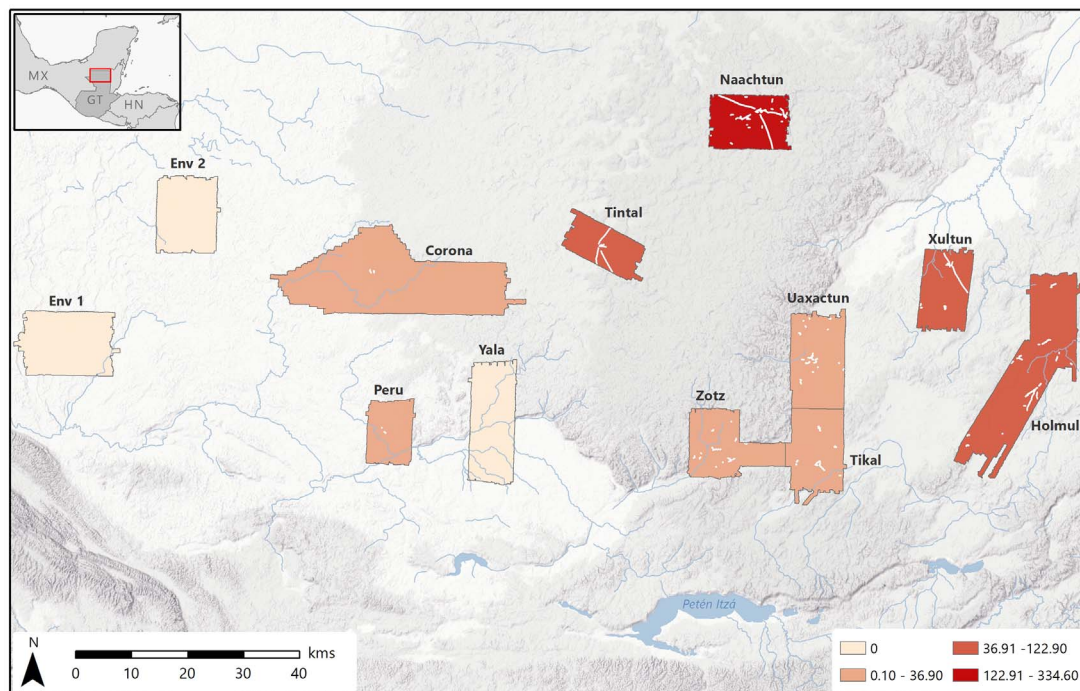
Maya deepened, dammed, or bermed natural depressions to capture rainwater running off stucco-paved surfaces (111). Lidar data now show a greater extent of such large features. Large, round, and bermed reservoirs were built within



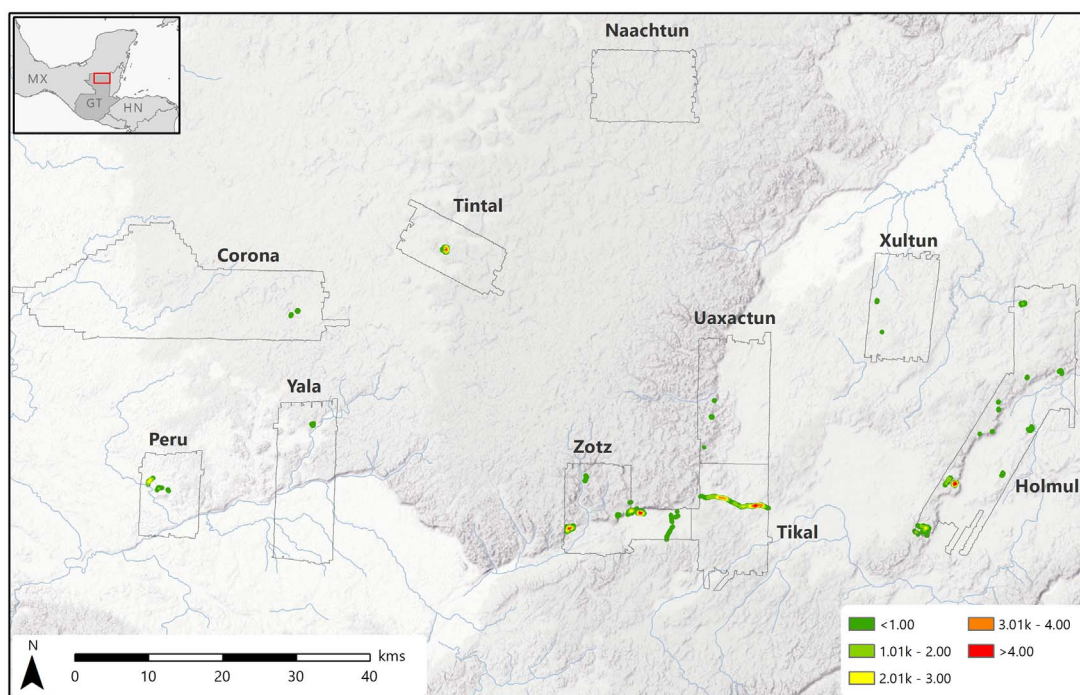
**Table 10. Defensive data for individual defended areas.** Summary of defensive data for individual defended areas and perimeter statistics. Bold lines separate different lidar survey blocks (in order as Yala, Tikal, Uaxactun, Holmul, Zotz, Xultun, Perú, Corona, and Tintal).

Site/region	Defended area	Number of defensive features	Number of defensive systems	Area (ha)	Length of all built defenses (m)	Total perimeter length (m)	Length of built perimeter defenses (m)	Length of natural perimeter defenses (m)	Total length of perimeter defenses (m)	% of perimeter with built defenses	% of perimeter with natural defenses	% of total perimeter defended
E4 River Site	Site core	4	4	7.5	727	1,170	727	443	1,170	62%	38%	100%
Tikal*	City extent	139	10	10,677.4	18,603	15,675	12,072	3,418	15,490	77%	22%	99%
Tikal Polity	Group NW of Tikal Wall	6	2	70.4	1,027	3,670	900	2,763	3,663	25%	75%	100%
Atalaya	Atalaya core	2	1	13.6	198	1,996	198	1,787	1,985	10%	90%	99%
RS028	Refuge	2	1	3.2	132	764	132	631	764	17%	83%	100%
RS07	Eastern hilltop	2	2	1.7	452	539	452	87	539	84%	16%	100%
Cival	Cival core	5	3	28.2	1,080	1,974	1,080	361	1,441	55%	18%	73%
Dos Aguadas	Dos Aguadas	2	2	11.4	1,825	1,637	1,625	13	1,637	99%	1%	100%
	North E plateau											
Dos Aguadas	Dos Aguadas	3	1	23.2	1,341	2,160	528	1,585	2,113	24%	73%	98%
	North N plateau											
Dos Aguadas	Dos Aguadas	1	1	7.8	654	1,363	654	708	1,363	48%	52%	100%
	North W plateau											
Dos Aguadas	Dos Aguadas	14	7	52.1	3,391	3,660	1,650	1,861	3,511	45%	51%	96%
	South core											
Dos Aguadas*	Dos Aguadas	7	4	162.0	2,243	4,349	2,243	804	3,047	52%	18%	70%
	South macro											
Kanalna	Kanalna N refuge	2	1	7.1	287	1,280	116	1,164	1,280	9%	91%	100%
Kanalna	Kanalna S refuge	1	1	16.7	218	2,584	218	2,344	2,563	8%	91%	99%
La Sufricaya	La Sufricaya core	4	3	9.6	556	1,204	556	96	652	46%	8%	54%
Turca†	Turca core	1	1	—	99	—	—	—	—	—	—	—
Turca	Turca E refuge	2	1	2.8	271	695	148	547	695	21%	79%	100%
Witzna	Witzna core	3	1	44.0	547	4,254	200	3,011	3,211	5%	71%	75%
Witzna East	Witzna East	10	4	14.1	942	2,066	630	1,390	2,020	31%	67%	98%
Xmakabuton	Summit E of Polity drainage	6	4	2.9	881	746	412	181	594	55%	24%	80%
Xmakabuton	Summit W of Polity drainage	1	1	5.4	525	924	525	213	738	57%	23%	80%
El Zotz	El Diablo	15	4	21.6	3,347	2,403	1,829	471	2,300	76%	20%	96%
El Zotz	El Tejon	11	3	9.3	1,717	1,399	542	805	1,346	39%	58%	96%
El Zotz	Hill NW of El Diablo	5	1	0.9	332	480	148	332	480	31%	69%	100%
	Group											
El Zotz Polity‡	Drainage N of La Cuernavilla	6	1	10.4	437	1,106	—	—	—	—	—	—
El Zotz Polity	Hills at N drainage intersection	3	2	15.3	771	1,966	771	1,191	1,962	39%	61%	100%
El Zotz Polity	Hills W of La Cuernavilla	7	3	11.5	1,035	2,904	602	2,293	2,895	21%	79%	100%
La Cuernavilla	La Cuernavilla East	27	6	55.0	5,014	4,222	1,278	1,777	3,055	30%	42%	72%
La Cuernavilla	La Cuernavilla West	24	3	21.7	4,262	2,697	1,777	722	2,499	66%	27%	93%
Xultun Polity‡	City extent	21	1	—	298	—	—	—	—	—	—	—
Xultun Polity‡	Unknown	2	1	—	332	—	—	—	—	—	—	—
El Perú-Waka'	El Perú-Waka' core	3	2	164.1	1,100	5,193	1,100	2,151	3,251	21%	41%	63%
El Perú-Waka'	Territorial extent	1	1	—	1,240	—	—	—	—	—	—	—
	Polity‡											
El Achiotal Polity	Defensive enclosure	10	1	1.8	827	522	505	0	505	97%	0%	97%
El Achiotal Polity	Peninsula refuge	3	1	16.7	554	2,193	354	1,792	2,145	16%	82%	98%
Tintal	Tintal core	15	3	37.3	2,703	2,481	1,811	452	2,263	73%	18%	91%

\*The defended area extends beyond the limits of the lidar data, but a substantial portion is visible. All statistics are based on the visible perimeter and area, but may not be reliable metrics of the actual defended area. †There is no clearly discernible defended area for the defensive systems present. Area not included in comparative analyses. ‡The defended area extends beyond the limits of the lidar data and is too incomplete to include in comparative analyses.



**Fig. 9. Map of survey blocks showing causeway density patterns.** Causeways are shown in white; values represent cumulative causeway length (in meters) normalized by survey block area ( $\text{km}^2$ ).



**Fig. 10. Density of defensive features.** Density is expressed as linear meters of defensive features per square kilometer normalized by survey block area.

wetlands to serve the needs of the rural population, and drainage ditches were cut into the edges of reservoirs to control flow during the wettest periods.

Lidar-derived terrain models allow for a volumetric assessment of reservoirs, in measures likely to be conservative because of later infilling. Here, we present data only on the built reservoirs of large scale (i.e., those serving more than a single household) (29). In our sample, Tikal's

urban core had the largest amount of water storage capacity in human-made reservoirs. The Tikal Palace Reservoir alone could store  $31,000 \text{ m}^3$  of water, which would be sufficient to supply the inhabitants of the urban core for an entire year. Earlier cities such as Cival and Tintal were near natural depressions that provided generous amounts of water with little labor investment. For example, Tintal ringed a 2-km-wide sinkhole that could have contained

more than  $3 \text{ million m}^3$  of water in wet years. A 2.5-km canal carried overflow into a natural drainage, preventing rising waters from flooding the city.

In all cases, such features are monumental in scale and imply some form of centralized involvement in planning and execution. Notably, large-scale water infrastructure is not limited to city centers but occurs in the periurban and even rural zones surrounding major cities, and

in some cases integrates cities and their peripheries. These findings indicate that the construction of monumental reservoirs and, less commonly, canals in both city centers and rural areas involved some form of centralized, institutional coordination. Yet this observation does not negate the possibility that smaller-scale infrastructure was independently built by households and corporate groups (104). Rather, the PLI data suggest that water management was neither fully centralized (112) nor left completely to individual households.

Lidar data also reveal investment in regional and local interconnectivity. “Causeways,” elevated and paved roads of varied size, demonstrate economic and political integration by revealing formal connections between cities, smaller communities, and dispersed populations (113). Until recently, knowledge of formal political connections was restricted to epigraphic records of dynastic interaction and a small number of causeways (25, 26, 114–117). The lidar sample

identified ~106 linear kilometers of causeway construction dating to the Late Preclassic and Classic periods. Intersite causeways are primarily associated with cities that rose to prominence during the Preclassic period (Tintal, Cival, San Bartolo) or Early Classic (Naachtun) and are surprisingly absent from the Late Classic period. They can reach up to 22 km in length (Tintal-Mirador) and 10 to 20 m in width, linking cities to smaller centers nearby. Examples of these occur at Tintal, where multiple monumental causeways radiate toward other centers. Intra-site causeways are associated with Late Classic period centers of all sizes. These roads are typically only a few hundred meters long, providing grand entryways to public spaces within a community. Tikal has the widest causeways (80 m) and the largest amount of internal paved causeway area (0.19 km<sup>2</sup>). Overall, causeways are more widely distributed than had been previously appreciated, especially in the Late Classic period. Across the sample, there is a general trend of increased causeway den-

sity from west to east and from south to north (Fig. 9), coinciding with regional gradients in settlement density, agricultural intensification, and water management infrastructure.

Perhaps the most obvious example of infrastructural investment comes in the form of defensive or military features pointing to a high incidence of conflict in the Maya Lowlands. Such fortifications are found at a scale and quantity (Fig. 10) matched only by the Tikal earthworks and the monumental fortifications at Becan, both known since the 1960s (118–120). To a notable extent, settlement density does not correlate with regional defense, in that some of the most densely settled blocks, such as Naachtun, have no defensive features. Individual defensive features—bridges, ditches, ramparts, stone walls, and terraces—were constructed as components of “built defensive systems.” These combined with natural defenses to protect “defended areas.” There were five types of built defensive systems: landscape ditch-and-rampart (type 1), hilltop ditch-and-rampart (type 2),

**Table 11. Statistics for reinforcing defenses.** Bold lines define natural breaks (Jenks) in the percentages of additional defenses in relation to an area's perimeter.

Site/region	Defended area	Length of additional defenses (m)	% of additional defenses from total built	% of additional defenses from total perimeter
La Cuernavilla	La Cuernavilla East	3736	75%	122%
La Cuernavilla	La Cuernavilla West	2485	58%	99%
El Zotz	El Tejon	1176	68%	87%
Tintal	Tintal core	1811	67%	80%
Xmakabatun Polity	Summit E of drainage	469	53%	79%
El Zotz	El Diablo	1519	45%	66%
El Achiotal Polity	Defensive enclosure	321	39%	64%
Dos Aguadas	Dos Aguadas South core	1742	51%	50%
Tikal*	City extent	6531	35%	42%
Dos Aguadas	Dos Aguadas North N plateau	813	61%	38%
El Zotz	Hill NW of El Diablo Group	184	55%	38%
Turca	Turca E refuge	123	45%	18%
Witzna East	Witzna East	312	33%	15%
El Zotz Polity	Hills W of La Cuernavilla	433	42%	15%
Kanalna	Kanalna N refuge	171	60%	13%
Dos Aguadas	Dos Aguadas North E plateau	200	11%	12%
Witzna	Witzna core	347	63%	11%
El Achiotal Polity	Peninsula refuge	200	36%	9%
Tikal Polity	Group NW of Tikal Wall	127	12%	3%
RS028	Refuge	0	0%	0%
Dos Aguadas	Dos Aguadas North W plateau	0	0%	0%
RS07	Eastern hilltop	0	0%	0%
E4 River Site	Site core	0	0%	0%
El Zotz Polity	Hills at N drainage intersection	0	0%	0%
Atalaya	Atalaya core	0	0%	0%
Kanalna	Kanalna S refuge	0	0%	0%
Xmakabatun Polity	Summit W of drainage	0	0%	0%
Cival	Cival core	0	0%	0%
Dos Aguadas*	Dos Aguadas South macro	0	0%	0%
El Perú-Waka'	El Perú-Waka' core	0	0%	0%
La Sufricaya	La Sufricaya core	0	0%	0%

\*The defended area extends beyond the limits of the lidar data, but a substantial portion is visible.



contoured terrace (type 3), stand-alone rampart (type 4), and stone wall (type 5). These were unevenly distributed across the sample, reflecting variation in local geography, resources, and history.

Built defensive systems are found in 36 discrete locations across the sample, with 31 meeting the criteria for defended areas (Table 10). In most cases (61.3%), defenses consisted of more than the establishment of perimeter walls. One-third ( $n = 10$ ) of all defended areas displayed more defensive features within the perimeter than on the perimeter itself (Table 11). The Holmul and Zotz survey blocks exhibited the greatest effort in reinforcing defenses, suggesting a heavy investment in the protection of particular sites in those areas. We also identified “refuges,” highly defensible areas lacking any other perceptible architecture. Presumably, these were only used as a last resort and for brief stays. Clear examples appear at Kanalna and the El Achiotal promontory.

Several metrics were calculated to interpret the defensive data (29). We assessed the “perceived need” for defense of an area according to the percentage of the perimeter defended by built or natural defenses (Table 10), as well as how much defensibility factored into the siting of a settlement. The latter statistic is a preparedness index that factors out the natural defensibility of a location (Table 12). At some larger sites, such as Tikal and Tintal, substantial investment in building perimeter defenses around the urban core suggests that the need for such constructions emerged well after the initial establishment of the settlement. Finally, we assessed the cost of building defensive systems in relation to the total protected area (Table 13). The defensive features of greatest cost are located in the southern part of the sample, particularly the Early Classic Buenavista Valley linking El Zotz and Tikal.

At the regional level, the Zotz, Tikal, Holmul, and Perú blocks show the greatest concentration of defensive features (Table 14). However, both the Zotz and Tikal regions demonstrate the largest investment of built and combined defenses in the sample. By controlling for survey block size, a line density analysis (Fig. 10) highlights the higher investment in defense along a west-to-east route. The trend continues to the east with several fortified hilltops in the Holmul region. The greatest density of high-cost features extends along a corridor associated with a Central Mexican (Teotihuacan) military intrusion into the Maya Lowlands in the late fourth century CE (121).

Concluding remarks

Since its first application in Maya archaeology (122), lidar technology has enabled the evaluation of lowland Maya society on a large scale. The PLI results confirm that lidar technology represents a watershed event in archaeological survey of forested environments. In the central Maya Lowlands, lidar survey will become indispensable for settlement research because of

(i) the speed of the data-gathering process, (ii) the degree of detail attainable over large areas, and (iii) the ability to discern large and small features routinely undetected by traditional methods.

The PLI lidar survey provides a uniquely large and continuous dataset for the central Maya Lowlands that is replete with evidence for ancient structures, canals, terraces, causeways, and defensive features. These data also show great variability in the intensity and distribution of these features over space. In sum, the PLI data unambiguously support the notion that the lowland Maya constructed a variable and contentious landscape in which a regionally interconnected network of densely populated and defended cities was sustained by an array of agricultural practices that optimized land productivity, resource diversity, and sustainability on a much grander scale than previously thought.

These preliminary conclusions should encourage research that, among many possibilities, (i) quantifies labor investments and land productivity at every scale; (ii) addresses how proj-

ects of agricultural intensification were organized and managed, with implications for long-term sustainability; (iii) assesses the extent of economic interdependence within large regions by studying differences between urban and rural populations; (iv) determines the relationship among settlement densities, land use, and monumental infrastructure; and (v) develops models for the types of warfare implied by defensive features.

Finally, by detecting most looters’ pits and assembling full inventories of ruins, lidar data offer a profound tool for cultural heritage management. Quantification of biomass and monitoring of biodiversity—an underexplored product of these lidar data—will also facilitate planning and management of regional conservation efforts and deepen our understanding of the relationship between ancient and modern environments.

Methods

The Pacunam Lidar Initiative collected lidar data over an area of 2144 km<sup>2</sup> of the Maya Biosphere

**Table 12. Degree of preparedness for each defended area.** The degree to which the need for defense was anticipated at the time of initial settlement. Scores closer to 0 indicate minimal preparedness.

Site/region	Defended area	Preparedness index
El Zotz	Hill NW of El Diablo Group	369
RS028	Refuge	199
El Zotz Polity	Hills W of La Cuernavilla	199
Turca	Turca E refuge	194
Kanalna	Kanalna N refuge	164
Kanalna	Kanalna S refuge	140
Atalaya	Atalaya core	132
El Achiotal Polity	Peninsula refuge	107
Witzna East	Witzna East	99
Dos Aguadas	Dos Aguadas North W plateau	91
El Zotz	El Tejon	87
El Zotz Polity	Hills at N drainage intersection	78
Witzna	Witzna core	68
Dos Aguadas	Dos Aguadas North N plateau	68
Xmakabatun Polity	Summit E of drainage	62
E4 River Site	Site core	59
RS07	Eastern hilltop	50
Xmakabatun Polity	Summit W of drainage	39
Tikal Polity	Group NW of Tikal Wall	39
Dos Aguadas	Dos Aguadas South core	36
La Cuernavilla	La Cuernavilla West	33
La Cuernavilla	La Cuernavilla East	32
El Zotz	El Diablo	22
El Perú-Waka'	El Perú-Waka' core	13
Cival	Cival core	13
Tintal	Tintal core	12
La Sufricaya	La Sufricaya core	10
Dos Aguadas*	Dos Aguadas South macro	5
Dos Aguadas	Dos Aguadas North E plateau	1
Tikal*	City extent	0
El Achiotal Polity	Defensive enclosure	0

\*The defended area extends beyond the limits of the lidar data, but a substantial portion is visible.

**Table 13. Cost index for each defended area.** Cost statistics for defending areas in relation to idealized forms.

Site/region	Defended area	Total built defenses (m)	z-score of TBD	Overbuild score	Efficiency score	Cost index
La Cuernavilla	La Cuernavilla West	4,262	0.73	2.58	0.61	3.08
La Cuernavilla	La Cuernavilla East	5,014	0.96	1.91	0.62	2.94
El Zotz	El Diablo	3,348	0.45	2.03	0.69	1.34
Tikal*	City extent	18,603	5.09	0.51	2.34	1.11
Dos Aguadas	Dos Aguadas South core	3,391	0.47	1.33	0.70	0.88
Tintal	Tintal core	2,703	0.26	1.25	0.87	0.37
Dos Aguadas*	Dos Aguadas South macro	2,243	0.12	0.50	1.04	0.06
Dos Aguadas	Dos Aguadas North E plateau	1,825	-0.01	1.52	0.73	-0.02
El Perú-Waka'	El Perú-Waka' core	1,100	-0.23	0.24	0.87	-0.06
El Zotz	El Tejon	1,717	-0.04	1.59	0.77	-0.09
Tikal Polity	Group NW of Tikal Wall	1,027	-0.25	0.35	0.81	-0.11
Atalaya	Atalaya core	198	-0.50	0.15	0.65	-0.12
RS028	Refuge	132	-0.52	0.21	0.83	-0.13
Kanalna	Kanalna S refuge	218	-0.50	0.15	0.56	-0.13
Cival	Cival Core	1,080	-0.24	0.57	0.95	-0.14
Dos Aguadas	Dos Aguadas North N plateau	1,341	-0.16	0.78	0.79	-0.16
Witzna	Witzna core	547	-0.40	0.23	0.55	-0.17
Kanalna	Kanalna N refuge	287	-0.48	0.30	0.74	-0.20
La Sufricaya	La Sufricaya core	556	-0.40	0.51	0.91	-0.22
El Achiotal Polity	Peninsula refuge	554	-0.40	0.38	0.66	-0.23
Turca	Turca E refuge	271	-0.48	0.45	0.86	-0.26
El Zotz Polity	Hills at N drainage intersection	771	-0.33	0.56	0.71	-0.26
Xmakabatun Polity	Summit W of drainage	525	-0.40	0.64	0.89	-0.29
Witzna East	Witzna East	942	-0.28	0.71	0.64	-0.31
E4 River Site	Site core	727	-0.34	0.75	0.83	-0.31
Dos Aguadas	Dos Aguadas North W plateau	654	-0.37	0.66	0.73	-0.33
RS07	Eastern hilltop	452	-0.43	0.97	0.86	-0.48
El Zotz Polity	Hills W of La Cuernavilla	1,035	-0.25	0.86	0.41	-0.52
Xmakabatun Polity	Summit E of drainage	881	-0.30	1.46	0.81	-0.53
El Achiotal Polity	Defensive enclosure	827	-0.31	1.73	0.91	-0.59
El Zotz	Hill NW of El Diablo Group	332	-0.46	0.99	0.70	-0.65

\*The defended area extends beyond limits of lidar data, but a substantial portion is visible.

**Table 14. Regional defensive statistics by lidar survey block.** TLD, total length defended; TBD, total built defenses.

Polygon	Major site	Lidar area (km <sup>2</sup> )	Number of defensive systems	km <sup>2</sup> of lidar per defensive system	Number of defensive features	km <sup>2</sup> of lidar per defensive feature	Total length of defended perimeters (m)	Length of defended perimeter per km <sup>2</sup> of lidar	Defended area (km <sup>2</sup> )	% lidar area defended	TLD (m)	TLD per km <sup>2</sup> of lidar	TBD length (m)	TBD per km <sup>2</sup> of lidar
E1		184.14	0	0.0	0	0.0	0	0	0.0	0.0%	0	0.0	0	0.0
E2		145.59	0	0.0	0	0.0	0	0	0.0	0.0%	0	0.0	0	0.0
E4		172.02	4	43.0	4	43.0	1,170	7	0.1	0.0%	1,170	6.8	727	4.2
1	Tikal	161.33	12	13.4	145	1.1	19,345	120	107.5	66.6%	25,811	160.0	19,630	121.7
2	Uaxactun	164.94	4	41.2	6	27.5	3,298	20	0.2	0.1%	3,287	19.9	782	4.7
8	Naachtun	135.26	0	0.0	0	0.0	0	0	0.0	0.0%	0	0.0	0	0.0
9	Holmul	308.66	35	8.8	62	5.0	28,895	94	3.9	1.3%	29,139	94.4	14,862	48.1
10	El Zotz	127.20	23	5.5	98	1.3	17,922	141	1.5	1.1%	24,507	192.7	16,915	133.0
11	Xultun	124.02	2	62.0	23	5.4	0	0	0.0	0.0%	630	5.1	630	5.1
12	El Perú	90.96	3	30.3	4	22.7	5,193	57	1.6	1.8%	4,491	49.4	2,340	25.7
13	La Corona	424.69	2	212.3	13	32.7	2,715	6	0.2	0.0%	3,173	7.5	1,381	3.3
15	Tintal	97.03	3	32.3	15	6.5	2,481	26	0.4	0.4%	3,155	32.5	2,703	27.9
Total		2,135.84	88	24.3	370	5.8	81,019	38	115.3	5.4%	95,363	44.6	59,971	28.1
Average		177.99	7.33	37.4	30.83	12.1	6,752	39	9.6	6.0%	7,947	47.4	4,998	31.1

Reserve (MBR) in northern Guatemala. These data were provided to the research consortium as bare-earth terrain models with grid resolution of 1 m. Given the coverage scale and resolution, these raster models make it possible to interpret aspects of ancient Maya society as warranted by visible remains of archaeological settlement. These data were analyzed to (i) estimate regional population levels, (ii) calculate agricultural capacity, (iii) determine settlement density patterns, and (iv) evaluate the extent of infrastructural investment in water storage, regional communication, and defense.

Terrain data and its derivatives were mostly manipulated and analyzed through GIS applications (ArcMap 10.4 and 10.6, QGIS 2.18, and GRASS 7.2, and the Relief Visualization Toolkit L3), enhanced by supplemental calculations and statistical analyses performed in Microsoft Excel and SPSS. Terrain models were visualized using multiple methods. Archaeological features were digitized manually, using existing site maps as training samples, augmented by targeted ground-truthing during the 2017 field season. All PLI member projects uploaded their manually derived feature datasets to a shared cloud server, where they were collated to create a single regional dataset for each feature class (e.g., buildings, causeways, terraces, etc.).

Population estimates are based on the total number of structures in the sample and include adjustments for Late Classic occupation, non-residential structures, structures not visible on the surface, contemporaneity of occupation within the Late Classic period, and number of persons per structure. Settlement density classes were calculated using the Heatmap tool in QGIS 2.18 and classified according to published descriptions of settlement density in the cores and peripheries of numerous sites in the central Lowlands. Agricultural capacity was estimated by combining zones of intensive cultivation (defined on the basis of built agricultural features) with sparsely settled and unmodified uplands to define the total area available for agriculture; this area was then multiplied by a standardized estimate of traditional Maya swidden productivity. Water storage capacity was calculated using the *r.lake* function in GRASS 7.2. Causeways were classified using published typologies and statistics calculated by “calculate geometries” in ArcMap 10.4. Defensive features were classified by an original typology, their distribution was analyzed using the Line Density tool in ArcMap 10.6, and inferences regarding their cost effectiveness and other metrics were derived from statistical analyses performed in Microsoft Excel.

See (29) for detailed methodology and analytical procedures for each of these analyses.

## REFERENCES AND NOTES

- A. H. Siemens, in *Pre-Hispanic Maya Agriculture*, P. D. Harrison, B. L. Turner II, Eds. (Univ. of New Mexico Press, 1978), pp. 117–143.
- T. Inomata et al., Archaeological application of airborne lidar with object-based vegetation classification and visualization techniques at the lowland Maya site of Ceibal, Guatemala. *Remote Sens.* **9**, 563 (2017). doi: [10.3390/rs9060563](https://doi.org/10.3390/rs9060563)
- A. F. Chase et al., Airborne lidar, archaeology, and the ancient Maya landscape at Caracol, Belize. *J. Archaeol. Sci.* **38**, 387–398 (2011). doi: [10.1016/j.jas.2010.09.018](https://doi.org/10.1016/j.jas.2010.09.018)
- K. Reese-Taylor et al., Boots on the ground at Yaxnohcah: Ground-truthing LiDAR in a complex tropical landscape. *Adv. Archaeol. Pract.* **4**, 314–338 (2016). doi: [10.7183/2326-3768.4.3.314](https://doi.org/10.7183/2326-3768.4.3.314)
- A. F. Chase et al., Ancient Maya regional settlement and inter-site analysis: The 2013 west-central Belize lidar survey. *Remote Sens.* **6**, 8671–8695 (2014). doi: [10.3390/rs6098671](https://doi.org/10.3390/rs6098671)
- A. F. Chase et al., The Use of LiDAR in Understanding the Ancient Maya Landscape: Caracol and Western Belize. *Adv. Archaeol. Pract.* **2**, 147–160 (2014). doi: [10.7183/2326-3768.4.3.314](https://doi.org/10.7183/2326-3768.4.3.314)
- J. Yaeger, M. K. Brown, B. Cap, Locating and dating sites using lidar survey in a mosaic landscape in western Belize. *Adv. Archaeol. Pract.* **4**, 339–356 (2016). doi: [10.7183/2326-3768.4.3.339](https://doi.org/10.7183/2326-3768.4.3.339)
- A. Ford, Using cutting-edge lidar technology at El Pilar Belize-Guatemala in discovering ancient Maya sites—there is still a need for archaeologists. *Res. Rep. Belizean Archaeol.* **12**, 271–280 (2014).
- B. L. Turner 2nd, J. A. Sabloff, Classic Period collapse of the Central Maya Lowlands: Insights about human-environment relationships for sustainability. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13908–13914 (2012). doi: [10.1073/pnas.1210106109](https://doi.org/10.1073/pnas.1210106109); pmid: [2292403](https://pubmed.ncbi.nlm.nih.gov/2292403/)
- N. Grube, in *A Comparative Study of Thirty City-State Cultures*, H. M. Hansen, Ed. (Royal Danish Academy of Sciences and Letters, 2000), pp. 547–565.
- D. L. Webster, in *The Archaeology of City-States: Cross-Cultural Approaches*, D. L. Nichols, T. H. Charlton, Eds. (Smithsonian Institution Press, 1997), pp. 135–154.
- D. A. Freidel, in *Peer-Polity Interaction and Socio-political Change*, C. Renfrew, J. F. Cherry, Eds. (Cambridge Univ. Press, 1986), pp. 93–108.
- T. P. Culbert, *Ancient Maya Wetland Agriculture* (Foundation for the Advancement of Mesoamerican Studies, 2005); [www.famsi.org/reports/94033/index.html](http://www.famsi.org/reports/94033/index.html)
- S. L. Fedick, in *The Managed Mosaic: Ancient Maya Agriculture and Resource Use*, S. L. Fedick, Ed. (Univ. of Utah Press, 1996), pp. 107–131.
- N. P. Dunning et al., Arising from the bajos: The evolution of a neotropical landscape and the rise of Maya civilization. *Ann. Assoc. Am. Geogr.* **92**, 267–283 (2002). doi: [10.1111/1467-8306.00290](https://doi.org/10.1111/1467-8306.00290)
- P. D. Harrison, in *Social Process in Maya Prehistory: Studies in Honour of Sir Eric Thompson*, N. Hammond, Ed. (Academic Press, 1977), pp. 469–508.
- J. L. Kunen, P. T. Culbert, V. Fialko, B. M. McKee, L. Grazioso, Bajo communities: A case study from the central Peten. *Cult. Agric.* **22**, 15–31 (2000). doi: [10.1525/cag.2000.22.3.15](https://doi.org/10.1525/cag.2000.22.3.15)
- M. D. Pohl, Ed., *Ancient Maya Wetland Agriculture: Excavations on Albion Island, Northern Belize* (Westview, 1990).
- N. P. Dunning, T. Beach, Soil erosion, slope management, and ancient terracing in the Maya Lowlands. *Lat. Am. Antiq.* **5**, 51–69 (1994). doi: [10.2307/971902](https://doi.org/10.2307/971902)
- A. F. Chase, D. Z. Chase, Scale and intensity in Classic period Maya agriculture: Terracing and settlement at the ‘garden city’ of Caracol, Belize. *Cult. Agric.* **20**, 60–77 (1998). doi: [10.1525/cag.1998.20.2.3.60](https://doi.org/10.1525/cag.1998.20.2.3.60)
- N. P. Dunning, T. Beach, D. Rue, The paleoecology and ancient settlement of the Petexbatun region, Guatemala. *Anc. Mesoam.* **8**, 255–266 (1997). doi: [10.1017/S0956536100001711](https://doi.org/10.1017/S0956536100001711)
- D. Z. Chase, A. F. Chase, W. A. Haviland, The Classic Maya city: Reconsidering the “Mesoamerican urban tradition”. *Am. Anthropol.* **92**, 499–506 (1990). doi: [10.1525/aa.1990.92.2.02a00210](https://doi.org/10.1525/aa.1990.92.2.02a00210)
- A. F. Chase, D. Z. Chase, Ancient Maya causeways and site organization at Caracol, Belize. *Anc. Mesoam.* **12**, 273–281 (2001). doi: [10.1017/S09565361010121097](https://doi.org/10.1017/S09565361010121097)
- W. J. Folan, Calakmul, Campeche: A centralized urban administrative center in the northern Peten. *World Archaeol.* **24**, 158–168 (1992). doi: [10.1080/00438243.1992.9980199](https://doi.org/10.1080/00438243.1992.9980199)
- S. Martin, N. Grube, *Chronicle of the Maya Kings and Queens* (Thames & Hudson, ed. 2, 2008).
- D. Z. Chase, A. F. Chase, Caracol, Belize, and changing perceptions of ancient Maya society. *J. Archaeol. Res.* **25**, 185–249 (2017). doi: [10.1007/s10814-016-9101-z](https://doi.org/10.1007/s10814-016-9101-z)
- B. L. Turner II, in *Culture, Form, and Place: Essays in Cultural and Historical Geography*, K. W. Mathewson, Ed. (Louisiana State University, 1993), pp. 57–88.
- J. C. Fernandez-Diaz et al., Capability assessment and performance metrics for the Titan multispectral mapping lidar. *Remote Sens.* **8**, 936 (2016). doi: [10.3390/rs8110936](https://doi.org/10.3390/rs8110936)
- See supplementary materials.
- K. Zakšek, K. Ostir, Z. Kokalj, Sky-View Factor as a relief visualization technique. *Remote Sens.* **3**, 398–415 (2011). doi: [10.3390/rs3020398](https://doi.org/10.3390/rs3020398)
- T. Chiba, S.-i. Kaneta, Y. Suzuki, Red relief image map: New visualization method for three dimensional data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **37**, 1071–1076 (2008).
- Ž. Kokalj, R. Hesse, *Airborne Laser Scanning Raster Data Visualization: A Guide to Good Practice* (Založba ZRC, Ljubljana, 2017).
- M. Doneus, Openness as visualization technique for interpretative mapping of airborne lidar derived digital terrain models. *Remote Sens.* **5**, 6427–6442 (2013). doi: [10.3390/rs5126427](https://doi.org/10.3390/rs5126427)
- D. S. Rice, in *Urbanism in the Preindustrial World: Cross-Cultural Approaches*, G. R. Storey, Ed. (Univ. of Alabama Press, 2006), pp. 252–276.
- D. S. Rice, T. P. Culbert, in *Pre-Columbian Population History in the Maya Lowlands*, T. P. Culbert, D. S. Rice, Eds. (Univ. of New Mexico Press, 1990), pp. 1–36.
- B. L. Turner II, in *Pre-Columbian Population History in the Maya Lowlands*, T. P. Culbert, D. S. Rice, Eds. (Univ. of New Mexico Press, 1990), pp. 301–324.
- A. Ford, R. Nigh, *The Maya Forest Garden: Eight Millennia of Sustainable Cultivation of the Tropical Woodlands* (Routledge, 2015).
- T. Inomata et al., Archaeological application of airborne LiDAR to examine social changes in the Ceibal region of the Maya lowlands. *PLOS ONE* **13**, e0191619 (2018). doi: [10.1371/journal.pone.0191619](https://doi.org/10.1371/journal.pone.0191619); pmid: [29466384](https://pubmed.ncbi.nlm.nih.gov/29466384/)
- R. E. W. Adams, in *Lowland Maya Settlement Patterns*, W. A. Ashmore, Ed. (Univ. of New Mexico Press, 1981), pp. 211–257.
- T. P. Culbert, *The New Maya*. *Archaeology* **51**, 48–51 (1998).
- R. J. Sharer, L. P. Traxler, *The Ancient Maya* (Stanford Univ. Press, 2006).
- S. G. Ortman, A. H. F. Cabaniss, J. O. Sturm, L. M. A. Bettencourt, The pre-history of urban scaling. *PLOS ONE* **9**, e87902 (2014). doi: [10.1371/journal.pone.0087902](https://doi.org/10.1371/journal.pone.0087902); pmid: [24533062](https://pubmed.ncbi.nlm.nih.gov/24533062/)
- P. D. Harrison, B. L. Turner II, Eds., *Pre-Hispanic Maya Agriculture* (Univ. of New Mexico Press, 1978).
- J. L. Kunen, Ancient Maya agricultural installations and the development of intensive agriculture in N.W. Belize. *J. Field Archaeol.* **28**, 325–346 (2001). doi: [10.1179/jfa.2001.28.3.4.325](https://doi.org/10.1179/jfa.2001.28.3.4.325)
- D. L. Lentz et al., Forests, fields, and the edge of sustainability at the ancient Maya city of Tikal. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18513–18518 (2014). doi: [10.1073/pnas.1408631111](https://doi.org/10.1073/pnas.1408631111); pmid: [25512500](https://pubmed.ncbi.nlm.nih.gov/25512500/)
- N. P. Dunning et al., in *Tikal: Paleoecology of an Ancient Maya City*, D. L. Lentz, V. L. Scarborough, N. P. Dunning, Eds. (Cambridge Univ. Press, 2015), pp. 95–123.
- R. E. W. Adams, Swamps, canals, and the location of ancient Maya cities. *Antiquity* **54**, 206–214 (1980). doi: [10.1017/S0003598X00043386](https://doi.org/10.1017/S0003598X00043386)
- R. E. W. Adams, W. E. Brown Jr., T. P. Culbert, Radar mapping, archaeology, and ancient Maya land use. *Science* **213**, 1457–1468 (1981). doi: [10.1126/science.213.4515.1457](https://doi.org/10.1126/science.213.4515.1457); pmid: [17780866](https://pubmed.ncbi.nlm.nih.gov/17780866/)
- R. E. W. Adams, T. P. Culbert, W. E. Brown Jr., P. D. Harrison, L. J. Levy, Rebuttal to Pope and Dahlin. *J. Field Archaeol.* **17**, 241–243 (1990).
- K. O. Pope, B. H. Dahlin, Ancient Maya wetland agriculture: New insights from ecology and remote sensing research. *J. Field Archaeol.* **16**, 87–106 (1989).
- N. P. Dunning, in *The Managed Mosaic: Ancient Maya Agriculture and Resource Use*, S. L. Fedick, Ed. (Univ. of Utah Press, 1996), pp. 53–68.
- K. O. Pope, B. H. Dahlin, Radar detection and ecology of ancient Maya canal systems—Reply to Adams et al. *J. Field Archaeol.* **20**, 379–383 (1993). doi: [10.2307/530065](https://doi.org/10.2307/530065)
- B. L. Turner II, in *Pre-Hispanic Maya Agriculture*, P. D. Harrison, B. L. Turner II, Eds. (Univ. of New Mexico Press, 1978), pp. 163–183.
- P. D. Harrison, in *Vision and Revision in Maya Studies*, F. S. Clancy, P. D. Harrison, Eds. (Univ. of New Mexico Press, 1990), pp. 99–113.
- T. P. Culbert, L. Levi, L. Cruz, in *Vision and Revision in Maya Studies*, F. S. Clancy, P. D. Harrison, Eds. (Univ. of New Mexico Press, 1990), pp. 115–124.



56. T. P. Culbert, L. J. Levi, B. M. McKee, J. L. Kunen, in *IX Simposio de Investigaciones Arqueológicas en Guatemala*, J. P. Laporte, H. L. Escobedo, Eds. (Instituto Nacional de Antropología e Historia, Guatemala, 1996), pp. 51–57.
57. B. H. Dahlin, A. Dahlin, Platforms in the *akalche* at El Mirador, Peten, Guatemala and their implications. *Geoarchaeology* **9**, 203–237 (1994). doi: [10.1002/gea.3340090303](https://doi.org/10.1002/gea.3340090303)
58. B. H. Dahlin, in *Actes du XLII<sup>e</sup> Congrès International des Américanistes*, vol. 8, pp. 305–312 (1979).
59. A. H. Siemens, D. E. Puleston, Ridged fields and associated features in southern Campeche: New perspectives on the Lowland Maya. *Am. Antiq.* **37**, 228–239 (1972). doi: [10.2307/278209](https://doi.org/10.2307/278209)
60. D. E. Puleston, in *Social Process in Maya Prehistory: Studies in Honour of Sir Eric Thompson*, N. Hammond, Ed. (Academic Press, 1977), pp. 449–467.
61. B. L. Turner II, Prehistoric intensive agriculture in the Mayan lowlands. *Science* **185**, 118–124 (1974). doi: [10.1126/science.185.4146.118](https://doi.org/10.1126/science.185.4146.118); pmid: [17810497](https://pubmed.ncbi.nlm.nih.gov/17810497/)
62. B. L. Turner II, P. D. Harrison, Prehistoric raised-field agriculture in the Maya lowlands. *Science* **213**, 399–405 (1981). doi: [10.1126/science.213.4506.399](https://doi.org/10.1126/science.213.4506.399); pmid: [17760172](https://pubmed.ncbi.nlm.nih.gov/17760172/)
63. B. L. Turner II, P. D. Harrison, Eds., *Pulltrouser Swamp: Ancient Maya Habitat, Agriculture, and Settlement in Northern Belize* (Univ. of Texas Press, 1983).
64. E. Lemonnier, B. Vannière, Agrarian features, farmsteads, and homesteads in the Rio Bec nuclear zone, Mexico. *Anc. Mesoam.* **24**, 397–413 (2013). doi: [10.1017/S0956536113000242](https://doi.org/10.1017/S0956536113000242)
65. P. D. Harrison, in *The Managed Mosaic: Ancient Maya Agriculture and Resource Use*, S. L. Fedick, Ed. (Univ. of Utah Press, 1996), pp. 177–190.
66. M. D. Pohl et al., Early agriculture in the Maya Lowlands. *Lat. Am. Antiq.* **7**, 355–372 (1996). doi: [10.2307/972264](https://doi.org/10.2307/972264)
67. J. S. Jacob, Ancient Maya wetland agricultural fields in Cobweb Swamp, Belize: Construction, chronology, and function. *J. Field Archaeol.* **22**, 175–190 (1995).
68. M. D. Pohl, in *Ancient Maya Wetland Agriculture*, M. D. Pohl, Ed. (Waveland, Boulder, CO, 1990), pp. 1–20.
69. S. L. Fedick, A. Ford, The prehistoric agricultural landscape of the central Maya Lowlands: An examination of local variability in a regional context. *World Archaeol.* **22**, 18–33 (1990). doi: [10.1080/00438243.1990.9980126](https://doi.org/10.1080/00438243.1990.9980126)
70. P. D. Harrison, in *Pre-Hispanic Maya Agriculture*, P. D. Harrison, B. L. Turner II, Eds. (Univ. of New Mexico Press, 1978), pp. 247–253.
71. N. P. Dunning et al., Temple mountains, sacred lakes, and fertile fields: Ancient Maya landscapes in northwestern Belize. *Antiquity* **73**, 650–660 (1999). doi: [10.1017/S0003598X0006525X](https://doi.org/10.1017/S0003598X0006525X)
72. S. Kepecs, S. Boucher, in *The Man-aged Mosaic: Ancient Maya Agriculture and Resource Use*, S. L. Fedick, Ed. (Univ. of Utah Press, 1996), pp. 69–91.
73. K. J. Johnston, The intensification of pre-industrial cereal agriculture in the tropics: Boserup, cultivation lengthening, and the Classic Maya. *J. Anthropol. Archaeol.* **22**, 126–161 (2003). doi: [10.1016/S0278-4165\(03\)00013-8](https://doi.org/10.1016/S0278-4165(03)00013-8)
74. A. S. Z. Chase, J. Weishampel, Using LiDAR and GIS to investigate water and soil management in the agricultural terracing at Caracol, Belize. *Adv. Archaeol. Pract.* **4**, 357–370 (2016). doi: [10.7183/2326-3768.4.3.357](https://doi.org/10.7183/2326-3768.4.3.357)
75. U. M. Cowgill, Soil fertility, population, and the ancient Maya. *Proc. Natl. Acad. Sci. U.S.A.* **46**, 1009–1011 (1960). doi: [10.1073/pnas.46.8.1009](https://doi.org/10.1073/pnas.46.8.1009); pmid: [16590706](https://pubmed.ncbi.nlm.nih.gov/16590706/)
76. U. M. Cowgill, Soil fertility and the ancient Maya. *Trans. Conn. Acad. Arts Sci.* **42**, 1–56 (1961).
77. U. M. Cowgill, An agricultural study of the southern Maya Lowlands. *Am. Anthropol.* **64**, 273–286 (1962). doi: [10.1525/aa.1962.64.2.02a00030](https://doi.org/10.1525/aa.1962.64.2.02a00030)
78. R. E. Griffin, thesis, Pennsylvania State University (2012).
79. N. B. Schwartz, A note on 'weights, measures' and swidden. *Cult. Agric.* **27**, 9–12 (1985).
80. N. B. Schwartz, A. R. Corzo M, Swidden counts: A Petén, Guatemala, milpa system: production, carrying capacity, and sustainability in the southern Maya Lowlands. *J. Anthropol. Res.* **71**, 69–93 (2015). doi: [10.3998/jar.0521004.0071.104](https://doi.org/10.3998/jar.0521004.0071.104)
81. J. D. Nations, R. B. Nigh, The evolutionary potential of Lacandon Maya sustained-yield tropical forest agriculture. *J. Anthropol. Res.* **36**, 1–30 (1980). doi: [10.1086/jar.36.1.3629550](https://doi.org/10.1086/jar.36.1.3629550)
82. A. Ford, K. C. Clarke, S. Morlet, Calculating Late Classic Lowland Maya population for the upper Belize River area. *Res. Rep. Belizean Archaeol.* **8**, 75–87 (2011).
83. A. Ford, R. Nigh, Origins of the Maya forest garden: Maya resource management. *J. Ethnobiol.* **29**, 213–236 (2009). doi: [10.2993/0278-0771-29.2.213](https://doi.org/10.2993/0278-0771-29.2.213)
84. W. A. Haviland, Tikal, Guatemala, and Mesoamerican urbanism. *World Archaeol.* **2**, 186–197 (1970). doi: [10.1080/00438243.1970.9979473](https://doi.org/10.1080/00438243.1970.9979473)
85. W. A. Ashmore, Ed., *Lowland Maya Settlement Patterns* (Univ. of New Mexico Press, 1981).
86. T. P. Culbert, D. S. Rice, Eds., *Precolumbian Population History in the Maya Lowlands* (Univ. of New Mexico Press, 1990).
87. D. E. Puleston, *The Settlement of Survey of Tikal* (University Museum, Univ. of Pennsylvania, 1983).
88. A. Ford, *Population Growth and Social Complexity: An Examination of Settlement and Environment in the Central Maya Lowlands* (Arizona State University, 1986).
89. D. L. Webster, W. T. Sanders, P. van Rossum, A simulation of Copán population history and its implications. *Anc. Mesoam.* **3**, 185–197 (1992). doi: [10.1017/S095653610000239X](https://doi.org/10.1017/S095653610000239X)
90. D. Simon, Urban environments: Issues on the peri-urban fringe. *Annu. Rev. Environ. Resour.* **33**, 167–185 (2008). doi: [10.1146/annurev.enviro.33.021407.093240](https://doi.org/10.1146/annurev.enviro.33.021407.093240)
91. W. R. Bullard Jr., Maya settlement pattern in northeastern Peten, Guatemala. *Am. Antiq.* **25**, 355–372 (1960). doi: [10.2307/277519](https://doi.org/10.2307/277519)
92. W. A. Haviland, in *Lowland Maya Settlement Patterns*, W. A. Ashmore, Ed. (Univ. of New Mexico Press, 1981), pp. 89–120.
93. A. F. Chase, D. Z. Chase, in *Perspectives on Ancient Maya Rural Complexity*, G. Iannone, S. V. Connell, Eds. (Cotsen Institute of Archaeology, Los Angeles, 2003), pp. 108–118.
94. S. R. Hutson, *The Ancient Urban Maya: Neighborhoods, Inequality, and Built Form* (Univ. Press of Florida, 2016).
95. D. B. Marken, in *Classic Maya Politics of the Southern Lowlands: Integration, Interaction, Dissolution*, D. B. Marken, J. L. Fitzsimmons, Eds. (Univ. Press of Colorado, 2015), pp. 123–166.
96. J. P. Baron, thesis, University of Pennsylvania (2013).
97. T. Barrientos Q., M. A. Canuto, D. Stuart, L. Auld-Thomas, M. Lamoureux-St-Hilaire, in *XXIX Simposio de Investigaciones Arqueológicas en Guatemala*, B. Arroyo, L. Méndez Salinas, L. Paiz, Eds. (Ministerio de Cultura y Deportes, Instituto de Antropología e Historia, Asociación Tikal, Guatemala, 2016), pp. 1–23.
98. M. A. Canuto, T. Barrientos Q., in *La Corona Notes* (2013); [www.mesoweb.com/LaCorona/LaCoronaNotes01.pdf](http://www.mesoweb.com/LaCorona/LaCoronaNotes01.pdf).
99. M. A. Canuto, T. Barrientos Q., in *XXII Simposio de Investigaciones Arqueológicas en Guatemala*, B. Arroyo, L. Méndez Salinas, L. Paiz, Eds. (Ministerio de Cultura y Deportes, Instituto de Antropología e Historia, Asociación Tikal, Guatemala, 2018), pp. 303–314.
100. S. Martin, "Wives and Daughters on the Dallas Altar." *Mesoweb* (2008); [www.mesoweb.com/articles/Martin/Wives&Daughters.pdf](http://www.mesoweb.com/articles/Martin/Wives&Daughters.pdf).
101. D. Stuart et al., in *XXVII Simposio de Investigaciones Arqueológicas en Guatemala*, B. Arroyo, L. Méndez Salinas, A. Rojas, Eds. (Ministerio de Cultura y Deportes, Instituto de Antropología e Historia, Asociación Tikal, Guatemala, 2014), pp. 435–448.
102. M. A. Canuto, T. Barrientos Q., La Corona: Un acercamiento a las políticas del reino Kaan desde un centro secundario del noroeste de Petén. *Estud. Cult. Maya XXXVIII*, 14–43 (2011).
103. V. L. Scarborough, Ecology and ritual: Water management and the Maya. *Lat. Am. Antiq.* **9**, 135–159 (1998). doi: [10.2307/971991](https://doi.org/10.2307/971991)
104. A. S. Z. Chase, Beyond elite control: Residential reservoirs at Caracol, Belize. *Wiley Interdiscip. Rev. Water* **3**, 885–897 (2016). doi: [10.1002/wat2.1171](https://doi.org/10.1002/wat2.1171)
105. E. Akpinar-Ferrand, N. P. Dunning, D. L. Lentz, J. G. Jones, Use of aguadas as water management sources in two southern Maya Lowland sites. *Anc. Mesoam.* **23**, 85–101 (2012). doi: [10.1017/S0956536112000065](https://doi.org/10.1017/S0956536112000065)
106. J. L. Bremer, Householders as water managers: A comparison of domestic-scale water management practices from two central Maya Lowland sites. *Anc. Mesoam.* **29**, 197–217 (2018). doi: [10.1017/S0956536117000244](https://doi.org/10.1017/S0956536117000244)
107. K. L. Davis-Salazar, Late Classic Maya water management and community organization at Copan, Honduras. *Lat. Am. Antiq.* **14**, 275–299 (2003). doi: [10.2307/3557561](https://doi.org/10.2307/3557561)
108. K. J. Johnston, Lowland Maya water management practices: The household exploitation of rural wells. *Geoarchaeology* **19**, 265–292 (2004). doi: [10.1002/gea.10117](https://doi.org/10.1002/gea.10117)
109. V. L. Scarborough, in *Economic Aspects of Water Management in the Prehispanic New World*, V. L. Scarborough, B. L. Isaac, Eds. (JAI Press, 1993), pp. 17–69.
110. E. Weiss-Krejci, T. Sabbas, The potential role of small depressions as water storage features in the central Maya Lowlands. *Lat. Am. Antiq.* **13**, 343–357 (2002). doi: [10.2307/972115](https://doi.org/10.2307/972115)
111. V. L. Scarborough et al., Water and sustainable land use at the ancient tropical city of Tikal, Guatemala. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12408–12413 (2012). doi: [10.1073/pnas.1202881109](https://doi.org/10.1073/pnas.1202881109); pmid: [22802627](https://pubmed.ncbi.nlm.nih.gov/22802627/)
112. V. L. Scarborough, in *Managed Mosaic: Ancient Maya Agriculture and Resource Use*, S. L. Fedick, Ed. (Univ. of Utah Press, 1996), pp. 304–314.
113. J. Kantner, in *Movement, Connectivity, and Landscape Change in the Ancient Southwest: The 20th Anniversary Southwest Symposium*, M. C. Nelson, C. Strahacker, Eds. (Univ. Press of Colorado, 2011), pp. 363–374.
114. I. Graham, *Archaeological Explorations in El Peten, Guatemala* (Middle American Research Institute, Tulane University, 1967).
115. R. D. Hansen, in *Function and Meaning in Classic Maya Architecture*, S. D. Houston, Ed. (Dumbarton Oaks Research Library and Collection, Washington, DC, 1998), pp. 49–122.
116. E. Hernández, T. Schreiner, C. Morales Aguilar, in *XXVI Simposio de Investigaciones Arqueológicas en Guatemala*, B. Arroyo, L. Méndez Salinas, Eds. (Museo Nacional de Arqueología y Etnología, Guatemala, 2013), pp. 939–950.
117. J. M. Shaw, *Maya sacbeob: Form and function*. *Anc. Mesoam.* **12**, 261–272 (2001). doi: [10.1017/S0956536101121048](https://doi.org/10.1017/S0956536101121048)
118. D. E. Puleston, D. W. Callender Jr., Defensive earthworks at Tikal. *Expedition* **9**, 40–48 (1967).
119. D. Webster et al., The great Tikal earthwork revisited. *J. Field Archaeol.* **32**, 41–64 (2007). doi: [10.1179/00934690791071700](https://doi.org/10.1179/00934690791071700)
120. D. L. Webster, in *Archaeological Investigations on the Yucatan Peninsula*, M. A. L. Harrison, R. Wauchope, Eds. (Tulane University, 1975), pp. 123–127.
121. D. Stuart, in *Mesoamerica's Classic Heritage: From Teotihuacan to the Aztecs*, D. Carrasco, L. Jones, S. Sessions, Eds. (Univ. Press of Colorado, 2000), pp. 465–514.
122. A. F. Chase, D. Z. Chase, J. F. Weishampel, Lasers in the jungle. *Archaeology* **63**, 27–29 (2010).

## ACKNOWLEDGMENTS

The Pacunam LiDAR Initiative (PLI) was made possible because of planning, funding, and management by the Fundación Patrimonio Cultural y Natural Maya-PACUNAM. PLI's technology provider for the collection and processing of the 2016 lidar dataset was the NSF National Center for Airborne Laser Mapping (NCALM). Permissions for lidar collection and field work were granted to NCALM-Pacunam and participating archaeological projects by the Guatemalan Ministry of Culture and Sports of Guatemala. We thank three anonymous reviewers for their careful reading and helpful comments on earlier drafts of the manuscript. **Funding:** Supported by the Fundación Patrimonio Cultural y Natural Maya-PACUNAM through contributions from the Hiltz Foundation and Pacunam's Guatemalan members (Cerveceria Centro Americana, Grupo Campollo, Cementos Progreso, Blue Oil, Asociación de Azucareros de Guatemala-ASAZGUA, Grupo Occidente, Banco Industrial, Walmart Guatemala, Citi, Samsung, Disagro, Cofino Stahl, Claro, CEG, Agroamerica, and Fundación Pantaleón). Additional funding for related fieldwork and data processing was provided by the Middle American Research Institute at Tulane University, Alphasword Foundation, Louisiana Board of Regents, CNRS, LabEx DynamiTe, the Slovak Research and Development Agency APVV-0864-12, and Scientific Grant Agency VEGA 1/0858/17. **Author contributions:** M.A.C., F.E.-B., and T.G.G. contributed to the planning of the research; M.A.C., F.E.-B., T.G.G., M.J.A., M.K., D.M., P.N., and R.S. supervised the work; J.C.F.-D. collected and processed the lidar data; M.A.C., F.E.-B., T.G.G., D.M., P.N., L.A.-T., C.C., D.C., C.R.C., T.D., T.L., A.T., and A.V. performed the analyses of the lidar data; M.A.C., F.E.-B., T.G.G., S.D.H., L.A.-T., and J.C.F.-D. drafted the manuscript; and M.A.C., F.E.-B., T.G.G., L.A.-T., and T.L. designed the figures. All authors discussed the results and commented on the manuscript. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** Summary data are available in the tables included in the main text or the supplementary materials. Files containing the data derived from the lidar analyses used in this paper are available for download from <http://doi:10.5061/dryad.k51j708>.

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6409/eaau0137/suppl/DC1](http://www.sciencemag.org/content/361/6409/eaau0137/suppl/DC1)  
Materials and Methods  
Supplementary Text  
Figs. S1 and S2  
Tables S1 to S4  
References (123–144)

10.1126/science.aau0137

## REPORT

## BIOSYNTHESIS

# Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom

John K. Brunson<sup>1,2\*</sup>, Shaun M. K. McKinnie<sup>1\*</sup>, Jonathan R. Chekan<sup>1</sup>, John P. McCrow<sup>2</sup>, Zachary D. Miles<sup>1</sup>, Erin M. Bertrand<sup>2,3</sup>, Vincent A. Bielinski<sup>4</sup>, Hanna Luhavaya<sup>1</sup>, Miroslav Oborník<sup>5</sup>, G. Jason Smith<sup>6</sup>, David A. Hutchins<sup>7</sup>, Andrew E. Allen<sup>2,8†</sup>, Bradley S. Moore<sup>1,9†</sup>

Oceanic harmful algal blooms of *Pseudo-nitzschia* diatoms produce the potent mammalian neurotoxin domoic acid (DA). Despite decades of research, the molecular basis for its biosynthesis is not known. By using growth conditions known to induce DA production in *Pseudo-nitzschia multiseries*, we implemented transcriptome sequencing in order to identify DA biosynthesis genes that colocalize in a genomic four-gene cluster. We biochemically investigated the recombinant DA biosynthetic enzymes and linked their mechanisms to the construction of DA's diagnostic pyrrolidine skeleton, establishing a model for DA biosynthesis. Knowledge of the genetic basis for toxin production provides an orthogonal approach to bloom monitoring and enables study of environmental factors that drive oceanic DA production.

Oceanic harmful algal blooms (HABs) are intensifying in frequency and severity in association with climate change (1–3). This phenomenon was exemplified in the summer of 2015 when the North American Pacific coast experienced the largest HAB ever recorded, spanning the Aleutian Islands of Alaska to the Baja peninsula of Mexico (fig. S1) (4). This bloom caused widespread ecological and economic devastation, resulting in the deaths of marine mammals and the closure of beaches and fisheries. The dominant algal species in the 2015 HAB were pennate diatoms of the globally distributed genus *Pseudo-nitzschia*, which are often associated with high production of the excitatory glutamate receptor agonist domoic acid (DA; 1). Mammalian consumption of DA-contaminated shellfish exerts its toxicity at the AMPA and

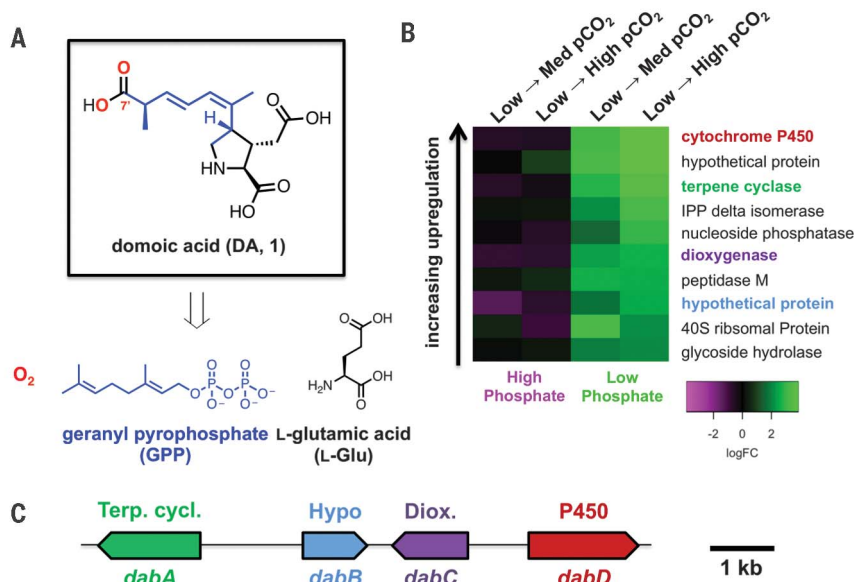
kainate ionotropic glutamate receptors of the central nervous system (5, 6). In humans, a high, single-exposure dose of DA can cause amnesic shellfish poisoning (ASP), which involves symptoms of amnesia, seizures, coma, and in extreme cases, death. Even chronic, low-level consumption of DA may lead to kidney damage, cognitive

deficit, and impairment of fetal development, making DA outbreaks an important human health problem (7–10). Similar neurotoxic symptoms have been observed in birds and marine mammals such as sea lions, which suffer spatial memory impairment linked to DA consumption, likely leading to increases in sea lion strandings (11).

Although abiotic (12) and biotic (13) factors have been shown to affect toxicity in culture, the biological and physicochemical mechanisms underlying DA production in *Pseudo-nitzschia* are unclear. Moreover, not all *Pseudo-nitzschia* blooms produce DA (12). An understanding of the genetic basis of DA biosynthesis in diatoms would facilitate determination of the cellular pathway that controls oceanic DA production and thereby diatom toxicity.

Stable isotope experiments (14, 15) suggest that DA is composed of glutamic acid (Glu) and geranyl pyrophosphate (GPP) building blocks. No pathway that involves these starting materials to construct the characteristic pyrrolidine ring of DA has been previously characterized, and thus, our ability to predict candidate biosynthetic genes using traditional bioinformatic methods was limited. However, we postulated that a redox enzyme, such as a cytochrome P450 (CYP450), generates the 7'-carboxylic acid of DA through three successive oxidations, a reaction common in all branches of life (Fig. 1A) (16).

To identify putative DA biosynthetic genes, we examined patterns of transcriptional activity under previously established conditions—phosphate limitation and elevated CO<sub>2</sub> (17)—that stimulate DA production. We created a differential expression dataset for nearly 20,000 *Pseudo-nitzschia multiseries* RNA transcripts (table S1). A small



**Fig. 1. Identification of domoic acid biosynthetic genes from transcriptomic and genomic data.**

(A) Structure of DA (1) and its proposed biosynthetic building blocks (14, 15). (B) Ten most up-regulated *P. multiseries* transcripts under previously reported DA-induction conditions (17). Differential expression changes are each shown with reference to the P<sub>CO2</sub> conditions listed across the top of the heatmap, with phosphate concentrations held constant. (C) DA biosynthesis (*dab*) gene cluster in the *P. multiseries* genome.

<sup>1</sup>Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA. <sup>2</sup>Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA 92037, USA. <sup>3</sup>Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada. <sup>4</sup>Synthetic Biology and Bioenergy Group, J. Craig Venter Institute, La Jolla, CA 92037, USA. <sup>5</sup>Institute of Parasitology, University of South Bohemia and Biology Center CAS, Branišovská 31, 370 05 České Budějovice, Czech Republic. <sup>6</sup>Moss Landing Marine Laboratories, 8272 Moss Landing Road, Moss Landing, CA 95039, USA. <sup>7</sup>Marine and Environmental Biology, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089, USA. <sup>8</sup>Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92037, USA. <sup>9</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

\*These authors contributed equally to this work.

†Corresponding author. Email: bsmoore@ucsd.edu (B.S.M.); aallen@jvci.org (A.E.A.)

fraction (~500; 2.5%) showed consistent up-regulation under phosphate limitation (fig. S2 and table S2), and further refinement of this subset to include genes that were also up-regulated with increasing partial pressure of CO<sub>2</sub> (Pco<sub>2</sub>) highlighted only 43 (0.22%) transcripts (table S3). A CYP450 gene showed the highest fold change of all analyzed transcripts under the high Pco<sub>2</sub> and low-phosphate DA-inducing condition (Fig. 1B). This transcript was the only annotated CYP450 gene out of 20 total within the *P. multiseri* genome (220 Mbp; Joint Genome Institute accession no. PRJNA32659) that showed increased transcription (fig. S3 and table S4). When mapped to the public *P. multiseri* genome, this CYP450 was localized to a compact genomic island spanning ~8 kb that possesses three other similarly up-regulated genes that are indicative of canonical gene clustering more typically observed in bacterial- and fungal-specialized metabolism (Fig. 1C) (18). Genomic organization in diatoms is not generally typified by clustering of metabolic genes (19, 20). However, clusters of transcriptionally coregulated genes, sensitive to Fe and Si limitation, have been reported (21, 22). The annotated gene functions suggest involvement in terpenoid and redox biochemistry, which are two predicted hallmarks for DA biosynthesis. These gene candidates for DA biosynthetic (Dab) enzymes were annotated as *dabA* (terpene cyclase), *dabB* (hypothetical protein), *dabC* [ $\alpha$ -ketoglutarate ( $\alpha$ KG)-dependent dioxygenase], and *dabD* (CYP450) (Fig. 1C). We independently sequenced this gene cluster from an environmental isolate of *P. multiseri* in order to validate its conservation (GenBank accession no. MH202990).

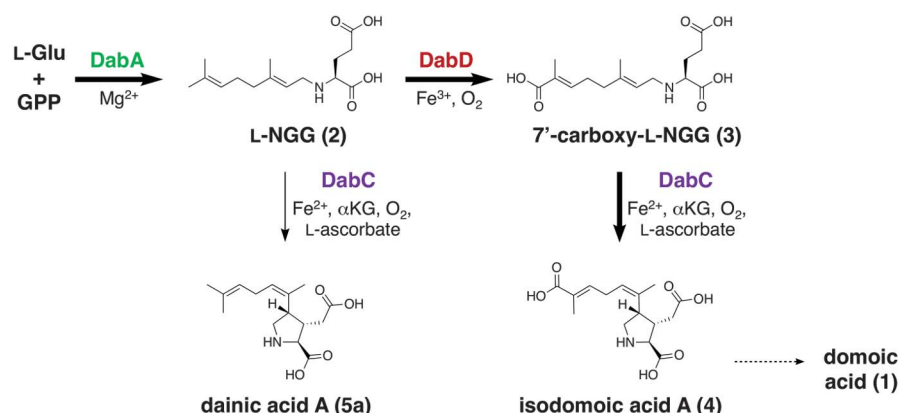
Despite the rarity of the pyrrolidine kainoid skeleton in nature, the putative *dab* gene functions enabled us to initially map our genes onto the suggested biosynthetic pathway (14, 15). We hypothesized that DabA catalyzes the first committed step of DA biosynthesis: *N*-prenylation of L-Glu with GPP to form *N*-geranyl-L-glutamic acid (L-NGG; **2**). DabC and DabD would perform subsequent oxidative reactions (Fig. 2).

We began in vitro validation of the *dab* genes with *dabA*, which contains a chloroplast transit peptide sequence and an intron but low similarity to any characterized protein in the National Center for Biotechnology Information (NCBI) database. Structural prediction by using Phyre2 (23) suggested that DabA may possess a terpene cyclase-like fold, and phylogenetic analysis intimated an evolutionary history related to red algal and bacterial genes (fig. S4). Among diatoms, *dabA* appears restricted to *Pseudo-nitzschia* spp. We expressed recombinant His<sub>6</sub>-DabA without the N-terminal transit peptide in *Escherichia coli* and purified the enzyme using Ni<sup>2+</sup> affinity and size-exclusion chromatography (fig. S5). In vitro, DabA catalyzes the *N*-geranylation of L-Glu to form L-NGG (**2**) in a Mg<sup>2+</sup>-dependent manner (Fig. 3A). DabA interrogation with structurally similar substrates shows modest promiscuity toward prenyl pyrophosphates but high specificity for Glu (fig. S6). Recently, L-NGG was isolated in low abundance from the DA-containing

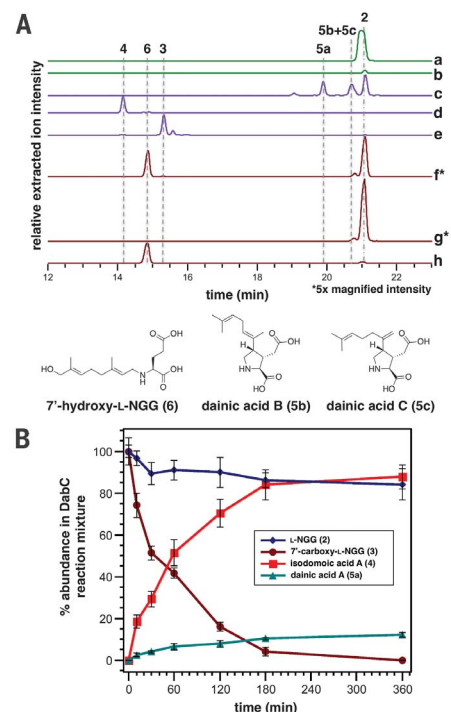
red alga *Chondria armata* and shown through labeling experiments in *P. multiseri* to be a precursor to DA (24). These observations further implicate *dabA* as a gene that encodes a major step in DA biosynthesis.

To investigate the subsequent transformations of L-NGG, we individually interrogated the activity of the DabC and DabD oxygenases toward this substrate. In the presence of Fe<sup>2+</sup>, L-ascorbic acid, and cosubstrate  $\alpha$ KG, recombinant DabC purified from *E. coli* cyclized L-NGG to form three pyrrolidine ring-containing molecules: 7-methyl-isodomoic acids A, B, and C, termed dainic acids A, B, and C (**5a** to **5c**), respectively (Fig. 3A and figs. S7 and S8), of which **5a** and **5b** had been recently isolated from *C. armata* (24). DabC, however, showed minimal cyclization of other similarly *N*-prenylated glutamic acids (fig. S7).

Enzymatic synthesis of the dainic acids was made more amenable through a one-pot coupled assay using L-Glu, GPP, DabA, DabC, and their requisite cofactors and cosubstrates (fig. S9). Despite DabC generating much of the structural diversity observed within the kainoid metabolites (25), the slow rate of L-NGG consumption and failure to go to completion led us to suspect that this enzyme may instead act on an oxidized substrate, placing DabD oxidation ahead of DabC cyclization in the DA biosynthetic pathway. Incubation of L-NGG with *Saccharomyces cerevisiae* microsomes that possess coexpressed transmembrane proteins DabD (fig. S10) and *P. multiseri* CYP450 reductase (*PmCPR1*) generated small but reproducible quantities of 7'-hydroxy-L-NGG (**6**) and 7'-carboxy-L-NGG (**3**) (Fig. 3A and fig. S11), which was validated through comparison with synthetic



**Fig. 2.** Domoic acid (**1**) biosynthetic pathway based on *dab* gene annotations and in vitro enzyme activities.



**Fig. 3.** In vitro characterization of Dab enzymes. (A) Relative intensities of negative ionization-extracted ion chromatogram liquid chromatography-mass spectrometry traces [(282.1711, 298.1660, 312.1453, 280.1554, 310.1296)  $\pm$  0.01 mass/charge ratio] for in vitro Dab reactions. a, DabA, L-Glu (500  $\mu$ M), GPP (500  $\mu$ M), MgCl<sub>2</sub> (10 mM); b, DabA, L-Glu (500  $\mu$ M), GPP (500  $\mu$ M), without MgCl<sub>2</sub>; c, DabC, L-NGG (**2**; 500  $\mu$ M), FeSO<sub>4</sub>,  $\alpha$ KG, L-ascorbic acid; d, DabC, 7'-carboxy-L-NGG (**3**; 500  $\mu$ M), FeSO<sub>4</sub>,  $\alpha$ KG, L-ascorbic acid; e, DabC, 7'-carboxy-L-NGG (**3**; 500  $\mu$ M), FeSO<sub>4</sub>,  $\alpha$ KG, L-ascorbic acid, EDTA (1.0 mM); f, DabD and *PmCPR1* containing *S. cerevisiae* microsomes, L-NGG (**2**; 500  $\mu$ M), NADPH; g, empty vector *S. cerevisiae* microsomes, L-NGG (**2**; 500  $\mu$ M), NADPH; and h, synthetic 7'-hydroxy-L-NGG (**6**) standard. Additional enzyme and cosubstrate concentrations can be found in the supplementary materials. Asterisks indicate that relative extracted ion intensities for traces f and g have been 5 $\times$  magnified. (B) Comparison of in vitro DabC substrate consumption of L-NGG (**2**; 1.0 mM) and 7'-carboxy-L-NGG (**3**; 1.0 mM) and relative formation of respective dainic acid A (**5a**) and isodomoic acid A (**4**) products ( $n = 3$  replicate DabC experiments).



standards. Conversely, the dainic acid isomers were not substrates of the DabD P450 (fig. S11). 7'-carboxy-L-NGG incubation with DabC showed rapid cyclization to the *P. multiseriis* natural product isodomoic acid A (**4**) (Fig. 3A), which did not occur in the presence of metal chelator ethylenediaminetetraacetic acid (EDTA). When comparing both DabC in vitro substrates for physiological relevance, 7'-carboxy-L-NGG was consumed at a much faster rate than L-NGG, further suggesting that 7'-carboxy-L-NGG is an on-pathway intermediate (Fig. 3B and fig. S12). Our cumulative in vitro biochemical results imply a DA biosynthetic pathway that begins with the DabA-catalyzed geranylation of L-Glu to yield L-NGG, likely in the chloroplast. DabD then performs three successive oxidation reactions at the 7'-methyl of L-NGG to produce 7'-carboxy-L-NGG, which is then cyclized by DabC to generate the naturally occurring isodomoic acid A (Fig. 2). A putative isomerase likely converts isodomoic acid A to DA. We biochemically interrogated the coclustered *dabB* gene product but did not observe isomerase activity (fig. S13). Further examination of additionally up-regulated transcripts did not suggest an obvious candidate gene (fig. S14). We are actively investigating this final isomerization reaction to complete the pathway to DA.

In addition to *Pseudo-nitzschia* spp., the kainoid structure has been observed in other diatom, red macroalgal, and fungal compounds (fig. S15) (12, 26, 27). In silico application of the *dab* genes as a kainoid ring biosynthetic query identified *dabA* and *dabC* homologs in several red algal transcriptomes (28, 29), including the known kainic acid producer *Palmaria palmata* (fig. S16). Like DA, kainic acid (26) is a structurally related glutamate agonist constructed from a shorter terpene substrate without the need of the DabD P450 oxidation.

With the establishment of the DabACD-dependent biosynthetic pathway to isodomoic acid A in *P. multiseriis*, we next linked this discovery to the marine environment. We identified *dab* genes with the same genetic organization and high sequence identity in the genome of the known DA-producing *Pseudo-nitzschia multistriata* (fig. S17) (30). Moreover, of the eight publicly available *Pseudo-nitzschia* transcriptomes, only the

highly toxic DA-producing species *Pseudo-nitzschia australis* expressed the *dab* genes (figs. S18 and S19) (31). No *dab* homologs were found in any other sequenced microalgal genera. By virtue of its limited distribution, *dabA* thus presents an opportunity for genetic monitoring of the DA-producing capabilities of *Pseudo-nitzschia* blooms orthogonal to currently established mass spectrometry-based and enzyme-linked immunosorbent assay-based identification approaches. We anticipate that knowledge of the *dab* genes will allow for greater understanding of the basis and diversity of *Pseudo-nitzschia* toxicity, the physiological function of DA, and the environmental conditions that promote HAB formation so that we may better anticipate risk of exposure to this toxic marine natural product.

## REFERENCES AND NOTES

- M. L. Wells et al., *Harmful Algae* **49**, 68–93 (2015).
- S. M. McKibben et al., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 239–244 (2017).
- C. J. Gobler et al., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4975–4980 (2017).
- R. M. McCabe et al., *Geophys. Res. Lett.* **43**, 10366–10376 (2016).
- G. R. Stewart, C. F. Zorumski, M. T. Price, J. W. Olney, *Exp. Neurol.* **110**, 127–138 (1990).
- J. A. Larm, P. M. Beart, N. S. Cheung, *Neurochem. Int.* **31**, 677–682 (1997).
- J. A. Funk et al., *J. Am. Soc. Nephrol.* **25**, 1187–1197 (2014).
- K. A. Lefebvre et al., *Harmful Algae* **64**, 20–29 (2017).
- J. S. Ramsdell, T. S. Zabka, *Mar. Drugs* **6**, 262–290 (2008).
- L. M. Grattan et al., *Toxins (Basel)* **10**, 103 (2018).
- P. F. Cook et al., *Science* **350**, 1545–1547 (2015).
- A. Lelong, H. Hégaret, P. Soudant, S. S. Bates, *Phycologia* **51**, 168–216 (2012).
- M. P. Sison-Mangus, S. Jiang, K. N. Tran, R. M. Kudela, *ISME J.* **8**, 63–76 (2014).
- U. P. Ramsey, D. J. Douglas, J. A. Walter, J. L. Wright, *Nat. Toxins* **6**, 137–146 (1998).
- T. J. Savage, G. J. Smith, A. T. Clark, P. N. Saucedo, *Toxicol.* **59**, 25–33 (2012).
- X. Zhang, S. Li, *Nat. Prod. Rep.* **34**, 1061–1089 (2017).
- J. Sun et al., *Limnol. Oceanogr.* **56**, 829–840 (2011).
- M. H. Medema et al., *Nat. Chem. Biol.* **11**, 625–631 (2015).
- E. V. Armbrust et al., *Science* **306**, 79–86 (2004).
- C. Bowler et al., *Nature* **456**, 239–244 (2008).
- A. E. Allen et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10438–10443 (2008).
- G. Sapriel et al., *PLOS ONE* **4**, e7458 (2009).
- L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **10**, 845–858 (2015).
- Y. Maeno et al., *Sci. Rep.* **8**, 356 (2018).
- J. Clayden, B. Read, K. R. Hebditch, *Tetrahedron* **61**, 5713–5724 (2005).
- I. Nitta, H. Watase, Y. Tomiie, *Nature* **181**, 761–762 (1958).
- K. Konno, H. Shirahama, T. Matsumoto, *Tetrahedron Lett.* **24**, 939–942 (1983).
- G. W. Saunders, C. Jackson, E. D. Salomaki, *Mol. Phylogenet. Evol.* **119**, 151–159 (2018).
- N. Matasci et al., *Gigascience* **3**, 17 (2014).
- S. Basu et al., *New Phytol.* **215**, 140–156 (2017).
- P. J. Keeling et al., *PLOS Biol.* **12**, e1001889 (2014).

## ACKNOWLEDGMENTS

We acknowledge B. Duggan, A. Mrse, and Y. Su (all University of California, San Diego) for assistance and maintenance of nuclear magnetic resonance and high-resolution mass spectrometry machinery; W. Fenical (Scripps Institution of Oceanography) for helpful discussion and feedback; and M. Kahr (Scripps Institution of Oceanography) for providing processed satellite images. **Funding:** This work was supported by grants from the National Science Foundation (NSF OCE-1313747 to B.S.M., NSF-ANT-1043671 to A.E.A., and OCE 1538525 and OCE 1638804 to D.A.H.), the National Institute of Environmental Health Sciences (NIEHS P01-ES021921 to B.S.M.), the U.S. Department of Energy Genomics Science program (DE-SC0008593 and DE-SC0018344 to A.E.A.), the Gordon and Betty Moore Foundation (GBMF3828 to A.E.A. and GBMF4960 to G.J.S.), and the Czech Science Foundation (18-13458S to M.O.) and funding from the National Institutes of Health (Chemical Biology Interfaces—University of California, San Diego, training grant 5T32GM112584 to J.K.B.), the Dickinson-McCrink Fellowship (J.K.B.), the Natural Sciences and Engineering Research Council of Canada (NSERC-PDF to S.M.K.M.), and the Simons Foundation Fellowship of the Life Sciences Research Foundation (J.R.C.). **Author contributions:** J.K.B., S.M.K.M., J.R.C., A.E.A., and B.S.M. conceived the project, designed the experiments, analyzed the data, and wrote the paper, with input from all authors. J.K.B., S.M.K.M., J.R.C., and Z.D.M. performed protein expression and did the enzymology experiments. S.M.K.M. and J.K.B. performed chemical synthesis and structural characterization of enzymatic products. J.P.M., E.M.B., J.K.B., and A.E.A. performed the transcriptome analyses. V.A.B. designed vectors and expression constructs and assisted with sequencing. H.L. designed the yeast expression experiments. M.O. constructed the phylogeny. G.J.S. provided the *P. multiseriis* isolate 15091C3. D.A.H. provided *P. multiseriis* biomass for RNA sequencing from the DA induction experiments (17) and helped to design RNA sequencing experiments. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Transcriptome sequences are deposited in NCBI's Sequence Read Archive, BioProjectIDs SAMN08773590-SAMN08773622. The sequence of the *dab* cluster as amplified from *P. multiseriis* 15091C3 is deposited in GenBank, accession no. MH202990. Data used to construct fig. S1 are courtesy of the NASA Ocean Biology Processing Group (OBPG) and are available from <https://oceancolor.gsfc.nasa.gov>. All other data needed to evaluate the paper are present in the supplementary materials.

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6409/1356/suppl/DC1](http://www.sciencemag.org/content/361/6409/1356/suppl/DC1)  
Materials and Methods  
Supplementary Text  
Figs. S1 to S19  
References (32–64)  
Tables S1 to S4

3 May 2018; accepted 15 August 2018  
10.1126/science.aau0382

## ULTRAFAST OPTICS

# Ultrafast electro-optic light with subcycle control

David R. Carlson<sup>1\*</sup>, Daniel D. Hickstein<sup>1†</sup>, Wei Zhang<sup>1</sup>, Andrew J. Metcalf<sup>1</sup>, Franklyn Quinlan<sup>1</sup>, Scott A. Diddams<sup>1,2</sup>, Scott B. Papp<sup>1,2\*</sup>

Light sources that are ultrafast and ultrastable enable applications like timing with subfemtosecond precision and control of quantum and classical systems. Mode-locked lasers have often given access to this regime, by using their high pulse energies. We demonstrate an adaptable method for ultrastable control of low-energy femtosecond pulses based on common electro-optic modulation of a continuous-wave laser light source. We show that we can obtain 100-picojoule pulse trains at rates up to 30 gigahertz and demonstrate sub-optical cycle timing precision and useful output spectra spanning the near infrared. Our source enters the few-cycle ultrafast regime without mode locking, and its high speed provides access to nonlinear measurements and rapid transients.

Ultrafast lasers produce trains of femtosecond-duration light pulses and can operate as frequency combs to provide a time and frequency reference bridging the optical and microwave domains of the electromagnetic spectrum (1). Achieving phase control of these pulse trains to better than a single optical cycle has enabled diverse applications ranging from optical atomic clocks (2) to controlling quantum states of matter (3, 4). These capabilities have evolved over decades, and yet they still require the intrinsic stability of a suitable mode-locked resonator.

One alternative method that produces optical pulse trains without mode locking is electro-optic modulation (EOM) of a laser (5, 6). These pulse generators, or “EOM combs,” first gained interest nearly 50 years ago because of their simplicity, tunability, reliability, commercialization, and spectral flatness (7–10). Nevertheless, despite their broad appeal and decades of development, the fundamental goal of electronic switching with the optical-cycle precision needed to create ultrafast trains of EOM pulses has remained unmet, limited by thermodynamic noise and oscillator phase noise inherent in electronics.

Here, we report the generation of ultrafast and ultrastable electro-optic pulses without any mode locking. Our experiments demonstrate widely applicable techniques to mitigate electro-optic noise by relying on the quantum-limited optical processes of cavity transmission, nonlinear interferometry, and nonlinear optical pulse compression, as well as low-loss microwave interferometry. This results in phase control of ultrafast electro-optic fields with a temporal precision better than one cycle of the optical carrier. Because electro-

optic sources support pulse repetition rates greater than 10 GHz, our work opens up the regime of high-speed, ultrafast light sources, enabling sampling or excitation of high-speed transient events, as well as making precision measurements across octaves of bandwidth.

We demonstrate the performance of our ultrafast phase control by directly carving electro-optic pulse trains at 10 and 30 GHz with ~1-ps initial pulse durations and show that these pulses can be spectrally broadened to octave bandwidths and temporally compressed to less than three optical cycles (15 fs) in nanophotonic silicon-nitride ( $\text{Si}_3\text{N}_4$ , henceforth SiN) waveguides. To deliver a femtosecond source timed with subcycle precision, we introduce an EOM-comb configuration implementing high- $Q$  microwave-cavity stabilization of the 10-GHz electronic oscillator. This oscillator is phase-locked to the continuous-wave (CW) pump source via  $f - 2f$  stabilization of the carrier-envelope offset, enabling complete knowledge of the ~28,000 EOM-comb frequencies to 17 digits. Our implementation uses a cavity-stabilized CW laser to demonstrate subhertz-linewidth modes spanning the near infrared, but we note that more standard pump sources could achieve the same relative stability between the microwave source and optical carrier.

Our EOM comb is derived from a microwave source that drives an intensity modulator placed in series with multiple phase modulators to produce a 50%-duty-cycle pulse train with mostly linear frequency chirp (Fig. 1). In the spectral domain, this process results in a deterministic cascade of sidebands with prescribed amplitude and phase that converts the CW laser power into a frequency comb with a mode spacing given by the microwave driving frequency  $f_{\text{eo}}$ . The frequency of each resulting mode  $n$ , counted from the CW laser at frequency  $\nu_p$ , can then be expressed as  $\nu_n = \nu_p + n f_{\text{eo}}$ . Equivalently, the modes can be expressed as a function of the classic offset frequency  $f_0$  and repetition rate  $f_{\text{rep}}$  parameters as  $\nu_n = f_0 + n f_{\text{rep}}$ , where now the mode number  $n'$  is counted from zero frequency and  $f_{\text{rep}} = f_{\text{eo}}$ .

In order for the EOM comb to achieve ultra-stable coherence between  $\nu_p$  and  $f_{\text{eo}}$ , it is vital to keep the integrated phase noise of each mode below  $\pi$  radians. In the temporal domain, this corresponds to subcycle timing jitter, and for EOM combs this requirement becomes more difficult to achieve as the comb bandwidth is increased because of microwave-noise multiplication (11). For octave-spanning spectra at a 10-GHz repetition rate, this multiplication factor is  $n' \approx 20,000$  and corresponds to an 86-dB increase in phase noise. Thus, reaching the  $\pi$ -radian threshold with an EOM comb requires careful treatment of the noise at all Fourier frequencies.

As noted earlier (10), broadband thermal noise in the electronic components up to the Nyquist frequency causes the phase-coherence threshold to be exceeded. To compensate, a Fabry-Pérot cavity optically filters the broadband thermal noise fundamental to electro-optic modulation, resulting in a detectable carrier-envelope offset frequency. However, the cavity linewidth (typically a few megahertz) places a lower bound on the range of frequencies where this suppression is possible, and therefore, it is additionally necessary to investigate the use of low-noise microwave oscillators. This is especially important for the Fourier-frequency range between 100 kHz and the filter-cavity linewidth, where high-gain feedback is technically challenging.

In the stabilized EOM comb (a comprehensive system diagram is shown in fig. S1), we use a commercial dielectric-resonator oscillator (DRO) with a nominal operating frequency of 10 GHz and 0.1% tuning range to drive the modulators. Compared to other commercial microwave sources, the DRO offers improved phase-noise performance in the critical Fourier-frequency range between 100 kHz and 10 MHz. The DRO output is then amplified before driving the phase modulators to produce the typical comb spectrum shown in Fig. 2A.

After transmission through an optical-filter cavity to suppress thermal noise, the chirped-pulse output of the EOM comb is compressible to durations as short as 600 fs, depending on the initial spectral bandwidth. Unfortunately, pulse durations greater than ~200 fs pose problems for coherent supercontinuum broadening in nonlinear media with anomalous dispersion (12, 13). However, if the nonlinear material instead exhibits normal dispersion, broadening due to pure self-phase modulation is known to produce lower-noise spectra owing to the suppression of modulation instability (13). Consequently, we employ a two-stage broadening scheme using a normal-dispersion highly nonlinear fiber (HNLF) to achieve initial spectral broadening (14–16) and pulse compression to 100 fs (17), followed by an anomalous-dispersion SiN waveguide for broad spectrum generation.

High-repetition rate lasers ( $f_{\text{rep}} \geq 10$  GHz) produce lower pulse energies for the same average power, making it challenging to use nonlinear broadening to produce the octave bandwidths required for self-referencing. However, patterned nanophotonic waveguides have recently emerged

<sup>1</sup>Time and Frequency Division, National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA.

<sup>2</sup>Department of Physics, University of Colorado, 2000 Colorado Avenue, Boulder, CO 80309, USA.

\*Corresponding author. Email: david.carlson@nist.gov; scott.papp@nist.gov

†Present address: KMLabs, Inc., 4775 Walnut Street, Suite 102, Boulder, CO 80301, USA.

as a promising platform owing, in part, to their high nonlinearity and engineerable dispersion (18–20). Here, we demonstrate input-coupling efficiency to a SiN waveguide of up to 85% (17) that enables a broadband supercontinuum to be generated with pulses from high-repetition rate ultrafast sources. The spectra generated with our 10-GHz EOM comb spans wavelengths from 750 nm to beyond 2700 nm for two different waveguide geometries (Fig. 2B), producing a total integrated power of ~1.1 W. Individual comb lines across the entire bandwidth exhibit a high degree of extinction (50 dB at 1064 nm; see fig. S2 for data at 775, 1064, and 1319 nm) and do not exhibit any intermode artifacts such as sidebands, a common problem when mode filtering is used to convert low-repetition rate combs to high repetition rates (21).

To investigate the scalability to even higher repetition rates, we made additional supercontinuum measurements using a 30-GHz EOM comb, which produced 600-fs, 70-pJ pulses (Fig. 2D). Despite the three-times reduction in pulse energy compared to the 10-GHz comb, similar broadband spectra are readily obtained. In both cases, if the waveguide input pulse energy is kept below ~100 pJ, smooth spectra can be obtained with high power per comb mode.

For applications requiring very flat spectra over broad bandwidths, such as astronomical spectrograph calibration (22), the supercontinuum

light can be easily collected in a single-mode fiber and flattened with a single passive optical attenuator. Under these conditions, fluctuations in spectral intensity can be kept within  $\pm 3$  dB over wavelengths spanning from 850 to 1450 nm while delivering more than 10 nW per mode in the fiber at 10 GHz. Improved waveguide-to-fiber output coupling, or free-space collimation combined with an appropriate color filter, could further improve the power per mode.

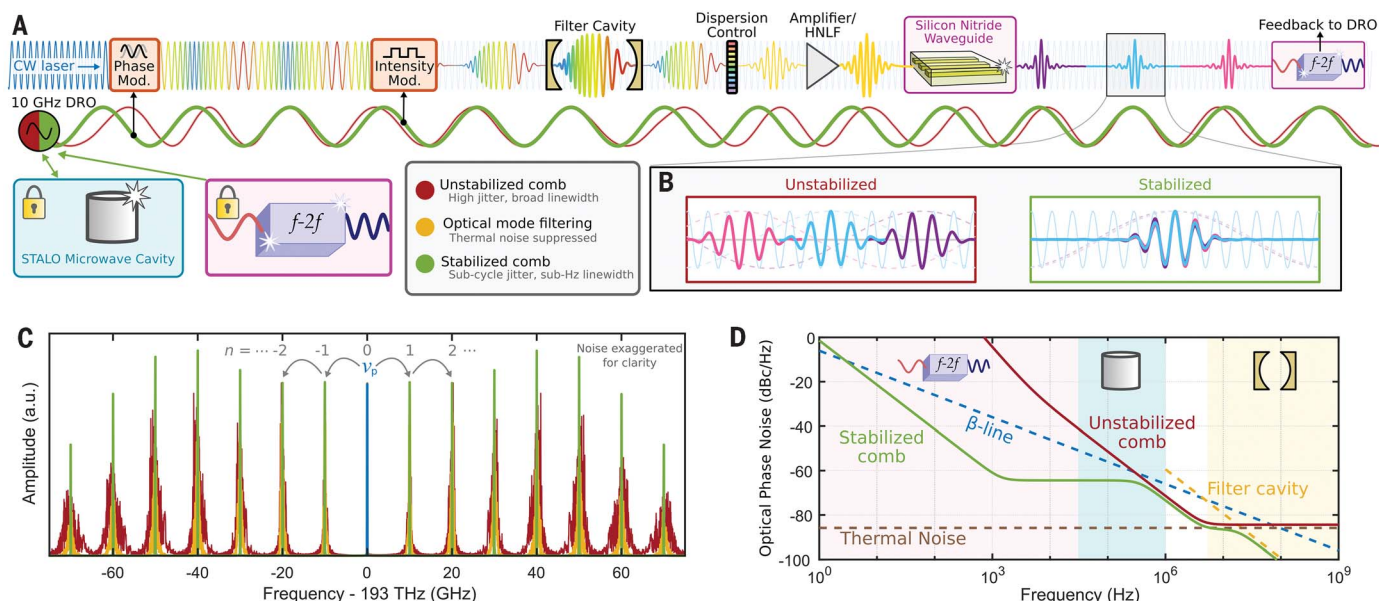
After broadening in the SiN waveguide, the offset frequency is detected with >30 dB signal-to-noise ratio (SNR), suggesting that the scheme of combining normal- and anomalous-dispersion media indeed allows us to overcome the difficulties of producing a coherent supercontinuum using pulses longer than a few hundred femtoseconds; see fig. S3 for SNR versus bandwidth. Stabilization of  $f_0$  is subsequently accomplished by feeding back to the frequency-tuning port of the DRO. However, owing to optical and electronic phase delay in this configuration, the feedback bandwidth is limited to ~200 kHz (Fig. 3A, blue curve) and thus is insufficient on its own to narrow the comb linewidth set by the multiplied microwave noise of the DRO.

To reach the  $\pi$ -radian threshold for phase coherence between the CW laser and electronic oscillator, the output of one high-power microwave amplifier is stabilized to an air-filled aluminum

microwave cavity in the stabilized-local-oscillator (STALO) configuration (23, 24) and yields an immediate reduction in phase noise of up to 20 dB at frequencies less than 500 kHz from the carrier.

In Fig. 3A, we use the  $\beta$ -line (25) to distinguish between regimes where the linewidth of the comb offset  $f_0$  is adversely affected (phase noise above the  $\beta$ -line) and where there is no linewidth contribution (phase noise below the  $\beta$ -line). Having phase noise below the  $\beta$ -line at all points is approximately equivalent to an integrated phase noise below  $\pi$  radians, and thus provides a convenient visual way to assess the impact of noise at different Fourier frequencies. For our EOM comb, the  $f_0$  phase noise remains below the  $\beta$ -line at all frequencies only when both the STALO lock and the  $f - 2f$  lock are used in tandem. Under these conditions, noise arising from the microwave oscillator does not contribute appreciably to the comb linewidth and thus, the CW laser stability is faithfully transferred across the entire comb bandwidth. Equivalently, integrating the phase noise of the fully locked  $f_0$  beat (1.17 rad, 10 Hz to 4 MHz) yields a pulse-to-pulse timing jitter of 0.97 fs (1.9 fs if limited by the  $\beta$  line between 4 MHz and 5 GHz) (17), indicating that the microwave envelope coherently tracks the optical carrier signal with subcycle precision.

The progression of offset-frequency stabilization is also shown by the beat frequencies as each



**Fig. 1. Carving femtosecond pulses from a continuous-wave (CW) laser with subcycle precision.** (A) A chirped pulse train is derived from a 1550-nm CW laser by electro-optic phase and intensity modulation driven by a 10-GHz dielectric resonant oscillator (DRO) that is locked to a high-Q microwave cavity in the stabilized-local-oscillator (STALO) configuration. The pulse train is then optically filtered by a Fabry-Pérot cavity to suppress electronic thermal noise on the comb lines before spectral broadening in highly nonlinear fiber (HNLF) followed by a silicon-nitride waveguide. Octave-spanning spectra allow detection of the comb offset frequency in an  $f - 2f$  interferometer that is used to stabilize the DRO output. (B) Without stabilization, the microwave-derived pulse train exhibits large pulse-to-pulse timing jitter relative

to the CW carrier. When the drive frequency is stabilized by feedback from the comb offset frequency and the STALO cavity, sub-optical cycle phase coherence between successive pulses is achieved. Note that the stabilized pulses are shown with zero carrier-envelope offset, though this is not generally the case. (C) In the frequency-domain picture, the unstabilized comb (red) exhibits large noise multiplication as the mode number  $n$  expands about zero. Mode filtering (yellow) suppresses high-frequency thermal noise. The fully stabilized comb lines (green) appear as  $\delta$ -functions because the CW-laser stability is transferred across the entire comb bandwidth. (D) Optical phase noise picture of the comb, showing the effects of the  $f - 2f$  stabilization, STALO cavity, and filter cavity.



lock is turned on (Figs. 3, B to D). The coherent carrier seen in the offset frequency when fully stabilized (Fig. 3D) indicates that phase coherence has been achieved between individual comb lines across the entire available spectral bandwidth. The accuracy and precision of the stabilized EOM comb were determined by beating the 10-GHz repetition rate against the 40th har-

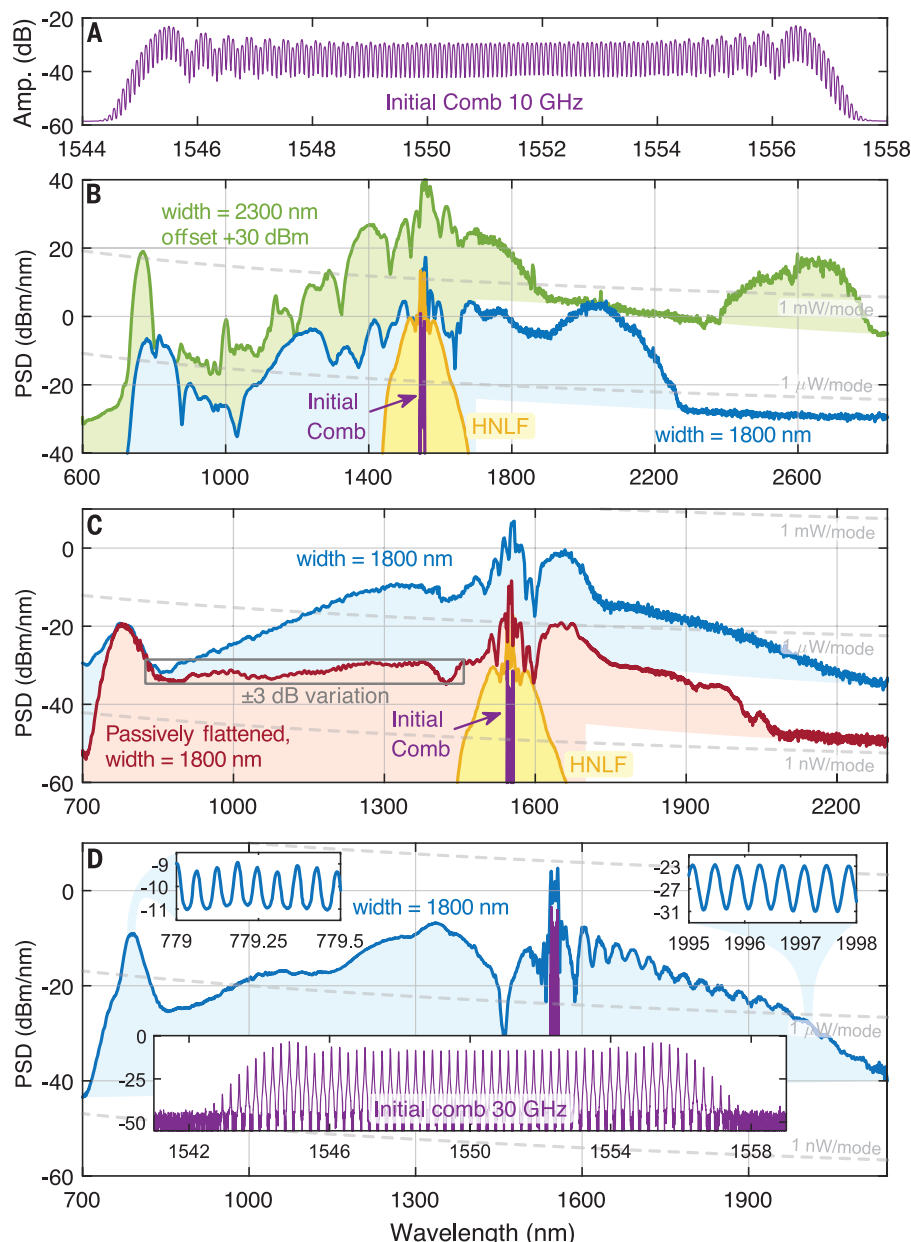
monic of an independent mode-locked laser operating at 250 MHz (17). After 2000 s of averaging, a fractional stability of  $3 \times 10^{-17}$  was obtained with no statistically significant frequency offset observed. This level of accuracy represents an improvement of more than three orders of magnitude over previously demonstrated EOM-comb systems (10) and is likely only limited here by

averaging time and out-of-loop path differences between the two combs.

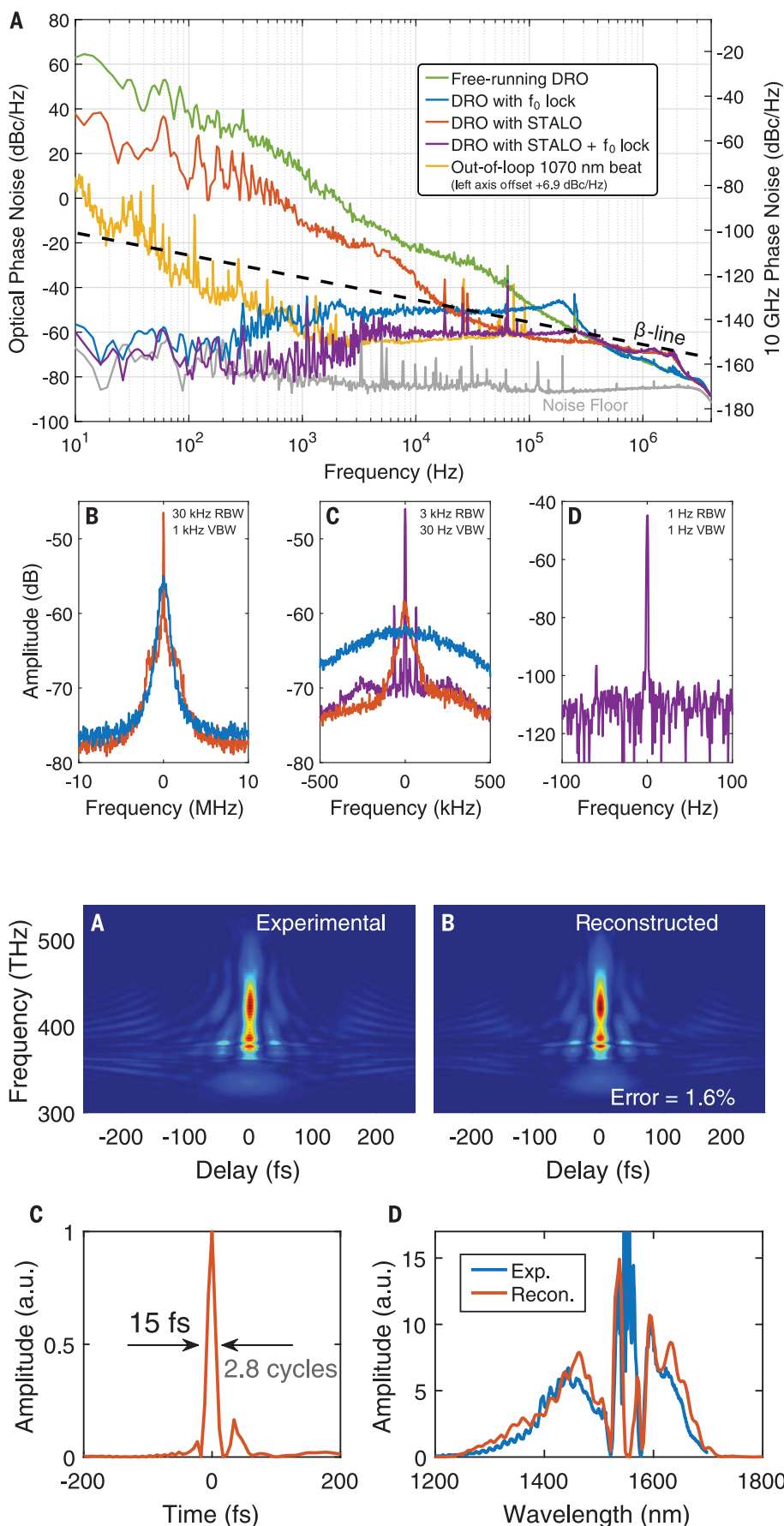
To further show the versatility of the EOM comb as an ultrafast source, we describe how to create pulses that have durations lasting only a few cycles of the optical field. Pulses in this regime can provide direct access to the carrier-envelope phase and high peak intensities but require a well-controlled output spectrum exhibiting a high degree of spectral flatness and coherence. However, achieving such pulses at gigahertz repetition rates with mode-locked lasers is technically challenging. Still, high-repetition rate sources of few-cycle pulses could be valuable for applications like optically controlled electronics (26, 27), where both fast switching speeds and peak intensity are important. Similarly, coherent Raman imaging of biological samples can benefit from transform-limited ultrashort pulses (28), but the acquisition speed for broadband spectra is typically restricted by the megahertz-rate mode-locked laser sources that are used. Extending to higher repetition rates could reduce measurement dead time and also prevent sample damage due to high peak powers (29).

The use of optical modulators to directly carve a train of ~1-ps pulses from a CW laser provides an effective method for generating clean few-cycle pulses thanks to the soliton self-compression effect (30, 31). To achieve this, the pulse power and chirp incident on the SiN waveguide are adjusted such that the launched pulse approaches the threshold peak intensity for soliton fission near the output facet of the chip. A normal-dispersion single-element aspheric lens is then used to out-couple the light without introducing appreciable higher-order dispersion, and a 2-cm-long rod of fused silica recompresses the pulse to near its transform limit. Figure 4 shows the reconstructed pulse profile obtained through frequency-resolved optical gating (FROG) (32). Pulse durations of 15 fs (2.8 optical cycles, full width at half maximum) and out-coupled pulse energies in excess of 100 pJ (1 W average power) are readily achievable at a repetition rate of 10 GHz.

The combination of high-repetition rate pulse trains, ultrastable broadband frequency synthesis, few-cycle pulse generation, and extensible construction in our EOM-comb system provides a versatile ultrafast source with other additional practical benefits. For instance, these combs could also support further photonic integration through complementary metal-oxide-semiconductor (CMOS)-compatible modulators (33), alignment-free construction, the use of commercially sourceable components, and straightforward user customization. Moreover, whereas the optical and microwave cavities currently limit the broad tuning capability of the repetition rate, the ~300 THz of comb bandwidth places a mode within 5 GHz of any spectral location in this range. By overcoming several experimental challenges related to broadening and stabilizing noisy picosecond-duration pulses, our techniques are widely applicable to existing technologies with demanding requirements, such as chip-based microresonators (34) or semiconductor lasers (35).



**Fig. 2. High-repetition-rate supercontinuum.** (A) Spectrum of the 10-GHz EOM comb directly after generation. (B) Ten-gigahertz supercontinuum spectra spanning from 750 to 2750 nm for two different silicon-nitride waveguide widths. The spectral intensity is scaled to intrawaveguide levels. Also shown is the spectrum of the first-stage highly nonlinear fiber (HNLF). (C) Ten-gigahertz supercontinuum optimized for spectral smoothness by reducing incident power (blue). Between 830 and 1450 nm, a flat spectrum ( $\pm 3$  dB) is produced by a single passive optical attenuator (red). (D) Supercontinuum spectrum from a 30-GHz EOM comb. Top insets show that comb coherence is maintained across the entire spectrum (optical SNRs are spectrometer limited). Bottom inset shows initial spectrum of the 30-GHz EOM comb. The y axes in both (C) and (D) show the power spectral density (PSD) obtained in the output fiber.



**Fig. 3. EOM-comb phase noise.** (A) Optical phase noise of the comb offset frequency measured at 775 nm (left axis) and scaled to the 10-GHz repetition rate (right axis) under different locking conditions. Prestabilizing the free-running RF oscillator (DRO) using a high-Q microwave cavity in the stabilized-local-oscillator (STALO) configuration lowers the phase noise by up to 20 dB at frequencies below 500 kHz. When servo feedback from the optical  $f_0$  signal is engaged, a tight phase lock is achieved that suppresses low-frequency noise. The  $\beta$ -line indicates the level above which phase noise causes an increase in the comb linewidth. When both the STALO and  $f_0$  locks are engaged, the phase noise remains below the  $\beta$ -line at all frequencies, indicating that the coherence of the CW pump laser is faithfully transferred across the entire comb spectrum. (B to D)  $f_0$  RF beats showing the effects of each feedback loop. A coherent carrier signal is observed (D) only when both the STALO lock and direct  $f_0$  feedback are engaged.

**Fig. 4. Few-cycle pulse generation.** (A) Experimental and (B) reconstructed FROG traces. (C) Reconstructed temporal pulse profile with a full width at half maximum duration of 15 fs (2.8 optical cycles). (D) Comparison of reconstructed and experimental spectra. The quasi-CW spectral wings of the initial comb spectrum near 1550 nm do not contribute appreciably to the pulse and thus are not seen in the reconstructed spectrum. At least 75% of the total optical power is concentrated in the compressed pulse. More sophisticated amplitude and phase compensation of the initial comb spectrum could allow an even larger fraction of the power to be compressed (36).

## REFERENCES AND NOTES

1. S. T. Cundiff, J. Ye, *Rev. Mod. Phys.* **75**, 325–342 (2003).
2. A. D. Ludlow, M. M. Boyd, J. Ye, E. Peik, P. Schmidt, *Rev. Mod. Phys.* **87**, 637–701 (2015).
3. C. Brif, R. Chakrabarti, H. Rabitz, *New J. Phys.* **12**, 075008 (2010).
4. D. S. Jin, J. Ye, *Chem. Rev.* **112**, 4801–4802 (2012).
5. T. Kobayashi, T. Sueta, Y. Cho, Y. Matsuo, *Appl. Phys. Lett.* **21**, 341–343 (1972).
6. M. Kourogi, K. Nakagawa, M. Ohtsu, *IEEE J. Quantum Electron.* **29**, 2693–2701 (1993).
7. V. Torres-Company, A. M. Weiner, *Laser Photonics Rev.* **8**, 368–393 (2014).
8. J. Li, X. Yi, H. Lee, S. A. Diddams, K. J. Vahala, *Science* **345**, 309–313 (2014).
9. G. Millot *et al.*, *Nat. Photonics* **10**, 27–30 (2015).
10. K. Beha *et al.*, *Optica* **4**, 406 (2017).
11. F. L. Walls, A. DeMarchi, *IEEE Trans. Instrum. Meas.* **24**, 210–217 (1975).
12. J. M. Dudley, G. Genty, S. Coen, *Rev. Mod. Phys.* **78**, 1135–1184 (2006).
13. K. R. Tamura, H. Kuhota, M. Nakazawa, *IEEE J. Quantum Electron.* **36**, 773–779 (2000).
14. C.-B. Huang, S.-G. Park, D. E. Leaird, A. M. Weiner, *Opt. Express* **16**, 2520–2527 (2008).
15. R. Wu, V. Torres-Company, D. E. Leaird, A. M. Weiner, *Opt. Express* **21**, 6045–6052 (2013).
16. V. Ataie, E. Myslivets, B. P.-P. Kuo, N. Alic, S. Radic, *J. Lightwave Technol.* **32**, 840–846 (2014).
17. Supplementary materials are available online.
18. A. R. Johnson *et al.*, *Opt. Lett.* **40**, 5117–5120 (2015).
19. M. A. G. Porcel *et al.*, *Opt. Express* **25**, 1542–1554 (2017).
20. D. R. Carlson *et al.*, *Opt. Lett.* **42**, 2314–2317 (2017).
21. D. A. Braje, M. S. Kirchner, S. Osterman, T. Fortier, S. A. Diddams, *Eur. Phys. J. D* **48**, 57–66 (2008).
22. T. Steinmetz *et al.*, *Science* **321**, 1335–1337 (2008).
23. G. J. Dick, J. Saunders, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **37**, 339–346 (1990).
24. A. S. Gupta *et al.*, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **51**, 1225–1231 (2004).
25. G. Di Domenico, S. Schilt, P. Thomann, *Appl. Opt.* **49**, 4801–4807 (2010).
26. T. Rybka *et al.*, *Nat. Photonics* **10**, 667–670 (2016).
27. C. Kernetzky *et al.*, *Nat. Commun.* **9**, 2471 (2018).
28. C. H. Camp Jr., M. T. Cicerone, *Nat. Photonics* **9**, 295–305 (2015).
29. K. J. Mohler *et al.*, *Opt. Lett.* **42**, 318–321 (2017).
30. G. P. Agrawal, *Nonlinear Fiber Optics* (Academic Press, 2013).
31. L. F. Mollenauer, R. H. Stolen, J. P. Gordon, W. J. Tomlinson, *Opt. Lett.* **8**, 289–291 (1983).
32. R. Trebino *et al.*, *Rev. Sci. Instrum.* **68**, 3277–3295 (1997).
33. G. T. Reed, G. Mashanovich, F. Y. Gardes, D. J. Thomson, *Nat. Photonics* **4**, 518–526 (2010).
34. T. Herr *et al.*, *Nat. Photonics* **8**, 145–152 (2013).
35. B. W. Tilma *et al.*, *Light Sci. Appl.* **4**, e310 (2015).
36. Z. Jiang *et al.*, *IEEE J. Quantum Electron.* **43**, 1163–1174 (2007).

## ACKNOWLEDGMENTS

We thank H. Timmers for assistance making the FROG measurements, A. Hati and C. Nelson for discussions on microwave stabilization, L. Chang for helping lay out the

waveguide lithography masks, K.V. Reddy and S. Patil for discussion of erbium amplifiers, and F. Baynes for constructing the cavity-stabilized laser. **Funding:** This research is supported by the Air Force Office of Scientific Research (AFOSR) under award no. FA9550-16-1-0016, the Defense Advanced Research Projects Agency (DARPA) DODOS program, the National Aeronautics and Space Administration (NASA), the National Institute of Standards and Technology (NIST), and the National Research Council (NRC). This work is a contribution of the U.S. government and is not subject to copyright in the U.S.A. **Author contributions:** The experiment was planned by D.R.C., D.D.H., S.A.D., and S.B.P. The combs were operated by D.R.C. and D.D.H. The optical filter cavity was designed and constructed by W.Z. The 30-GHz measurements were assisted by A.J.M. The data were analyzed by D.R.C., D.D.H., F.Q., S.A.D., and S.B.P. The manuscript was prepared by D.R.C. with input from all coauthors. **Competing interests:** NIST has a pending patent application related to this work. D.H. is currently employed by KMLabs, Inc. **Data and materials availability:** All data are present in the paper or supplementary materials.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6409/1358/suppl/DC1  
Materials and Methods  
Supplementary Text  
Figs. S1 to S5  
References (37–46)

23 April 2018; accepted 1 August 2018  
10.1126/science.aat6451



## ORGANIC CHEMISTRY

# Interrupted carbonyl-olefin metathesis via oxygen atom transfer

Jacob R. Ludwig<sup>1</sup>, Rebecca B. Watson<sup>1\*</sup>, Daniel J. Nasrallah<sup>1\*</sup>, Joseph B. Gianino<sup>2</sup>, Paul M. Zimmerman<sup>1</sup>, Ren A. Wiscons<sup>1</sup>, Corinna S. Schindler<sup>1†</sup>

Some of the simplest and most powerful carbon-carbon bond forming strategies take advantage of readily accessible ubiquitous motifs: carbonyls and olefins. Here we report a fundamentally distinct mode of reactivity between carbonyls and olefins that differs from established acid-catalyzed carbonyl-ene, Prins, and carbonyl-olefin metathesis reaction paths. A range of epsilon, zeta-unsaturated ketones undergo Brønsted acid-catalyzed intramolecular cyclization to provide tetrahydrofluorene products via the formation of two new carbon-carbon bonds. Theoretical calculations and accompanying mechanistic studies suggest that this carbocyclization reaction proceeds through the intermediacy of a transient oxetane formed by oxygen atom transfer. The complex polycyclic frameworks in this product class appear as common substructures in organic materials, bioactive natural products, and recently developed pharmaceuticals.

Carbonyl and olefin functionalities generally react together through carbonyl-ene (1–5) or Prins (5–10) pathways upon activation with strong Brønsted or Lewis acids. Both pathways can be accessed from similar substrates **1**, depending on the choice of catalyst and conditions (Fig. 1A). Upon coordination to a carbonyl, Lewis acids can induce a change in polarization that enhances reactivity. Different Lewis acids can subtly alter the charge distribution in the resulting Lewis acid–substrate complex (11–14) to generate a continuum between a stepwise mechanism involving carbocationic intermediate **2** and a concerted mechanism via a six-membered transition state **3** (I, Fig. 1A) (3). The carbonyl-ene reaction can proceed through either route to generate homoallylic alcohol **4** upon addition of an electrophilic carbonyl to an alkene with concomitant transfer of an allylic hydrogen atom. In comparison, the Prins reaction proceeds through intermediate carbocation **2** that is subsequently captured by an exogenous nucleophile to provide the corresponding alcohol **5** (II, Fig. 1A) (5). In a further modulation of the reactivity between carbonyls and olefins, we have recently reported an iron(III)-catalyzed carbonyl-olefin ring-closing metathesis reaction of substrate **1** to provide cyclohexene product **7** (III, Fig. 1A) (15). Mechanistic studies supported a concerted, asynchronous [2+2]-cycloaddition reaction that does not rely on carbocation intermediates to form oxetane **6**, which subsequently fragments in an asynchronous, concerted retro-[2+2]-cycloaddition to provide the corresponding cyclopentene and -hexene metathesis products (16). In this context, Lewis acid activation opens access to intermediate oxetanes that are otherwise restricted to photochemical

[2+2] cycloadditions such as the Paternò-Büchi reaction (17–19).

We herein report that these fundamental acid-mediated transformations between carbonyls and olefins can be expanded to include an additional mode of reactivity (IV, Fig. 1A). This transformation resembles the carbonyl-olefin metathesis reaction in that it also proceeds via an oxetane (**8**); however, its fragmentation pathway is interrupted to result in the formation of an intermediate carbocation. This interrupted carbonyl-olefin metathesis path relies on Brønsted acid activation of carbonyl and olefin functionalities to yield a complex, carbocyclic framework **9** (20, 21) upon formation of two carbon-carbon bonds. Whereas this multiple bond-forming process selectively yields the tetrahydrofluorene product **9**, the presence of a Brønsted acid has previously been reported to be detrimental to many Lewis acid-catalyzed carbonyl-ene and Prins reactions, resulting in undesired polymerization and isomerization of olefins (22). Furthermore, this interrupted carbonyl-olefin metathesis reaction provides access to tetrahydrofluorene product **9** in a single synthetic transformation, whereas current strategies rely on multistep sequences and precious metal-catalyzed cycloisomerizations (23, 24). Tetrahydrofluorenes represent key structural elements in materials science (**10**) (25) and are ubiquitous core structures in biologically active natural products (**11**) (26). Additionally, tetrahydrofluorenes are found in recently developed pharmaceuticals [Merck's ERβ agonist **12** (27) and ledipasvir **13** (28), part of Harvoni, Gilead's two-component treatment against hepatitis C].

During our investigations into the iron(III)-catalyzed carbonyl-olefin metathesis reaction, we discovered a complementary mode of reactivity of aryl ketone **14**, depending on the choice of iron(III) catalyst (Fig. 1B). Specifically, when **14** was converted under the optimal reaction conditions developed for carbonyl-olefin metathesis using 5 mole % (mol %) iron(III) chloride (FeCl<sub>3</sub>)

in dichloroethane, the desired metathesis product **16** was obtained in 71% yield. We initially postulated that iron(III) triflate [Fe(OTf)<sub>3</sub>] could function as a stronger Lewis acid catalyst to further improve the yield of the carbonyl-olefin metathesis product **16**. However, subjecting the same aryl ketone **14** to 10 mol % Fe(OTf)<sub>3</sub> resulted in the formation of a new product, in 55% yield, that was identified as tetrahydrofluorene **17**. We hypothesized that this new compound likely arose from an intermediate carbocation that subsequently underwent Friedel-Crafts alkylation with the pendant aromatic ring to form the tricyclic core. This outcome is in stark contrast to the reactivity observed in FeCl<sub>3</sub>-catalyzed carbonyl-olefin metathesis, which proceeds via an asynchronous, concerted mechanism and does not involve carbocations as essential intermediates. As such, this result indicated that altering the catalytic system, which we ultimately discovered to be Brønsted acid catalyzed, provides access to a distinct reaction pathway relying on carbocation intermediates. The combined importance of the tetrahydrofluorene products obtained along with the distinct reactivity observed upon Brønsted acid catalysis of aryl ketone **14** prompted us to further optimize this reaction, explore the substrate scope, and investigate the reaction mechanism of this transformation.

Thorough optimization involved the evaluation of various Lewis and Brønsted acids, solvents, and reaction times and ultimately led to the following optimal reaction conditions: 5 mol % triflic acid (TfOH) at 80°C in degassed benzene (see supplementary materials for details).

Initial efforts to explore the scope of this transformation focused on investigating the effect of varying the olefin substitution (Fig. 2A). The trisubstituted olefins **18** and **19** provided the corresponding tetrahydrofluorene product in 85 and 86% yield, respectively. Phenyl-substituted olefin **20** and exocyclic olefins **21** to **23** proved to be viable substrates, with the latter leading to spirocyclic tetrahydrofluorene products. Specifically, piperidine **23** bearing a para-trifluoromethyl phenylsulfonamide underwent efficient cyclization in 87% yield, demonstrating the potential for the incorporation of heteroatoms into the polycyclic scaffold. Aryl ketones **24** and **25**, incorporating either a 1,2-di-substituted olefin or a 1,1-di-chlorinated olefin, failed to undergo the desired transformation, suggesting that two carbon-based substituents are required to increase the nucleophilicity of the olefin.

The optimized reaction conditions for the formation of tetrahydrofluorenes proved general for various electronically and sterically differentiated aryl ketone substrates (Fig. 2B). Electron-rich aryl ketones bearing hydroxyl, methoxy, or dioxole functionalities underwent the desired transformation in good to excellent yields (**17**, **26**, **27**, **31**, **32**, **36**, and **42**, Fig. 2B). Specifically, di- and trimethoxy-substituted aryl ketones **26** and **42** provided the desired products in 78 and 90% yield, respectively. A silyl-protected phenol underwent the desired cyclization with advantageous in situ deprotection to result in 92%

<sup>1</sup>Department of Chemistry, University of Michigan, Ann Arbor, MI, USA. <sup>2</sup>Advanced Materials and Systems Research, BASF, Wyandotte, MI, USA.

\*These authors contributed equally to this work.

†Corresponding author. Email: corinnas@umich.edu

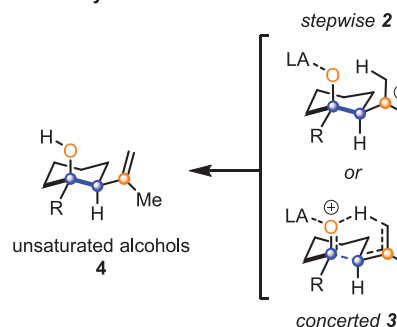
yield of **27**; subsection of the corresponding unprotected phenol also afforded efficient conversion to the desired tetrahydrofluorene **27** in 71% yield. Aryl ketones bearing electron-neutral *tert*-butyl, methyl, naphthyl, or phenyl substitution at the aromatic moiety provided the corresponding products in up to 94% yield (**28**, **37**, **43**, and **46**, Fig. 2B). Electron-deficient aryl ketones bearing chlorine or fluorine substitution reacted to form

**41** and **47** in 66 and 88% yield, respectively. The para position of the aromatic subunit was thoroughly investigated and found to be widely electronically tolerant (**27** to **29**, **32**, **33**, **35**, **37**, **39** to **41**, **44**, **46**, and **47**, Fig. 2B). Similarly, substitution at the meta and ortho positions of the aromatic subunit led to the desired tetrahydrofluorene products; however, in the case of meta-substituted substrates, a mixture of regioisomers

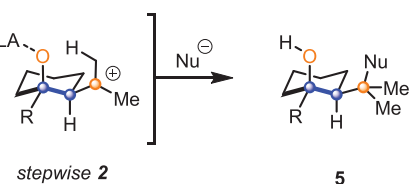
**34** and **36** was observed. The optimized reaction conditions also proved efficient for heteroaromatic ketones, resulting in the formation of benzothiophene **30** in 73% yield and thiophene **38** in 57% yield. Aryl ketones that contained heteroatoms distal from the reactive sites proved viable substrates for the desired transformation, resulting in the corresponding tetrahydrofluorene products in up to 59% yield (**29**, **33**, and **35**, Fig. 2B).

## A Reactivity Modes of Carbonyls and Olefins Under Acidic Conditions

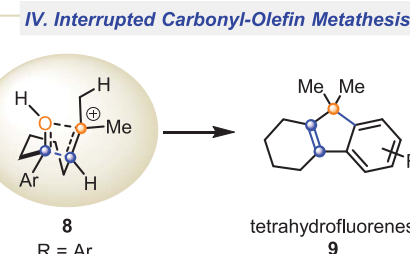
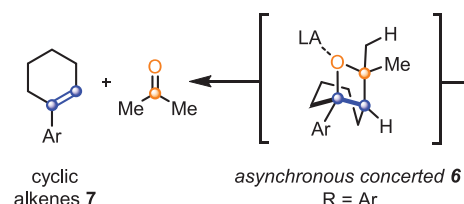
### I. Carbonyl-Ene Reaction



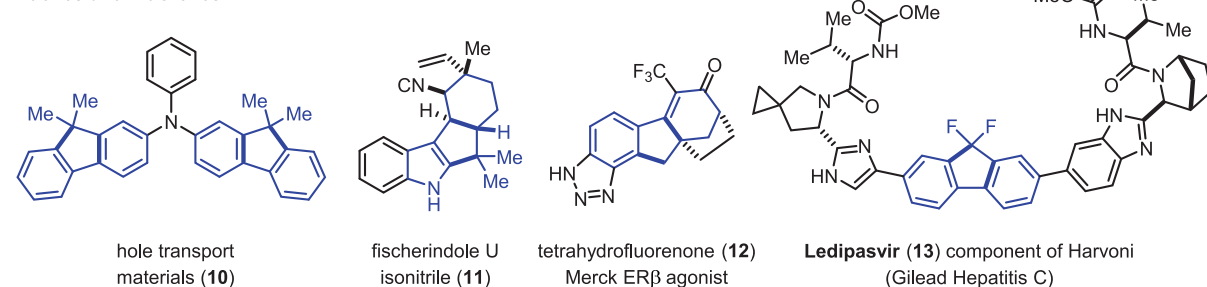
### II. Prins Reaction



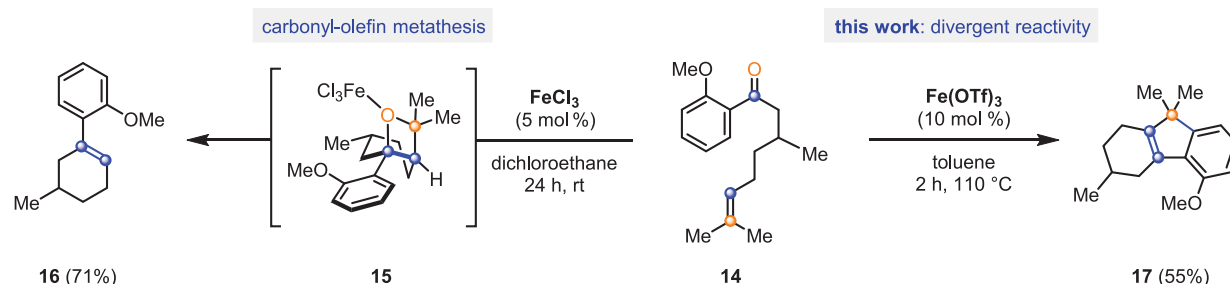
### III. Carbonyl-Olefin Metathesis



### Indenes and Fluorenes

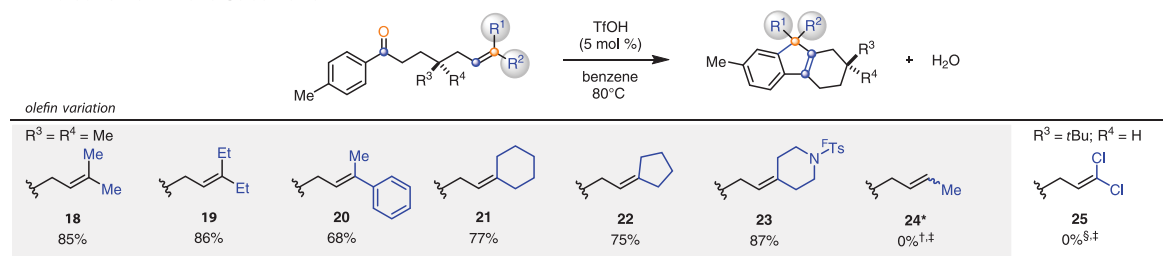


## B Divergent Reactivity of Aryl Ketones Under Lewis Acid and Brønsted Acid Catalysis

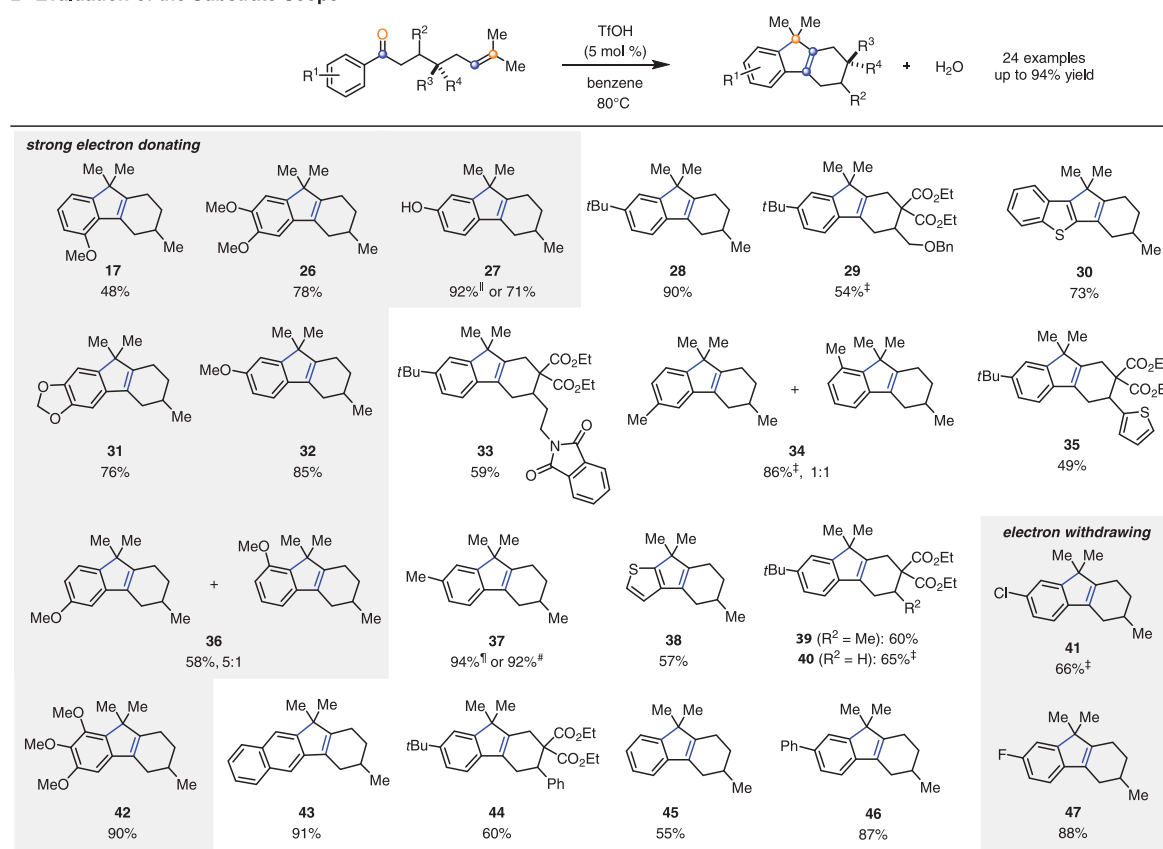
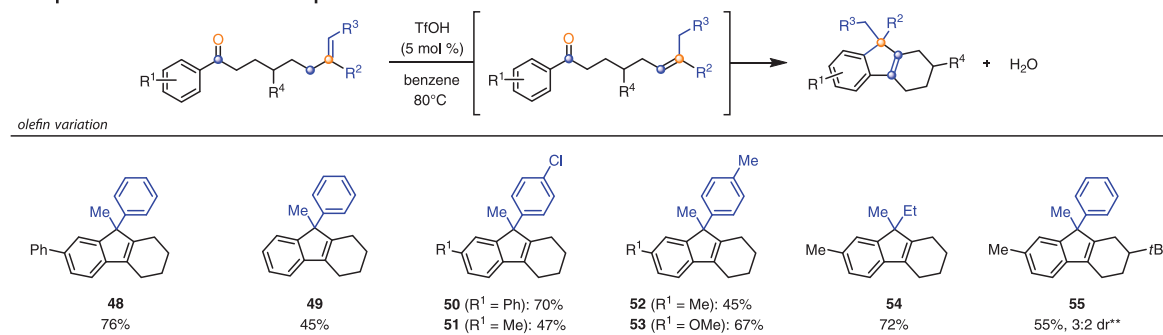


**Fig. 1. Chemistry of carbonyls and olefins.** (A) Fundamental acid-catalyzed reactivity modes between carbonyls and olefins. (I) Carbonyl-ene reaction. (II) Prins reaction. (III) Carbonyl-olefin metathesis.; (IV) This work: interrupted carbonyl-olefin metathesis. The importance of the products accessible in this reactivity mode is showcased below. LA, Lewis acid; R, hydrogen, alkyl, or aryl; Me, methyl; Nu, nucleophile; Ar, aryl. (B) Complementary reactivity modes of aryl ketone **14** are accessible depending on the choice of iron(III)-derived Lewis acid catalyst. The use of  $\text{FeCl}_3$  leads to the formation of the carbonyl-olefin metathesis product **16**, whereas  $\text{Fe}(\text{OTf})_3$  results in tetrahydrofluorene **17**. rt, room temperature.

## A Evaluation of Alkene Substitution



## B Evaluation of the Substrate Scope

C Expansion of the Substrate Scope via *in situ* Isomerization of Alkenes

**Fig. 2. Substrate scope.** Conditions: Ketone (1.0 equivalent; 0.02 M), TfOH (5 mol %) in benzene at 80°C. See supplementary materials for additional experimental details. **(A)** Variation of the alkene substitution. Et, ethyl; <sup>†</sup>Ts, (4-(trifluoromethyl)phenyl)sulfonyl; tBu, *tert*-butyl. **(B)** Variation of the aryl ketone and carbon tether. Bn, benzyl; Ph, phenyl. **(C)** In situ isomerization of alkenes to expand substrate scope. Footnotes: \*4.3:1 mixture of *E* and *Z* alkenes. †81% of starting material **24** is recovered.

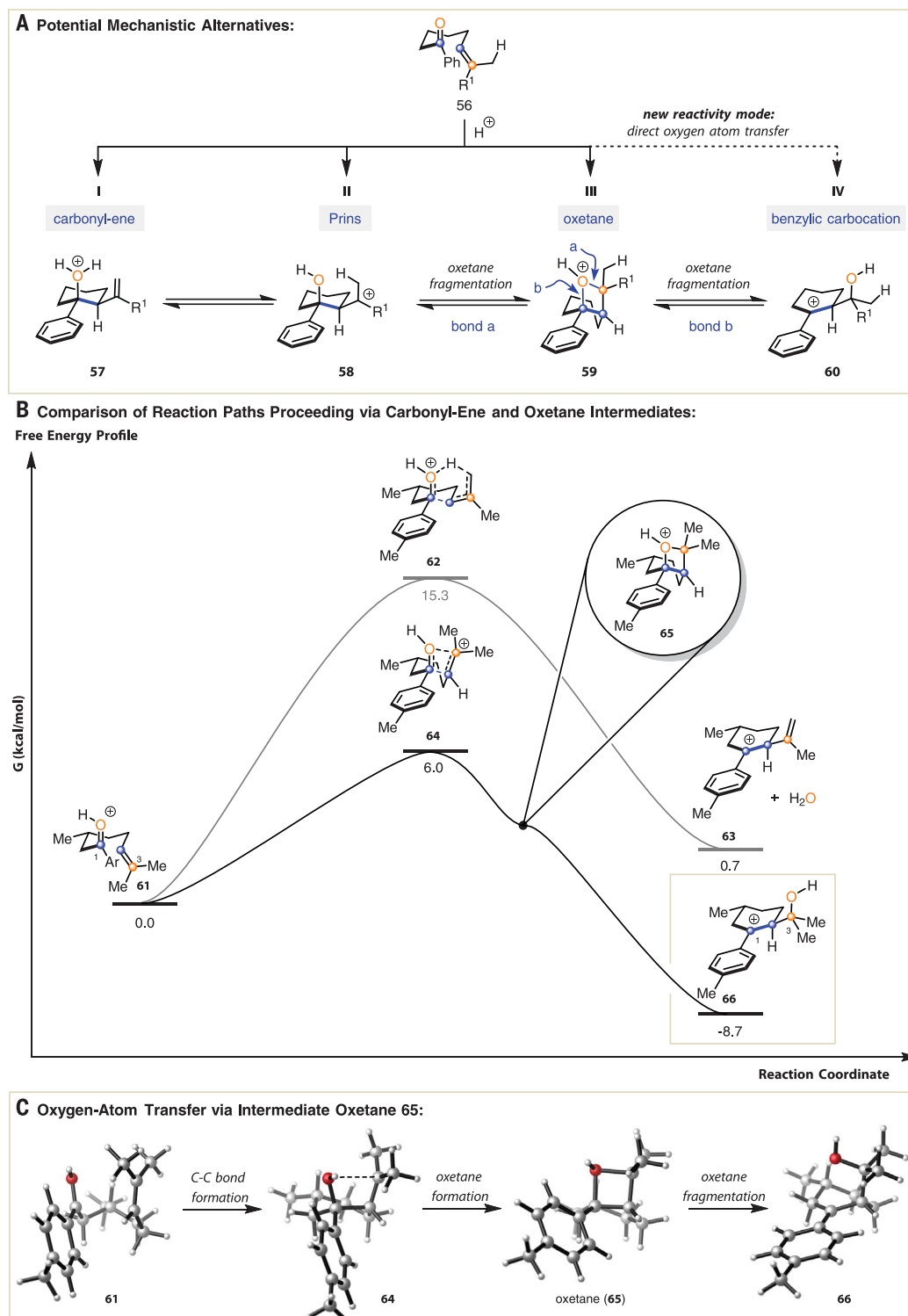
‡Percent yield and percent conversion determined by <sup>1</sup>H nuclear magnetic resonance (NMR) using dimethyl terephthalate or mesitylene as an internal standard. §98% of starting material **25** is recovered. ||Starting material TBS (*tert*-butyldimethylsilyl) protected. ¶Percent yield determined by gas chromatography using dodecane as an internal standard. #Reaction run on a 10.2-mmol scale. \*\*Percent yield and diastereomeric ratio (dr) determined by <sup>1</sup>H NMR using dimethyl terephthalate as an internal standard.



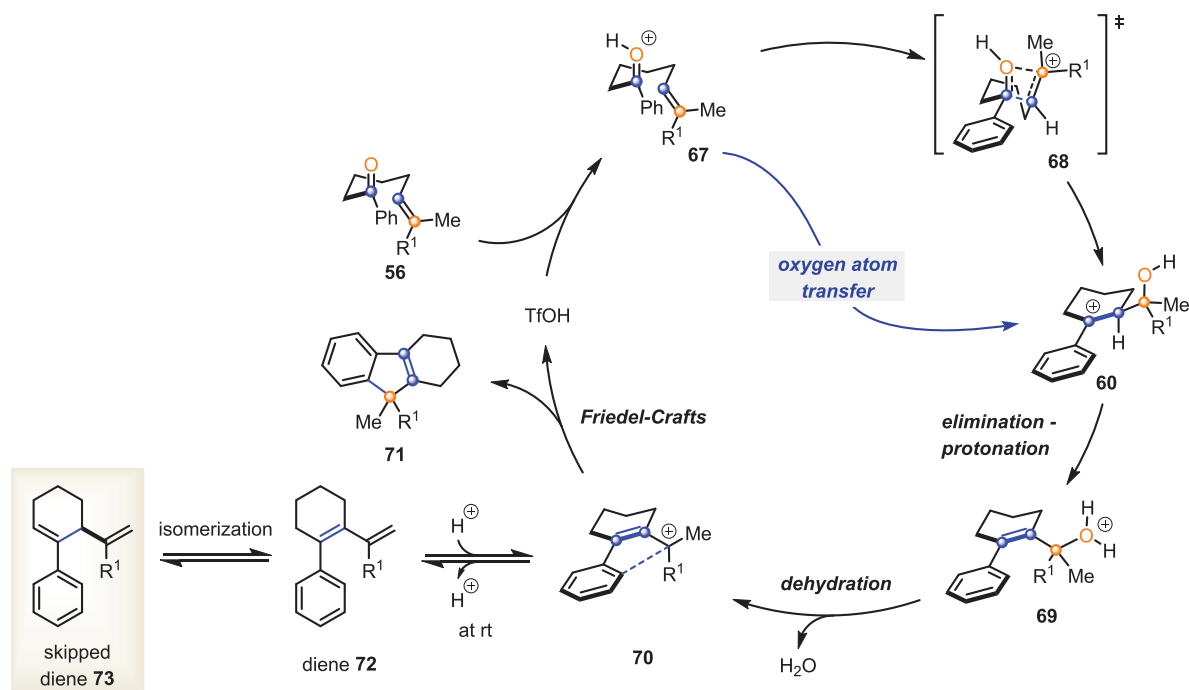
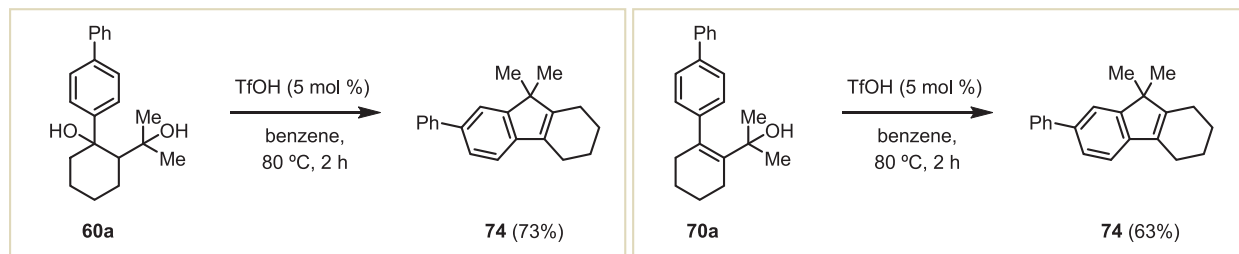
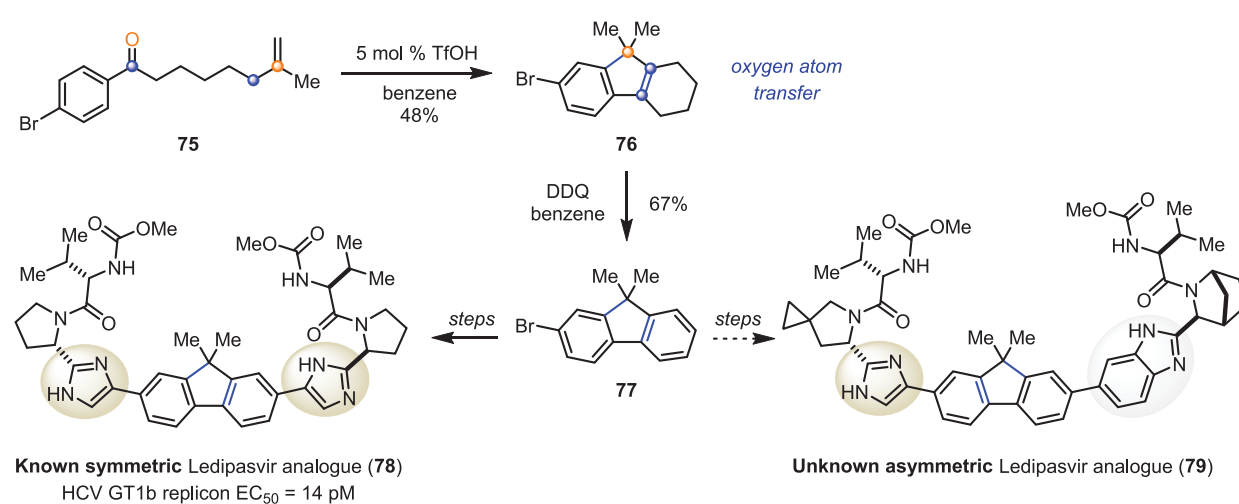
Advantageously, benzyl ether **29** and phthalimide **33** can be deprotected to provide handles for further elaboration. *Tert*-butyl-substituted aromatic substrates bearing distinct functionalities along the carbon backbone resulted in the for-

mation of the desired products **39**, **40**, and **44** in up to 65% yield, whereas the substrate bearing minimal substitution also proved viable, providing the desired tetrahydrofluorene **45** in 55% yield. Taken together, the substrate scope sug-

gests that the electronics of the aryl ketone and sterics on the substrate tether do not alter the efficiency of this transformation. The optimized reaction conditions also proved amenable to gram-scale synthesis of **37**.



**Fig. 3. Mechanistic investigations.** (A) Possible mechanistic alternatives in the Brønsted acid-catalyzed formation of tetrahydrofluorenes from aryl ketones. (B) Density functional theory studies of the reaction pathway comparing carbonyl-ene reactivity to a pathway relying on a transient oxetane. G, free energy. (C) Three-dimensional representations of intermediates and transition states.

**A Mechanistic Hypothesis for Interrupted Carbonyl-Olefin Metathesis:****B Mechanistic Probe Molecules for Carbocations 60 and 70:****C Synthesis of Intermediates Toward Ledipasvir Analogues via Interrupted Carbonyl-Olefin Metathesis:**

**Fig. 4. Proposed mechanism and application.** (A) Mechanistic hypothesis for interrupted carbonyl-olefin metathesis reaction. (B) Investigation of mechanistic probe molecules. (C) Synthesis of intermediates to access known and new ledipasvir analogs. HCV, hepatitis C virus.

Subsequent efforts focused on further expanding the scope of this transformation to include 1,1-disubstituted alkenes that are readily accessible via Wittig olefination or hydroarylation strategies (Fig. 2C). Isomerization of these alkenes in situ under Brønsted acid catalysis results in the corresponding 1,2,2-trisubstituted analogs that subsequently enable facile access to tetrahydrofluorene products bearing distinct substitution at the central five-membered ring. Specifically, aryl ketones bearing electron-rich and neutral substituents proved viable substrates in the isomerization-cyclization sequence and provided the corresponding tetrahydrofluorene products in up to 76% yield (**48** to **55**, Fig. 2C). Varying the electronics of the alkene itself was tolerated with both electron-poor (**50** and **51**) and -neutral (**52** and **53**) styrene derivatives, which underwent the desired transformation in up to 70% yield. This in situ isomerization-cyclization sequence is not only limited to terminal styrene derivatives but also tolerates the alkene-bearing aliphatic substituents (**54**). Modest diastereoselectivity (3:2 diastereomeric ratio) was observed for tetrahydrofluorene **55**, demonstrating the potential for this mode of reactivity to be used in the development of stereoselective methods.

On the basis of the literature precedent of transformations between carbonyls and olefins, we initially considered a mechanistic hypothesis relying on a carbonyl-ene reaction to form alcohol **57** upon nucleophilic addition between the carbonyl and olefin functionalities of **56** (Fig. 3A). However, this initial mechanistic hypothesis proved inconsistent with experimental data (supplementary materials), and other potential mechanistic alternatives were evaluated. In addition to a concerted carbonyl-ene reaction path (I, Fig. 3A), intermediate carbocation **58** could result from a nucleophilic addition between the carbonyl and olefin moieties in **56**, in accordance with established Prins reactivity (II, Fig. 3A). A third alternative would be the formation of oxetane **59** in analogy to the recently established carbonyl-olefin metathesis reaction proceeding via the asynchronous, concerted formation of intermediate oxetanes (III, Fig. 3A). Intermediates **57**, **58**, and **59** could also interconvert in reversible transformations. Finally, Brønsted acid catalysis of aryl ketone **56** could give rise to a fourth intermediate, benzylic carbocation **60**, as the product of a direct oxygen atom transfer relying on the initial formation of oxetane **59** (IV, Fig. 3A). Unlike in carbonyl-olefin metathesis, which proceeds via a retro-[2+2]-cycloaddition of oxetane **59**, this fragmentation is interrupted to result in carbocation **60** (fragmentation of bond b in **59**).

These distinct mechanistic scenarios were subsequently investigated computationally (unrestricted B97-D density functional and 6-31+G\* basis set) to determine the viability of their transition states and corresponding minimal-energy pathways. The quantum chemical simulations based on the growing string method (29) revealed two possible reaction paths (see supplementary materials for computational details): (i) a concerted carbonyl-ene pathway via transition

state **62**, following the initial mechanistic hypothesis (Fig. 3B), and (ii) a single-elementary step pathway passing through transition state **64** and oxetane **65** (Fig. 3B) (30). The latter path was found to have an energy barrier 9.3 kcal/mol lower than that of the carbonyl-ene reaction, which suggests that it is the preferred reaction path. This lower-energy pathway yields benzylic carbocation intermediate **66** and constitutes a direct oxygen atom transfer between two carbons (C-1 and C-3 in **66**, Fig. 3B). Electronically, conversion of **61** to **66** is enabled by an asynchronous, concerted path that is best conceptualized as two distinct transitions connected by an unstable oxetane intermediate **65**. Figure 3, B and C, highlights the asynchronous nature of this path by showing the transition state **64** as the highest-energy point, which forms oxetane **65** and subsequently fragments through an energetically favorable ring-opening to result in benzylic carbocation intermediate **66**. The second electronic change in this reaction path is barrierless, due to the instability of the protonated oxetane **65** when compared to its fragmentation product **66**.

This Brønsted acid-catalyzed mode of reactivity complements the previously established Lewis acid-catalyzed carbonyl-olefin metathesis reaction that relies on intermediate oxetanes. However, under Brønsted acid catalysis, fragmentation of the transient oxetane interrupts the carbonyl-olefin metathesis pathway and results in a new reactive intermediate, benzylic carbocation **66**. As such, the transience of **65** suggests a direct oxygen atom transfer that represents a distinct reactivity mode between carbonyls and olefins to provide benzylic carbocations (Fig. 3A, pathway IV).

Taking into account the experimental and computational results obtained, we propose the following reaction mechanism for the Brønsted acid-catalyzed interrupted carbonyl-olefin metathesis reaction (Fig. 4A). Protonation of aryl ketone **56** initiates intramolecular oxygen atom transfer via transition state **68** to form intermediate benzylic carbocation **60**. Elimination and subsequent protonation of the resulting allylic alcohol provides **69**, which can then undergo dehydration to produce carbocation **70**. This highly stabilized allylic carbocation undergoes a final Friedel-Crafts alkylation to form the tetrahydrofluorene product **71**. This hypothesis was subsequently tested by the independent synthesis of two probe molecules—specifically, tertiary alcohols **60a** and **70a** (Fig. 4B). Diol **60a** and allylic alcohol **70a** are both able to undergo a Friedel-Crafts alkylation to provide tetrahydrofluorene product **74** upon treatment with TfOH, which supports carbocations **60** and **70** as potential intermediates. However, at lower reaction temperatures, Friedel-Crafts alkylation does not proceed and carbocation **70** is quenched via elimination to result in diene **72**. Further isomerization of diene **72** provides an experimentally observed skipped diene **73** as a shunt product (supplementary materials). Alternative pathways for the formation of skipped diene **73** were investigated computationally but were found to be higher in energy. Upon exposure to the optimized

reaction conditions, skipped diene **73** reengages in the reaction pathway to give rise to the tetrahydrofluorene product exclusively (see supplementary materials for experimental details).

The tetrahydrofluorene products obtained in our one-step, multiple bond-forming transformation can be readily oxidized to the corresponding fluorene compounds in up to 99% yield using DDQ (2,3-dichloro-5,6-dicyano-1,4-benzoquinone) (supplementary materials). The synthetic value of this cyclization-oxidation sequence has been demonstrated in the synthesis of a key fluorene intermediate toward a biologically active ledipasvir analog (Fig. 4C) (28). Importantly, the interrupted carbonyl-olefin metathesis reaction enables rapid entry to aromatic fluorene moieties bearing distinctive substitution patterns that are difficult to access with currently available synthetic methods. Specifically, symmetric and asymmetric analogs are accessible using the same fluorene core. Under the optimized reaction conditions, aryl ketone **75** yields tetrahydrofluorene **76**, which upon subsequent oxidation results in fluorene **77**. This intermediate (**77**) can be further advanced to known symmetric ledipasvir derivative **78**, with a hepatitis C virus GT1b replicon EC<sub>50</sub> (median effective concentration) value of 14 pM, or an unknown asymmetric analog **79** (28).

The developed interrupted carbonyl-olefin metathesis reaction complements the repertoire of well-established reactions between carbonyls and olefins and provides entry into the formation of complex, polycyclic tetrahydrofluorenes in a single synthetic step relying on TfOH as an inexpensive catalyst.

## REFERENCES AND NOTES

- K. Alder, F. Pascher, A. Schmitz, *Br. Dtsch. Chem. Ges. B. Ser.* **76**, 27–53 (1943).
- M. L. Clarke, M. B. France, *Tetrahedron* **64**, 9003–9031 (2008).
- K. Mikami, M. Shimizu, *Chem. Rev.* **92**, 1021–1050 (1992).
- C.-Y. Ho, K. D. Schleicher, T. F. Jamison, *Synlett* **2009**, 2565–2582 (2009).
- B. B. Snider, in *Comprehensive Organic Synthesis*, B. M. Trost, I. Fleming, Eds. (Pergamon, 1991), vol. 2, pp. 527–562.
- O. Kriewitz, *Br. Dtsch. Chem. Ges.* **32**, 57–60 (1899).
- H. J. Prins, *Chem. Weekblad* **16**, 1510–1526 (1919).
- E. Arundale, L. A. Mikeska, *Chem. Rev.* **51**, 505–555 (1952).
- C. Olier, M. Kaafarani, S. Gastaldi, M. P. Bertrand, *Tetrahedron* **66**, 413–445 (2010).
- I. M. Pastor, M. Yus, *Curr. Org. Chem.* **11**, 925–957 (2007).
- L. M. Stephenson, M. Orfanopoulos, *J. Org. Chem.* **46**, 2200–2201 (1981).
- H. Kwart, M. Brechbiel, *J. Org. Chem.* **47**, 5409–5411 (1982).
- B. B. Snider, E. Ron, *J. Am. Chem. Soc.* **107**, 8160–8164 (1985).
- D. A. Singleton, C. Hang, *J. Org. Chem.* **65**, 895–899 (2000).
- J. R. Ludwig, P. M. Zimmerman, J. B. Gianino, C. S. Schindler, *Nature* **533**, 374–379 (2016).
- J. R. Ludwig et al., *J. Am. Chem. Soc.* **139**, 10832–10842 (2017).
- E. Paternò, *Gazz. Chim. Ital.* **39**, 237–250 (1909).
- G. Büchi, C. G. Inman, E. S. Lipinsky, *J. Am. Chem. Soc.* **76**, 4327–4331 (1954).
- J. A. Porco Jr., S. L. Schreiber, in *Comprehensive Organic Synthesis*, B. M. Trost, I. Fleming, Eds. (Pergamon, 1991), vol. 5, pp. 151–192.
- The formation of a related tetrahydrofluorene has previously been described as a by-product in 38% yield using catalytic



- Bi(OTf)<sub>3</sub>. However, the reaction was assumed to proceed following a carbonyl-ene reaction path and subsequent Friedel-Crafts alkylation. See (21).
21. P. Tremel *et al.*, *New J. Chem.* **39**, 7453–7458 (2015).
  22. B. B. Snider, *Acc. Chem. Res.* **13**, 426–432 (1980).
  23. G. Lemi re *et al.*, *Angew. Chem. Int. Ed.* **45**, 7596–7599 (2006).
  24. S. M. Wilkerson-Hill, C. M. Lavados, R. Sarpong, *Tetrahedron* **72**, 3635–3640 (2016).
  25. A.-H. Zhou, F. Pan, C. Zhu, L.-W. Ye, *Chemistry* **21**, 10278–10288 (2015).
  26. Z. Lu, M. Yang, P. Chen, X. Xiong, A. Li, *Angew. Chem. Int. Ed.* **53**, 13840–13844 (2014).
  27. M. L. Maddess *et al.*, *Org. Process Res. Dev.* **18**, 528–538 (2014).
  28. J. O. Link *et al.*, *J. Med. Chem.* **57**, 2033–2046 (2014).
  29. P. M. Zimmerman, *J. Comput. Chem.* **36**, 601–611 (2015).
  30. Attempts to probe the Prins reaction pathway computationally did not support a tertiary carbocation but instead implicated benzylic carbocation **66** via a transition state resembling an oxetane (**64**).

#### ACKNOWLEDGMENTS

**Funding:** This work was supported by the NIH National Institute of General Medical Sciences (R01-GM118644), the David and Lucile Packard Foundation, and the Alfred P. Sloan Foundation (fellowships to C.S.S.). J.R.L. and R.B.W. thank the National Science Foundation for predoctoral fellowships. P.M.Z. thanks the National Science Foundation (CHE-1551994). **Author contributions:** J.R.L., R.B.W., D.J.N., and C.S.S. devised the experiments and prepared the manuscript for publication. D.J.N. conducted the theoretical investigations. J.B.G. performed initial experiments, P.M.Z. provided guidance in theoretical investigations, and R.A.W. conducted x-ray crystallographic

studies. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The supplementary materials contain complete experimental and spectral details for all new compounds and all reactions reported herein. Crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre under reference CCDC-1584828.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6409/1363/suppl/DC1](http://www.sciencemag.org/content/361/6409/1363/suppl/DC1)  
 Materials and Methods  
 Figs. S1 to S7  
 Tables S1 to S4  
 References (31–64)  
 NMR Spectra  
 20 December 2017; accepted 25 July 2018  
 10.1126/science.aar8238

## ORGANIC CHEMISTRY

# Arylsulfonylacetamides as bifunctional reagents for alkene aminoarylation

Timothy M. Monos\*, Rory C. McAtee\*, Corey R. J. Stephenson†

Alkene aminoarylation with a single, bifunctional reagent is a concise synthetic strategy. We report a catalytic protocol for the addition of arylsulfonylacetamides across electron-rich alkenes with complete anti-Markovnikov regioselectivity and excellent diastereoselectivity to provide 2,2-diarylethylamines. In this process, single-electron alkene oxidation enables carbon-nitrogen bond formation to provide a key benzylic radical poised for a Smiles-Truce 1,5-aryl shift. This reaction is redox-neutral, exhibits broad functional group compatibility, and occurs at room temperature with loss of sulfur dioxide. As this process is driven by visible light, uses readily available starting materials, and demonstrates convergent synthesis, it is well suited for use in a variety of synthetic endeavors.

The arylolethylamine motif is conserved in dopamine, serotonin, and many opioid receptor drugs responsible for modulating pain sensation and treating neurobehavioral disorders (Fig. 1A) (1, 2). In light of the opioid epidemic, the climate surrounding opioid pain medications is conflicted. It is noteworthy that frontline medications treating opioid addiction contain such arylolethylamine substructures (naltrexone and buprenorphine) (3–5). With this rationale, continued drug development in the arylolethylamine chemical space is necessary for general hit-to-lead exploration and the discovery of new and safer medicines. Conventional methods to synthesize arylolethylamines use multistep homologation and reductive amination sequences. Alternatively, alkene aminoarylation, particularly of anethole and other biomass-derived alkenes, allows for direct access to this medicinally desirable functionality. The development of methodologies to rapidly construct two new bonds (C–C and C–N) in a single operation from feedstock chemicals can improve and expedite the discovery of new arylolethylamine-based small-molecule therapeutics.

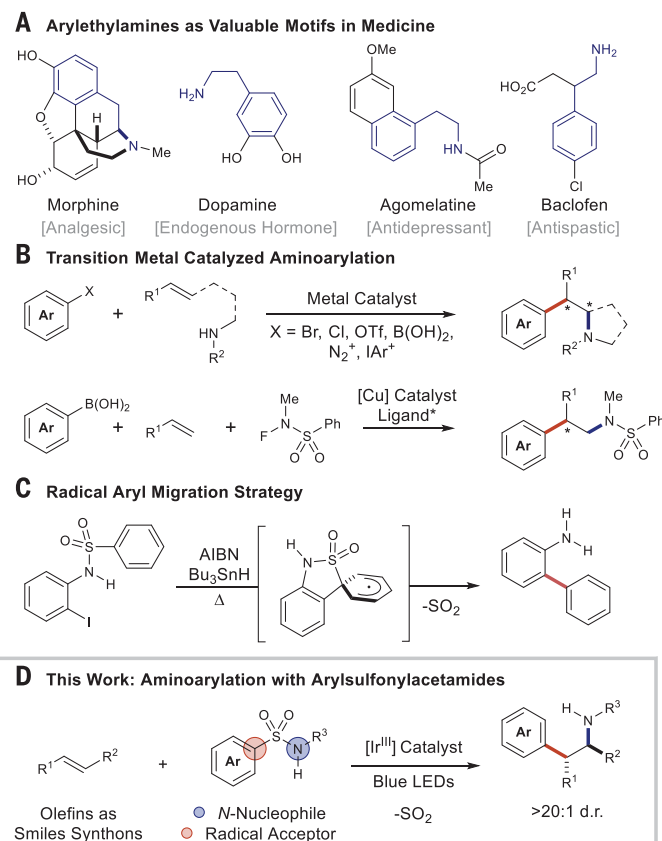
Alkene aminoarylation has been demonstrated with palladium (6, 7), copper (8–10), nickel (11, 12), and gold (13), in which alkenes are activated by the transition metal to facilitate a stereoselective amine cyclization, followed by a two-electron metal-mediated arylation event (Fig. 1B). The metals used in these aminoarylation platforms control stereoselectivity and activate the alkene for reactivity while suppressing protodemetalation or  $\beta$ -hydride elimination pathways that hinder desired C–C bond formation. Amides and amines are more nucleophilic than the alkene coupling partner; thus, elevated temperatures are often necessary to facilitate ligand substitution to

unite the reactants in the initial amination event (14). Despite robust investigation, these methods are generally limited by the need for directing groups and intramolecular reaction designs that restrict the products to pyrrolidine and piperidine structures. Recently, transformations effecting intermolecular aminoarylation and carboamination have been accomplished in which the alkene is decoupled from the arylation and amination reagents. In one case, Lin and Liu demonstrated an enantioselective copper(I)-catalyzed

aminoarylation of vinyl arenes relying upon preoxidized sulfonamide reactants (*N*-fluoro-*N*-methylbenzenesulfonamide) (Fig. 1B) (9). Separately, Rovis and Piou demonstrated an intermolecular carboamination using *N*-enoxyphthalimides and Rh(III) catalysis (15).

Photocatalysis and radical-based chemistry have proven similarly influential in alkene difunctionalization. The simplest strategy is Meerwein aminoarylation, a Markovnikov-selective reaction that begins with the reductive generation of a radical from a suitable precursor (arene diazonium salt or diaryliodonium salt) followed by radical-polar crossover and carbocation trapping with acetonitrile solvent (16). These reactions are regioselective but are devoid of stereoselectivity. Photocatalytic *anti*-Markovnikov-selective alkene hydro- and carboamination reactions were recently demonstrated by Knowles (17–21) and Nicewicz (22–24). These approaches represent contrasting C–N bond formation strategies while using a common catalytic cycle. Knowles and co-workers have demonstrated both aminium radical cation and amidyl radical generation for the addition to olefins. In both cases, nitrogen-centered radicals couple with  $\pi$ -systems to generate  $\beta$ -amino radicals that are rapidly trapped with an H-atom transfer reagent. Successful H-atom transfer reagents are minimally nucleophilic to prevent thiolene reactivity. Nitrogen radical-based chemistry is particularly challenging because both alkene addition and allylic H-atom abstraction are kinetically competitive processes (25); thus, success often requires excesses of the alkene component

**Fig. 1. Strategies to access arylolethylamines.** (A) Arylolethylamines as valuable motifs. (B) Current approaches toward arylolethylamines with transition metal catalysis. (C) The Smiles-Truce rearrangement as an aryl migration strategy. (D) The method proposed herein. Me, methyl; Ph, phenyl; AIBN, azobisisobutyronitrile; HSnBu<sub>3</sub>, tributyltin hydride.

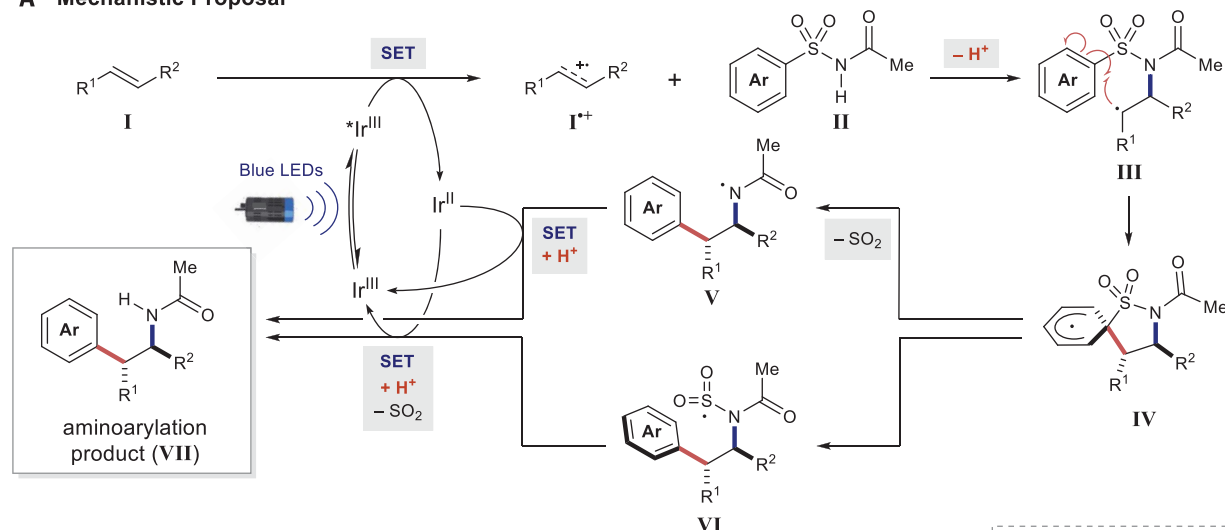


Department of Chemistry, University of Michigan, Ann Arbor, MI, USA.

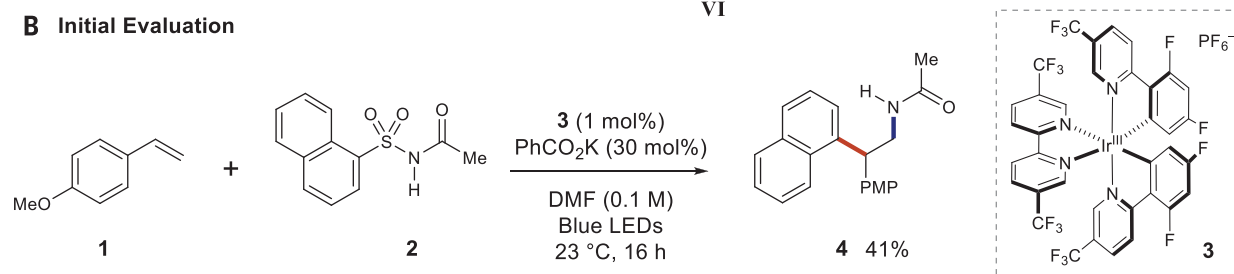
\*These authors contributed equally to this work.

†Corresponding author. Email: crjsteph@umich.edu

## A Mechanistic Proposal



## B Initial Evaluation



**Fig. 2.** Proposed reaction design for aminoarylation with arylsulfonylacetyl amides. (A) Proposed reaction mechanism. (B) Initial reaction evaluation. PMP, *p*-methoxyphenyl.

or intramolecular amino-cyclization. Additionally, amine and amide oxidation generate a more reactive, but not a more nucleophilic, nitrogen atom. In contrast, Nicewicz and co-workers have targeted alkene single-electron oxidation, a process approximately as rapid as amide or amine oxidation. This approach benefits from converting the alkene to a more electrophilic species in solution, necessitating lower equivalents of the nitrogen nucleophile to conduct alkene difunctionalization.

To contrast the widely investigated field of transition metal-mediated aminoarylation and build on the successes of photocatalytic alkene difunctionalization chemistry, we were inspired by the possibility of a radical Smiles-Truce rearrangement to provide alkene aminoarylation products in a diastereoselective fashion. Traditionally, the Truce variant of the Smiles rearrangement is a nucleophilic aromatic substitution effected by benzylic lithiation of *ortho*-tolyl-arylsulfones (26). The rearrangement is more broadly applicable to *ipso*-substitution reactions with aryl sulfides, sulfoxides, sulfones, and amides. Pennell and Motherwell furthered the utility of this transformation by demonstrating that aryl radicals are also capable of the same arene transposition (27) (Fig. 1C). Although there are numerous intramolecular examples of radical Smiles-Truce reactions (28–33), many of these reactions use net reductive conditions, generate a stoichiometric

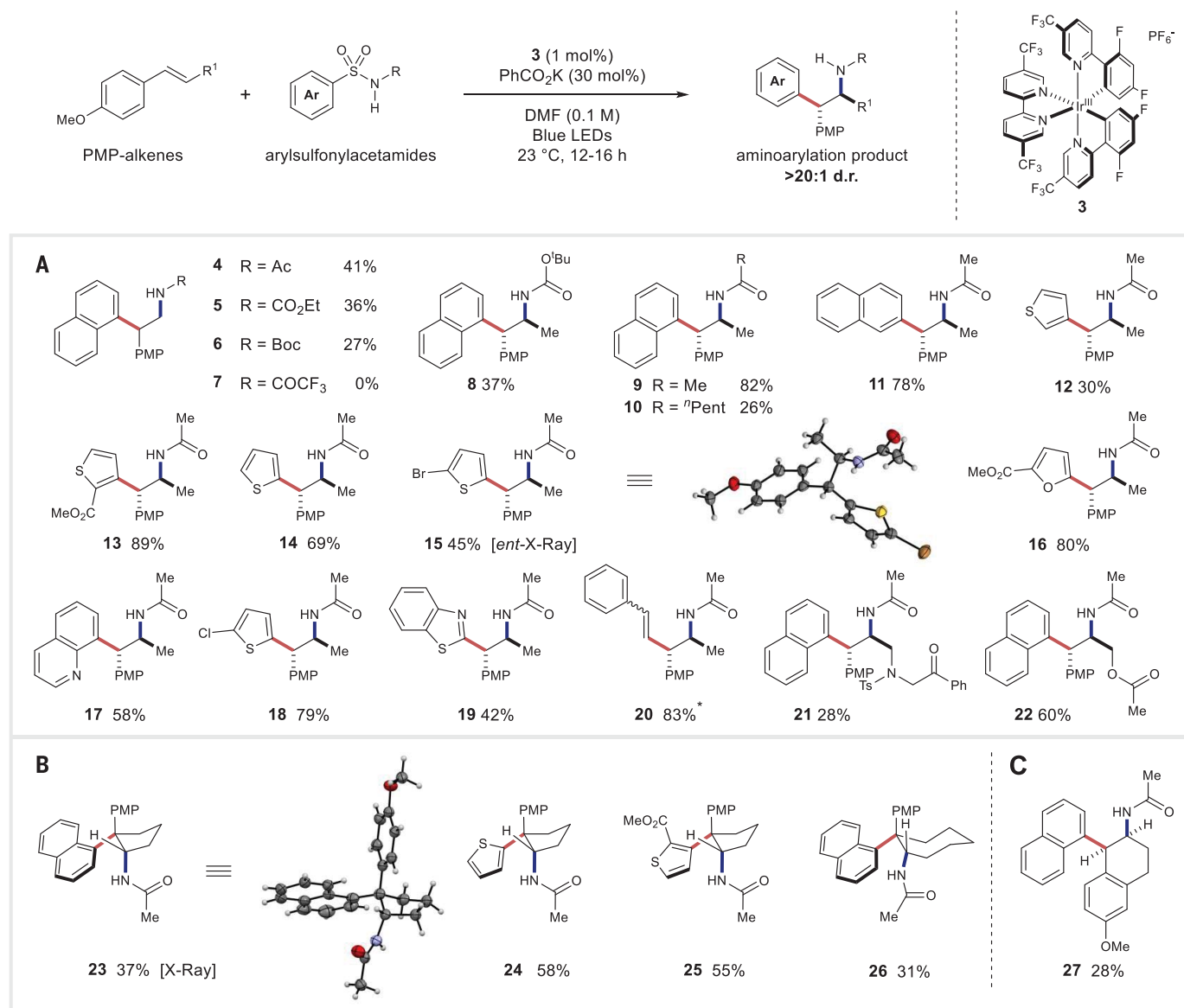
amount of waste, and rely on a substrate design that tethers the radical precursor to the aryl-sulfonate derivative. Realizing that this intramolecular tether can be formed via *in situ* oxidation of an alkene and subsequent nucleophilic trapping with an arylsulfonylacetyl amide (34), we sought to design a photocatalyzed radical Smiles-Truce reaction that showcases the utility of arylsulfonylacetyl amides as capable reagents for both C–N and C–C bond formation in aminoarylation (Fig. 1D).

A general catalytic cycle was postulated to begin with an oxidation event between a photoexcited catalyst ( $^*\text{Ir}^{\text{III}}$ ) and an alkene (I) (Fig. 2A) (23, 24). Single-electron oxidation of the alkene would enable nucleophilic addition of an arylsulfonylacetyl amide (II) to afford the desired  $\beta$ -aminoalkyl radical intermediate (III) (35–37). This radical is poised for regioselective cyclization onto the *ipso*-position of the appended arene to generate IV (38). Lastly, an entropically favored desulfonylation can proceed via two plausible pathways to generate the aminoarylation product, VII: (i) rapid radical desulfonylation from IV to generate nitrogen-centered radical V followed by catalyst turnover, or (ii) homolytic fragmentation of the  $\text{C}_{\text{Ar}}\text{--S}$  bond to furnish VI, which can turn over the catalyst and undergo desulfonylation to VII. Exploiting both the electronic activation of the sulfonylated arene unit and the tunable nucleophilicity of the nitrogen motif allows for this

photoredox catalysis platform to promote both the C–N and C–C bond-forming events with arylsulfonylacetyl amides.

To realize the proposed aminoarylation reactivity, we first conducted reaction optimization with vinyl anisole (1) [ $E_{\text{p}/2} = 1.6$  V versus saturated calomel electrode (SCE)] (39) and 1-naphthylsulfonylacetyl amide (2) (table S1). A potent photooxidant,  $[\text{Ir}(\text{dF}(\text{CF}_3)\text{ppy})_2](5,5'\text{-CF}_3\text{-bpy})\text{PF}_6$  (3) ( $\text{Ir}^{\text{III}/\text{II}} = 1.68$  V versus SCE in MeCN) (40) was initially selected for alkene radical cation formation (Fig. 2B). Early optimization experiments lent evidence to the chemoselectivity of this reaction; excess loading of arylsulfonylacetyl amide and base were unnecessary (table S1). Nearly equivalent stoichiometry between 1 and 2 afforded the highest yield for the optimization product 4. A base screen revealed potassium acetate, benzoate, and tribasic phosphate as superior bases to the less basic potassium trifluoroacetate and potassium phosphate (mono- or dibasic). The reaction was incompatible with pyridine or with stronger alkoxide bases, as photocatalyst decomposition was observed. Reaction dilution past 0.1 M slowed the rate of product formation, whereas reaction concentrations greater than 0.1 M inhibited product formation. Further optimization proved that less oxidizing photocatalysts such as  $[\text{Ru}(\text{bpy})_3]\text{Cl}_2$  ( $\text{Ru}^{\text{II}/\text{I}} = 0.77$  V versus SCE in MeCN),  $[\text{Ir}(\text{dF}(\text{CF}_3)\text{ppy})_2](\text{dtbbpy})\text{PF}_6$  ( $\text{Ir}^{\text{III}/\text{II}} = 0.89$  V versus SCE in MeCN),





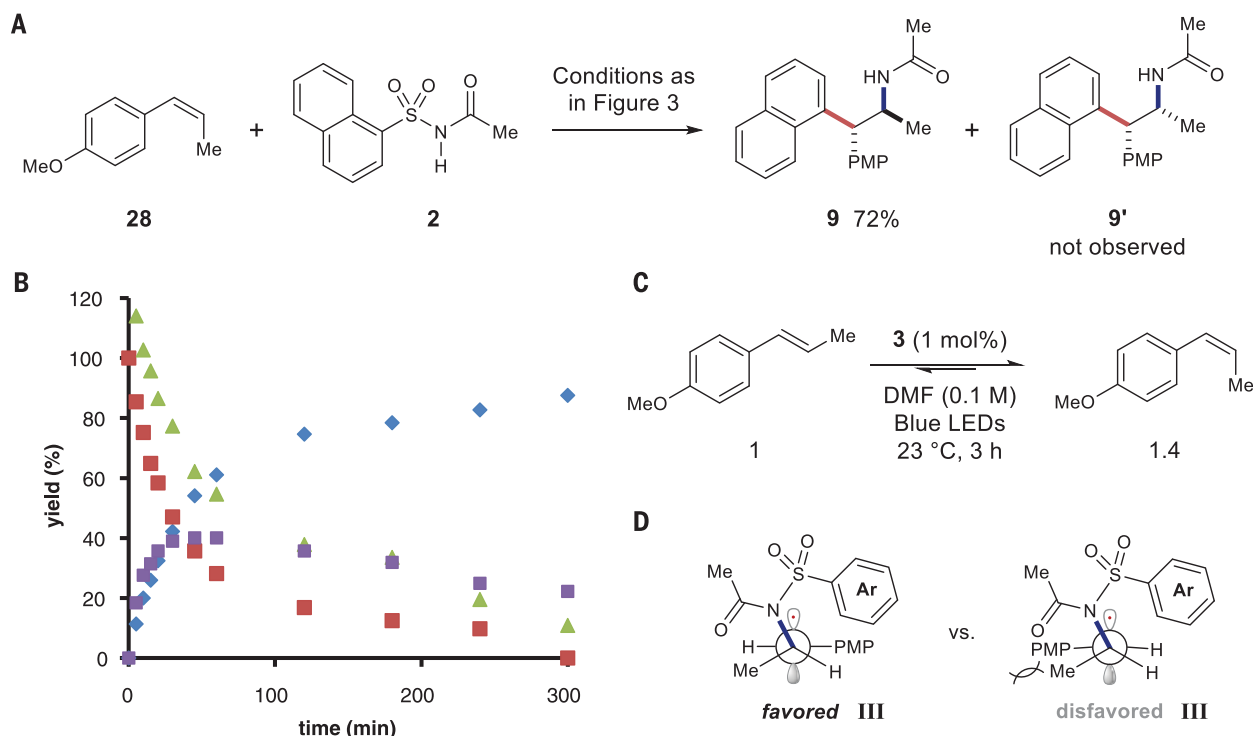
**Fig. 3. Exploration of substrate scope.** All yields are isolated yields. Relative configurations of products were assigned by analogy to **15** and **23**. (A) Evaluation of scope of aryl group. (B) Scope of cyclic *trans*-PMP alkenes. (C) Scope of a compatible cyclic *cis*-PMP alkene. Ac, acetyl; Boc, *tert*-butoxycarbonyl; <sup>n</sup>Pent, *n*-pentyl; Ts, tosyl. \*2:1 mixture of *E/Z* alkene diastereomers.

$[\text{Ir}(\text{ppy})_2(\text{dtbbpy})]\text{PF}_6$  ( $\text{Ir}^{\text{III}/\text{II}} = 0.31$  V versus SCE in MeCN) (**41**), were unable to catalyze this transformation. Use of Fukuzumi's catalyst ( $\text{PC}^*/\text{PC}^* = 1.88$  V versus SCE in MeCN) (**42**) did produce **4** in 13% yield. Finally, H-atom donor additives such as 1,4-cyclohexadiene and isopropanol did not improve on the established conditions for the optimization product **4**. Exclusion of either light or photocatalyst failed to promote aminoarylation (table S1). With the proof of concept established, we identified the acyl group, among a range of amides and carbamates, as the optimal activating group for the sulfonamide reagent in this transformation (Fig. 3, **4–7**). We reasoned that the acidity and the steric encumbrance of the sulfonamide activating group control the nucleophilicity of the arylsulfonylacetamide.

A substantial increase in aminoarylation was observed when using 1,2-disubstituted *p*-methoxyphenyl alkenes in comparison to **1** (Fig. 3A). This substitution allowed us to realize the aryl transfer of several groups including 1-naphthyl (**4–6**, **8–10**, **21**, **22**), 2-naphthyl (**11**), 3-thiophenyl (**12**, **13**), 2-thiophenyl (**14**, **15**, **18**), 2-furanyl (**16**), 8-quinolino (**17**), 2-benzothiazole (**19**), and β-styrene (**20**) all in greater than 20:1 diastereoselectivity. X-ray crystallographic analysis of **15** was found to show a *syn*-configuration between the 5-bromothiophene and the acetamide groups supporting the stereochemical assignment. Use of cyclic (*E*)-alkenes allowed for the synthesis of cyclic arylethylamines (**23–26**) containing two contiguous stereocenters, one of which is quaternary (Fig. 3B). Furthermore, the *cis*-diastereomer

**27** can be formed when a cyclic (*Z*)-alkene is used as the oxidizable alkene substrate partner (Fig. 3C). Preparation of arylethylamine **21** containing an *N*-tosyl amide showcases the chemoselective nature of this aminoarylation, and the successful isolation of **22** suggests that nucleophiles tethered to the alkene are well tolerated under the reaction conditions. The current aminoarylation conditions are not amenable to benzenesulfonylacetamides, likely as a result of the increased enthalpic barrier for dearomatization during the initial radical cyclization (fig. S3).

To provide mechanistic insight, we carried out several studies to understand the efficiency and high diastereoselectivity of this transformation. We hypothesized that both acyclic (*Z*)- and (*E*)-alkenes would convert to the same *trans*-aminoarylation



**Fig. 4. Experiments to probe reaction mechanism.** (A) Aminoarylation with (Z)-anethole (**28**). (B) Tracking reaction progress for aminoarylation with (Z)-anethole (**28**) [triangles, **2**; red squares, (Z)-anethole; diamonds, **9**; purple squares, (E)-anethole]. (C) Determination of the photostationary state for anethole isomers catalyzed by **3**. (D) Favored and disfavored conformations for intermediate **III**.

diastereomer as a result of bond rotation out-competing cyclization of intermediate **III**. Notably, performing the title aminoarylation with (Z)-anethole afforded a nearly identical yield of **9** (72%), in comparison to (E)-anethole (82%), whereas diastereomer **9'** was not observed (Fig. 4A). Reaction progress analysis by <sup>1</sup>H nuclear magnetic resonance spectroscopy of (Z)-anethole aminoarylation revealed that (E)-anethole is generated during the reaction (Fig. 4B). On the basis of this observation, we examined the rates of isomerization for each anethole isomer to the photostationary state (Fig. 4C). This revealed a photostationary state of 1.4:1 (Z:E), with the initial rate of (Z)-anethole isomerization being much faster than (E)-anethole isomerization (figs. S4 and S8 to S10) (43). Furthermore, initial rate analysis of aminoarylation shows alkene consumption to be slower (fig. S6) than (Z)-anethole isomerization (figs. S8 and S9). These data suggest that the diastereoselectivity arises from either (i) a kinetically favored generation of (E)-anethole radical cation and subsequent aminoarylation, or (ii) a thermodynamic preference of radical intermediate **III** to adopt an *anti*-periplanar conformation between the *para*-methoxyphenyl (PMP) and methyl substituents prior to cyclization (Fig. 4D). One other competing possibility is that the (E)-anethole radical cation reacts with **2** faster than does the (Z)-anethole radical cation.

Overall, these mechanistic details describe how the combination of a Smiles-Truce aryl transfer and radical cation chemistry can be combined into a highly diastereospecific alkene amino-

arylation. Given the current availability of sulfonamide building blocks along with the ubiquity of alkenes as feedstock substrates, we view the method to be a highly enabling platform for research efforts synthesizing the aryethylamine pharmacophore diastereoselectively in a single operation.

#### REFERENCES AND NOTES

1. J. W. Dalley, B. J. Everitt, *Semin. Cell Dev. Biol.* **20**, 403–410 (2009).
2. A. Zhang, J. L. Neumeyer, R. J. Baldessarini, *Chem. Rev.* **107**, 274–302 (2007).
3. National Institute on Drug Abuse, “Effective Treatments for Opioid Addiction” (2018); [www.drugabuse.gov/publications/effective-treatments-opioid-addiction/effective-treatments-opioid-addiction](http://www.drugabuse.gov/publications/effective-treatments-opioid-addiction/effective-treatments-opioid-addiction).
4. V. Spahn *et al.*, *Science* **355**, 966–969 (2017).
5. H. Ding *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5511–E5518 (2016).
6. D. M. Schultz, J. P. Wolfe, *Synthesis* **44**, 351–361 (2012).
7. Z. Liu *et al.*, *J. Am. Chem. Soc.* **139**, 11261–11270 (2017).
8. S. R. Chemler, *Org. Biomol. Chem.* **7**, 3009–3019 (2009).
9. D. Wang *et al.*, *J. Am. Chem. Soc.* **139**, 6811–6814 (2017).
10. C. E. Sleat, U. K. Tambar, *Angew. Chem. Int. Ed.* **56**, 5536–5540 (2017).
11. H.-B. Yang, S. R. Pathipati, N. Selander, *ACS Catal.* **7**, 8441–8445 (2017).
12. S. Z. Tasker, T. F. Jamison, *J. Am. Chem. Soc.* **137**, 9531–9534 (2015).
13. E. Tkatchouk, N. P. Mankad, D. Benitez, W. A. Goddard 3rd, F. D. Toste, *J. Am. Chem. Soc.* **133**, 14293–14300 (2011).
14. T. E. Müller, M. Beller, *Chem. Rev.* **98**, 675–704 (1998).
15. T. Piou, T. Rovis, *Nature* **527**, 86–90 (2015).
16. S. Kindt, M. R. Heinrich, *Synthesis* **48**, 1597–1606 (2016).
17. A. J. Musacchio *et al.*, *Science* **355**, 727–730 (2017).
18. A. J. Musacchio, L. Q. Nguyen, G. H. Beard, R. R. Knowles, *J. Am. Chem. Soc.* **136**, 12217–12220 (2014).
19. Q. Zhu, D. E. Graff, R. R. Knowles, *J. Am. Chem. Soc.* **140**, 741–747 (2018).
20. G. J. Choi, R. R. Knowles, *J. Am. Chem. Soc.* **137**, 9226–9229 (2015).
21. D. C. Miller, G. J. Choi, H. S. Orbe, R. R. Knowles, *J. Am. Chem. Soc.* **137**, 13492–13495 (2015).
22. K. A. Margrey, D. A. Nicewicz, *Acc. Chem. Res.* **49**, 1997–2006 (2016).
23. T. M. Nguyen, N. Manohar, D. A. Nicewicz, *Angew. Chem. Int. Ed.* **53**, 6198–6201 (2014).
24. N. J. Gesmundo, J.-M. M. Grandjean, D. A. Nicewicz, *Org. Lett.* **17**, 1316–1319 (2015).
25. S. Z. Zard, *Chem. Soc. Rev.* **37**, 1603–1618 (2008).
26. K. Plesniak, A. Zarecki, J. Wicha, *Top. Curr. Chem.* **275**, 163–250 (2007).
27. W. B. Motherwell, A. M. K. Pennell, *J. Chem. Soc. Chem. Commun.* **1991**, 877–879 (1991).
28. C. M. Holden, M. F. Greaney, *Chem. Eur. J.* **23**, 8992–9008 (2017).
29. A. R. P. Henderson, J. R. Kosowan, T. E. Wood, *Can. J. Chem.* **95**, 483–504 (2017).
30. T. J. Snape, *Chem. Soc. Rev.* **37**, 2452–2458 (2008).
31. I. Allart-Simon, S. Gérard, J. Sapi, *Molecules* **21**, 878–889 (2016).
32. J. J. Douglas, H. Albright, M. J. Sevrin, K. P. Cole, C. R. J. Stephenson, *Angew. Chem. Int. Ed.* **54**, 14898–14902 (2015).
33. J. J. Douglas, M. J. Sevrin, K. P. Cole, C. R. J. Stephenson, *Org. Process Res. Dev.* **20**, 1148–1155 (2016).
34. P. T. G. Rabet, S. Boyd, M. F. Greaney, *Angew. Chem. Int. Ed.* **56**, 4183–4186 (2017).
35. L. J. Johnston, N. P. Schepp, *J. Am. Chem. Soc.* **115**, 6564–6571 (1993).
36. N. P. Schepp, L. J. Johnston, *J. Am. Chem. Soc.* **118**, 2872–2881 (1996).
37. H.-C. Xu, K. D. Moeller, *J. Am. Chem. Soc.* **130**, 13542–13543 (2008).
38. Z.-M. Chen, X.-M. Zhang, Y.-Q. Tu, *Chem. Soc. Rev.* **44**, 5220–5245 (2015).
39. H. G. Roth, N. A. Romero, D. A. Nicewicz, *Synlett* **27**, 714–723 (2016).
40. G. J. Choi, Q. Zhu, D. C. Miller, C. J. Gu, R. R. Knowles, *Nature* **539**, 268–271 (2016).

41. C. K. Prier, D. A. Rankic, D. W. C. MacMillan, *Chem. Rev.* **113**, 5322–5363 (2013).  
42. S. Fukuzumi *et al.*, *J. Am. Chem. Soc.* **123**, 8459–8467 (2001).  
43. F. D. Lewis, M. Kojima, *J. Am. Chem. Soc.* **110**, 8660–8664 (1988).

#### ACKNOWLEDGMENTS

We thank J. W. Kampf for assistance with x-ray crystallographic analyses. **Funding:** Supported by National Institute of General Medical Sciences grant R01-GM096129, the Camille Dreyfus Teacher-Scholar

Award Program, and the University of Michigan. This material is based on work supported by a NSF Graduate Research Fellowship (grant DGE 1256260) (R.C.M.). **Author contributions:** T.M.M. and R.C.M. performed the experiments; T.M.M., R.C.M., and C.R.J.S. designed the experiments; T.M.M., R.C.M., and C.R.J.S. wrote the manuscript.

**Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** X-ray data for compounds **15** and **23** are available free of charge from the Cambridge Crystallographic Data Centre under CCDC 1572215 and 1572214, respectively.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6409/1369/suppl/DC1](http://www.sciencemag.org/content/361/6409/1369/suppl/DC1)  
Materials and Methods  
Figs. S1 to S11  
Tables S1 to S5  
References (44–54)

4 February 2018; accepted 25 July 2018  
10.1126/science.aat2117



## PERSISTENT CHEMICALS

# Predicting global killer whale population collapse from PCB pollution

Jean-Pierre Desforbes<sup>1\*</sup>, Ailsa Hall<sup>2\*</sup>, Bernie McConnell<sup>2</sup>, Aqquale Rosing-Asvid<sup>3</sup>, Jonathan L. Barber<sup>4</sup>, Andrew Brownlow<sup>5</sup>, Sylvain De Guise<sup>6,7</sup>, Igor Eulaers<sup>1</sup>, Paul D. Jepson<sup>8</sup>, Robert J. Letcher<sup>9</sup>, Milton Levin<sup>6</sup>, Peter S. Ross<sup>10</sup>, Filipa Samarra<sup>11</sup>, Gísli Víkingsson<sup>11</sup>, Christian Sonne<sup>1</sup>, Rune Dietz<sup>1\*</sup>

Killer whales (*Orcinus orca*) are among the most highly polychlorinated biphenyl (PCB)-contaminated mammals in the world, raising concern about the health consequences of current PCB exposures. Using an individual-based model framework and globally available data on PCB concentrations in killer whale tissues, we show that PCB-mediated effects on reproduction and immune function threaten the long-term viability of >50% of the world's killer whale populations. PCB-mediated effects over the coming 100 years predicted that killer whale populations near industrialized regions, and those feeding at high trophic levels regardless of location, are at high risk of population collapse. Despite a near-global ban of PCBs more than 30 years ago, the world's killer whales illustrate the troubling persistence of this chemical class.

The widespread industrial use of polychlorinated biphenyls (PCBs) during the 20th century led to ubiquitous contamination of the biosphere, with substantial harm among different wildlife populations (1). PCBs are toxic anthropogenic compounds shown to impair reproduction, disrupt the endocrine and immune systems, and increase the risk of cancer in vertebrates (2, 3). National and international regulatory actions succeeded in reducing PCB contamination of the environment primarily in the first decades after the bans (4); however, PCB concentrations remain high in many long-lived wildlife species because of their environmental persistence and efficient biological cycling (mother-calf transfer), as well as dietary shifts in some species over time to more contaminated prey (2, 5). For example, PCB concentrations are exceedingly high in the tissue of high-trophic level killer whales (*Orcinus*

*orca*) and other dolphin species (5, 6). It has been suggested that high PCB concentrations in killer whales may be contributing to observations of low recruitment and population decline, potentially leading to local extinctions (5, 7). To date, only one study, focusing on resident killer whales in western Canada, has investigated population risk from PCB exposure (8). Exposure modeling predicted protracted health risks in these resident populations over the next century, underscoring the vulnerability of this long-lived species to PCBs (9). With many killer whale populations facing growing conservation pressures, there is an urgent need to assess the impact of PCBs on global killer whale populations.

We compiled available data on blubber PCB concentrations [ $\Sigma$ PCBs, mg/kg lipid weight (lw)] in killer whales from populations around the world and compared these to established concentration-response relationships for reproductive impairment and immunotoxicity-related disease mortality using an individual-based model framework (8, 10). This model incorporates published killer whale fecundity and survival data to construct a stable age-structured baseline population. The model then simulates the accumulation and loss of PCBs in blubber through placental and lactation transfer to the fetus and calf, as well as prey ingestion after weaning. Simulated PCB concentrations are then evaluated against concentration-response relationships for calf survival and immune suppression. Immunity is linked to survival probability based on relationships between immune suppression and disease mortality (11). We then forecast the predicted effects of PCB exposure on killer whale population growth around the world over the next 100 years.

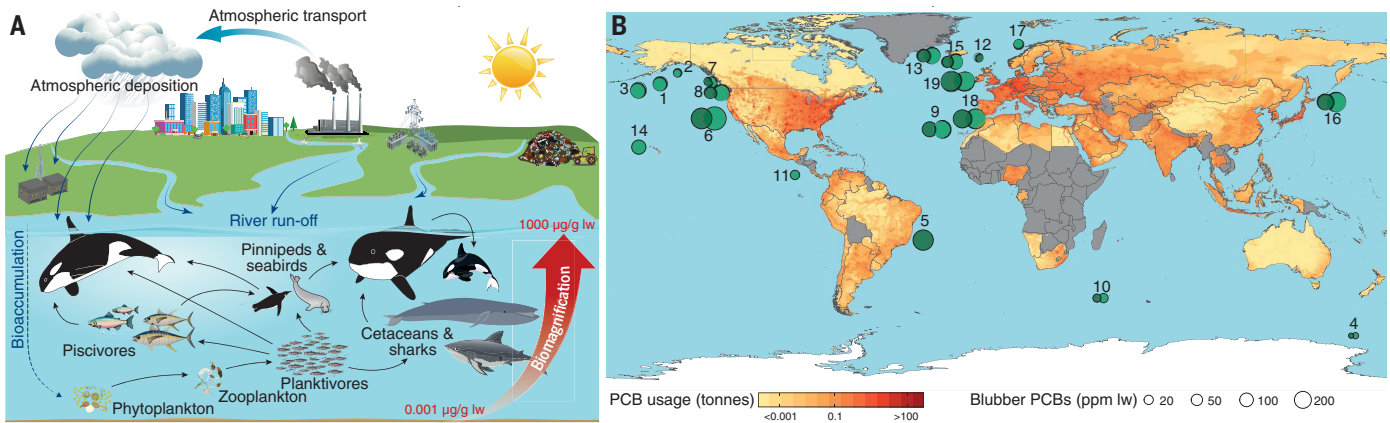
PCB concentrations in killer whales around the world reflect proximity to PCB production and usage, as well as diet and trophic level (Fig. 1 and

table S1). Global PCB production (1930 to 1993) was estimated to be between 1 and 1.5 million metric tons (tonnes), and mostly occurring in the United States (~50%), Russia (~13%), Germany (~12%), France (~10%), and the United Kingdom (5%) (12, 13). The global manufacture of PCBs corresponded well with the observed pattern of PCB levels in killer whale populations, which ranged widely from lowest values in Antarctica, <10 mg/kg lw (14), to values above 500 mg/kg lw in individuals near the highly industrialized areas of the Strait of Gibraltar, the United Kingdom, and the Northeast Pacific (5, 15, 16). Diet is an important contributor to PCB accumulation in killer whales via biomagnification across trophic levels, resulting in sharp differences between populations feeding on marine mammals, tuna (*Scombridae*), and sharks (*Selachimorpha*) and those feeding on lower-trophic level fish (Fig. 1 and table S1). This is exemplified in the Northeast Pacific where marine mammal-eating Bigg's killer whales carry 10- to 20-fold higher PCB burdens compared to fish-eating northern residents, despite sharing the same coastline (15, 17). Overall, females exhibit lower blubber PCB levels than males because of maternal sequestration to young during fetal development and lactation (18, 19). Exceptions have been reported in the most highly PCB-contaminated populations, including in the United Kingdom, Strait of Gibraltar (5), and Bigg's individuals in the Northeast Pacific (17), suggesting that PCBs may be limiting successful reproduction and consequently reducing the maternal loss of PCBs.

Model forecasting over the next 100 years shows the large potential impact of PCBs on population size and long-term viability of long-lived killer whales around the world (Fig. 2). Killer whale populations with similar PCB levels were grouped together and assigned to exposure groups (Fig. 2, C and D, and table S1) (10). The modeled reference (unexposed) population grew by 141% [interquartile range (25/75th) = 96.3 to 176.5%] over the 100-year simulation period. The least-contaminated populations (group 1) included Alaskan residents, Antarctica type C, Canadian Northern residents, Crozet Archipelago, Eastern Tropical Pacific, and Norwegian populations. These are estimated to accumulate 1 mg/kg lw of PCBs per year, resulting in median blubber concentrations of 7.9 (4.7 to 14.0) mg/kg lw and effects causing a population decrease of 8.8% (4.1 to 25.3%) or 15.4% (3.5 to 25.2%) relative to the reference population for reproductive effects alone or combined reproductive and immune effects, respectively. However, although relative population-level effects were observed for these low-exposed populations, the model still predicts a net doubling in their population size over 100 years (Fig. 2C and figs. S2 and S3). Annual PCB accumulation rates of 3, 6, 9, 15, 18, and 27 mg/kg are represented by exposure groups 2 through 7, which have incrementally greater blubber PCB levels (Fig. 2C and table S1). Alaskan offshore, Faroe Islands, and Iceland whales (group 2) have similar PCB burdens (13.9 to 41.5 mg/kg lw) and are predicted to have modest population growth over the 100-year simulation period, albeit at a reduced rate relative to the

<sup>1</sup>Department of Bioscience, Arctic Research Centre, Aarhus University, Frederiksborgvej 399, P.O. Box 358, 4000 Roskilde, Denmark. <sup>2</sup>Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, KY16 8LB, UK. <sup>3</sup>Greenland Institute of Natural Resources, P.O. Box 570, 3900 Nuuk, Greenland. <sup>4</sup>Centre for Environment, Fisheries and Aquaculture Science, Pakefield Road, Lowestoft NR33 0HT, UK. <sup>5</sup>Scottish Marine Animal Stranding Scheme, SRUC Veterinary Services Drummondhill, Stratherrick Road, Inverness IV2 4JZ, UK. <sup>6</sup>Department of Pathobiology and Veterinary Science, University of Connecticut, 61 North Eagleville Road, Storrs, CT 06269-3089, USA. <sup>7</sup>Connecticut Sea Grant, 1080 Shennecossett Road, Groton, CT 06340-6048, USA. <sup>8</sup>Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK. <sup>9</sup>Ecotoxicology and Wildlife Health Division, Environment and Climate Change Canada, National Wildlife Research Centre, Carleton University, Ottawa, Ontario K1A 0H3, Canada. <sup>10</sup>Ocean Wise Conservation Association, P.O. Box 3232, Vancouver, British Columbia V6B 3X8, Canada. <sup>11</sup>Marine and Freshwater Research Institute, Skúlagata 4, 101 Reykjavík, Iceland.

\*Corresponding author. Email: jpd@bios.au.dk (J.-P.D.); rdi@bios.au.dk (R.D.); ajh7@st-andrews.ac.uk (A.H.)



**Fig. 1. Global PCB concentrations in killer whales.** (A) Conceptual model of PCB bioaccumulation and magnification, leading to elevated PCB concentrations in killer whale populations. (B) Global overview of PCB concentrations in killer whale blubber (ppm, parts per million).

Light and dark green circles represent males and females, respectively. Also shown is population density–normalized cumulative global usage of PCBs per country from 1930 to 2000 (12). Number labels indicate populations with measured PCB concentrations (table S1).

**Table 1. Global assessment of population-level risk from PCB exposure.** Risk categories were set based on predicted growth rates ( $\lambda$ ) and significant difference by using a one-sample *t* test against a reference of no growth ( $\lambda = 1$ ): low risk ( $\lambda > 1$ , little to no effect on population growth), moderate risk ( $\lambda = 1$ , stagnant population growth), high risk ( $\lambda < 1$ , population decline).

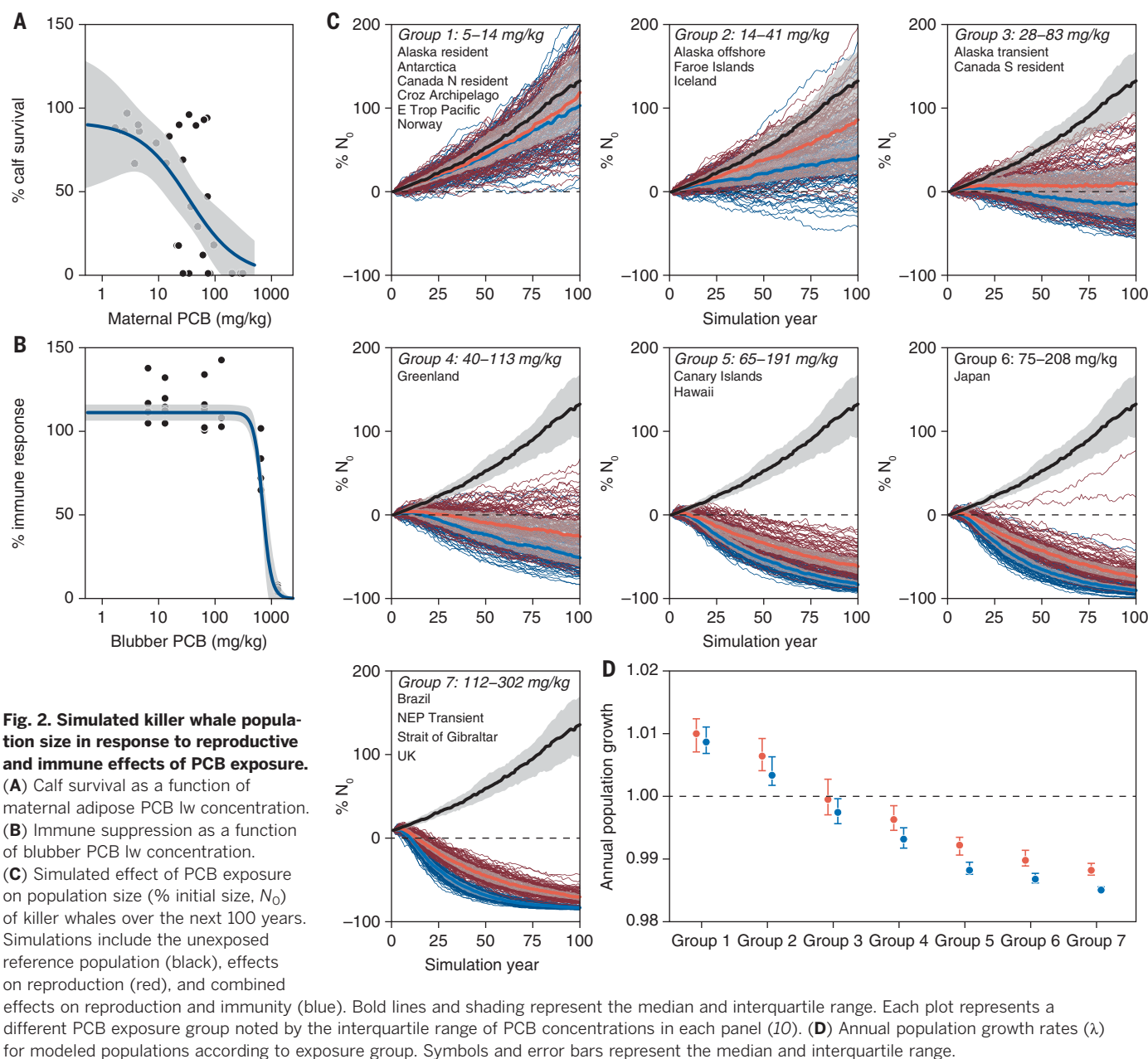
PCB risk	Population	Location	Population size	Protection status
<b>Low</b> ( $\lambda > 1$ )	Alaska offshore	North Pacific	>200*	None*
	Alaska resident	North Pacific	2347*	None*
	Antarctica type C	Southern Ocean	Unknown	Unknown
	Northeast Pacific	Northeast Pacific	290†	Threatened†
	Crozet Archipelago	South Indian Ocean	37–98‡	Unknown
	Eastern Tropical Pacific	Tropical Pacific	8500*	Unknown
	Faroe Islands	Northeast Atlantic	Unknown	Unknown
	Iceland	North Atlantic	376§	None§
	Norway	Northeast Atlantic	500–1100	Unknown
<b>Moderate</b> ( $\lambda = 1$ )	Alaska transient	North Pacific	587*	None/Depleted*
	Canada South resident	Northeast Pacific	78†	Endangered†
<b>High</b> ( $\lambda < 1$ )	Brazil	Southwest Atlantic	Unknown	Unknown
	Northeast Pacific Bigg's	Northeast Pacific	521*	None*/Threatened†
	Canary Islands	Atlantic Ocean	Unknown	Unknown
	Greenland	North Atlantic	Unknown	None
	Hawaii	Tropical Pacific	101*	None*
	Japan	Northwest Pacific	Unknown	Unknown
	Strait of Gibraltar	Mediterranean	36¶	Vulnerable¶
	United Kingdom	Northeast Atlantic	≤9#	None

\*National Oceanographic and Atmospheric Administration (NOAA) stock assessment reports ([www.fisheries.noaa.gov/species/killer-whale](http://www.fisheries.noaa.gov/species/killer-whale)); AT1 transients in Alaska are a subgroup considered depleted under the U.S. Marine Mammal Protection Act. †Government of Canada, Species at Risk Public Registry ([www.sararegistry.gc.ca/default.asp?lang=en&n=24F7211B-1](http://www.sararegistry.gc.ca/default.asp?lang=en&n=24F7211B-1)). ‡(27). §(28). ||(29). ¶(30). # (5).

reference population; modeled PCB effects on reproduction alone or in combination with immune suppression resulted in a population reduction of 22.6% (14.0 to 38.3%) or 40.5% (32.6 to 48.7%). Alaskan transient and Canadian Southern resident populations have similar PCB burdens (group 3: 28 to 83 mg/kg lw), and PCB effects are predicted to inhibit population growth or cause a gradual decline of ~15% (4.3 to 33.9%) for reproductive or

combined effects, respectively. These represent median reductions of 54.7 and 64.7% relative to unexposed populations. Greenland, Canary Islands, Hawaii, Japan, Brazil, Northeast Pacific Bigg's, Strait of Gibraltar, and U.K. populations all possess PCB levels above 40 mg/kg lw (Fig. 2C), and this level of exposure is predicted to cause population declines at various rates depending on the exposure group. Populations of Japan, Brazil, North-

east Pacific Bigg's, Strait of Gibraltar, and United Kingdom are all tending toward complete collapse in our modeled scenarios. To quantify and compare the global risk of PCB exposure in killer whales, we used population trajectories from the model to calculate potential annual population growth rates ( $\lambda$ ). The achievable growth rates, incorporating combined PCB effects on both reproduction and immune function, were



at or below the growth threshold ( $\lambda = 1$ ) for 10 of the 19 populations for which information on PCB exposure is currently available (Fig. 2D and Table 1). These results suggest that chronic exposure to persistent PCBs has the potential to affect long-term population viability in more than half of all studied killer whale populations. Of these, Alaskan transient and Canada Southern resident populations are at moderate risk of population-level effects ( $\lambda = 1$ ), whereas Brazilian, Northeast Pacific Bigg's, Canary Islands, Greenlandic, Hawaiian, Japanese, Strait of Gibraltar, and U.K. populations are at high risk of collapse over the next 100 years. The model predicted low PCB risk and stable population growth ( $\lambda > 1$ ) for the remaining nine populations (Fig. 2D and Table 1).

Our global assessment of PCB-related effects on the long-term viability of killer whale populations represents a fundamental advancement in our understanding of population impacts from chronic exposure to these legacy chemicals in a long-lived marine apex predator. More than 35 years after the onset of the ban on PCBs, killer whales still have PCB concentrations reported to be as high as 1300 mg/kg lw (20). Killer whales once thrived in all oceans of the world, but only those in the less-contaminated waters of the Arctic and Antarctic today appear to be able to sustain growth (Table 1) (7, 21). We had no PCB data for killer whales in the Gulf of Mexico, but even before the Deep Water Horizon oil spill in 2010, estimates for killer whales in the region are consistent with a progressive population collapse from 277 individ-

uals in 1991–1994, 133 in 1996–2001, 49 in 2003–2004, and only 28 in 2009 (22). Prey switching from low to high PCB-contaminated prey sources (e.g., fish to seals) has been documented in some killer whale populations like Northeast Scotland (United Kingdom) and Greenland (23, 24), which is likely to have important consequences for PCB exposure in these already vulnerable populations. Prey switching is likely a function of prey availability as fish stocks and seal populations fluctuate over time (23, 24). Our finding that a single chemical class (PCBs) may represent a substantial conservation threat to killer whales around the world raises concerns about the potential for other persistent contaminants to generate additional toxic injury in long-lived, high-trophic level aquatic species. Indeed, a long list of additional



known and as yet unmeasured contaminants are present in killer whale tissues, including biologically active compounds like perfluoroalkyl acids, brominated and organophosphate flame retardants, and polychlorinated naphthalenes (25), and although these are less well characterized, they may contribute to reproductive and immune failure or other health endpoints not included here.

The status-quo efforts to protect killer whales from conservation threats are likely to be impeded because PCBs have remained at levels associated with adverse health effects in at-risk populations over the past decades (5, 7, 9). Concerted efforts beyond those listed under the Stockholm Convention on Persistent Organic Pollutants (POPs) are urgently needed to reduce PCB exposure in vulnerable wildlife populations. It is estimated that more than 80% of global PCB stocks are yet to be destroyed, and at present rates of PCB elimination, many countries will not achieve the 2025 and 2028 targets as agreed upon under the Stockholm Convention on POPs (26). Although killer whale populations face other anthropogenic stressors such as prey limitation and underwater noise (21), our assessment here clearly demonstrates the high risk of collapse for many killer whale populations as a consequence of their PCB exposures alone.

## REFERENCES AND NOTES

1. A. M. Roos, B.-M. V. M. Backlin, B. O. Helander, F. F. Rigét, U. C. Eriksson, *Environ. Pollut.* **170**, 268–275 (2012).
2. R. J. Letcher *et al.*, *Sci. Total Environ.* **408**, 2995–3043 (2010).
3. S. H. Safe, *Crit. Rev. Toxicol.* **24**, 87–149 (1994).
4. AMAP, *AMAP Assessment 2002: Persistent Organic Pollutants in the Arctic* (Arctic Monitoring and Assessment Programme, Oslo, Norway, 2004).
5. P. D. Jepson *et al.*, *Sci. Rep.* **6**, 18573 (2016).
6. B. C. Balmer *et al.*, *Sci. Total Environ.* **409**, 2094–2101 (2011).
7. P. D. Jepson, R. J. Law, *Science* **352**, 1388–1389 (2016).
8. A. J. Hall *et al.*, *Environ. Pollut.* **233**, 407–418 (2018).
9. B. E. Hickie, P. S. Ross, R. W. Macdonald, J. K. B. Ford, *Environ. Sci. Technol.* **41**, 6613–6619 (2007).
10. Materials and methods are available as supplementary materials.
11. M. I. Luster *et al.*, *Fundam. Appl. Toxicol.* **21**, 71–82 (1993).
12. K. Breivik, A. Sweetman, J. M. Pacyna, K. C. Jones, *Sci. Total Environ.* **290**, 181–198 (2002).
13. K. Breivik, A. Sweetman, J. M. Pacyna, K. C. Jones, *Sci. Total Environ.* **377**, 296–307 (2007).
14. M. M. Krahn, R. L. Pitman, D. G. Burrows, D. P. Herman, R. W. Pearce, *Mar. Mamm. Sci.* **24**, 643–663 (2008).
15. P. S. Ross, G. M. Ellis, M. G. Ikonou, L. G. Barrett-Lennard, R. F. Addison, *Mar. Pollut. Bull.* **40**, 504–515 (2000).
16. M. M. Krahn *et al.*, *Mar. Environ. Res.* **63**, 91–114 (2007).
17. A. H. Buckman *et al.*, *Environ. Sci. Technol.* **45**, 10194–10202 (2011).
18. J.-P. W. Desforges, P. S. Ross, L. L. Loseto, *Environ. Toxicol. Chem.* **31**, 296–300 (2012).
19. A. Aguilar, A. Borrell, *Arch. Environ. Contam. Toxicol.* **27**, 546–554 (1994).
20. M. M. Krahn *et al.*, *J. Cetacean Res. Manag.* **6**, 175–189 (2004).
21. International Union for Conservation of Nature, The IUCN Red List of Threatened Species, Version 2015.4 (IUCN, Gland, 2015); [www.iucnredlist.org](http://www.iucnredlist.org).
22. S. A. Hayes *et al.*, US Atlantic and Gulf of Mexico Marine Mammal Stock Assessments–2016. NOAA Tech. Memo. NMFS NE 241.
23. D. Vongraven, A. Bisther, *J. Mar. Biol. Assoc. U. K.* **94**, 1357–1365 (2014).
24. A. D. Foote *et al.*, *Proc. Biol. Sci.* **280**, 20131481 (2013).
25. J.-P. Desforges *et al.*, *Environ. Sci. Technol.* **51**, 11431–11439 (2017).
26. United Nations Environment Programme, Consolidated assessment of efforts made towards the elimination of polychlorinated biphenyls (Switzerland, 2016); [https://wedocs.unep.org/bitstream/handle/20.500.11822/13664/Consolidated%20PCB%20Assessment\\_2016.pdf](https://wedocs.unep.org/bitstream/handle/20.500.11822/13664/Consolidated%20PCB%20Assessment_2016.pdf).
27. É. Poncelet, C. Barbraud, C. Guinet, *J. Cetacean Res. Manag.* **11**, 41–48 (2010).
28. F. I. P. Samarra *et al.*, *Mar. Biol.* **164**, 159 (2017).
29. S. Kuningas, T. Similä, P. S. Hammond, *J. Mar. Biol. Assoc. U. K.* **94**, 1–15 (2013).
30. R. Esteban *et al.*, *Ecol. Indic.* **66**, 291–300 (2016).

## ACKNOWLEDGMENTS

We thank all persons involved in the killer whale sampling necessary to determine PCB concentrations, as well as T. Christensen for help producing the manuscript figures. **Funding:** This research was supported by grants to J.-P.D. from the Canadian National Science and Engineering Research Council (NSERC) (PGSD3-443700-2013) and Aarhus University's Graduate School and Science and Technology (GSST) and Department of Bioscience; and by funding from the Danish DANCEA program (MST-112-00171 and MST-112-00199); the Defra, Scottish and Welsh Governments (for CSIP/SMAS/CEFAS); and the Icelandic Research Fund (i. Rannsóknasjóður; grant no. 120248042). B.M. was supported by funding from NERC (grant no. SMRU 10001). This paper is a contribution from the BONUS BALTHEALTH project, which has received funding from BONUS (Art. 185), funded jointly by the EU, Innovation Fund Denmark, Forschungszentrum Jülich GmbH, German Federal Ministry of Education and Research (grant no. FKZ 03F0767A), Academy of Finland (decision no. 311966), and Swedish Foundation for Strategic Environmental Research. **Author contributions:** J.-P.D., A.H., R.D., C.S., and I.E. designed the study. R.D., A.R.-A., G.V., F.S., R.J.L., J.L.B., A.B., P.S.R., and P.D.J. provided samples or performed the contaminant analysis. J.-P.D., M.L., S.D., I.E., and R.J.L. performed the contaminant cocktail extractions and immunological experiments. J.-P.D. collected the data. A.H. and B.M. developed and ran the model. J.-P.D. generated figures and performed all data analyses. All authors were involved in interpretation of results and writing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6409/1373/suppl/DC1](http://www.sciencemag.org/content/361/6409/1373/suppl/DC1)  
Materials and Methods  
Figs. S1 to S3  
Tables S1 and S2  
References (31–56)

8 February 2018; accepted 16 August 2018  
10.1126/science.aat1953

## EVO-DEVO

# An axial Hox code controls tissue segmentation and body patterning in *Nematostella vectensis*

Suonan He<sup>1</sup>, Florencia del Viso<sup>1</sup>, Cheng-Yi Chen<sup>1</sup>, Aissam Ikmi<sup>1,2</sup>, Amanda E. Kroesen<sup>1</sup>, Matthew C. Gibson<sup>1,3\*</sup>

Hox genes encode conserved developmental transcription factors that govern anterior-posterior (A-P) patterning in diverse bilaterian animals, which display bilateral symmetry. Although Hox genes are also present within Cnidaria, these simple animals lack a definitive A-P axis, leaving it unclear how and when a functionally integrated Hox code arose during evolution. We used short hairpin RNA (shRNA)-mediated knockdown and CRISPR-Cas9 mutagenesis to demonstrate that a Hox-Gbx network controls radial segmentation of the larval endoderm during development of the sea anemone *Nematostella vectensis*. Loss of Hox-Gbx activity also elicits marked defects in tentacle patterning along the directive (orthogonal) axis of primary polyps. On the basis of our results, we propose that an axial Hox code may have controlled body patterning and tissue segmentation before the evolution of the bilaterian A-P axis.

Cnidarians (corals, jellyfish, and sea anemones) occupy a key position within the animal phylogeny, serving as an essential out-group for understanding the evolution of developmental processes throughout Bilateria (1–3). The characteristic polyp bauplan is a unifying feature within Cnidaria, typified by a tubelike body column with a single oral opening surrounded by tentacles (4, 5). Bilaterian body plans are more diverse, exhibiting increasing levels of complexity that correlate with the expansion of genomically clustered Hox genes (6, 7). In arthropods and chordates, for instance, Hox genes are expressed along the anterior-posterior (A-P) axis

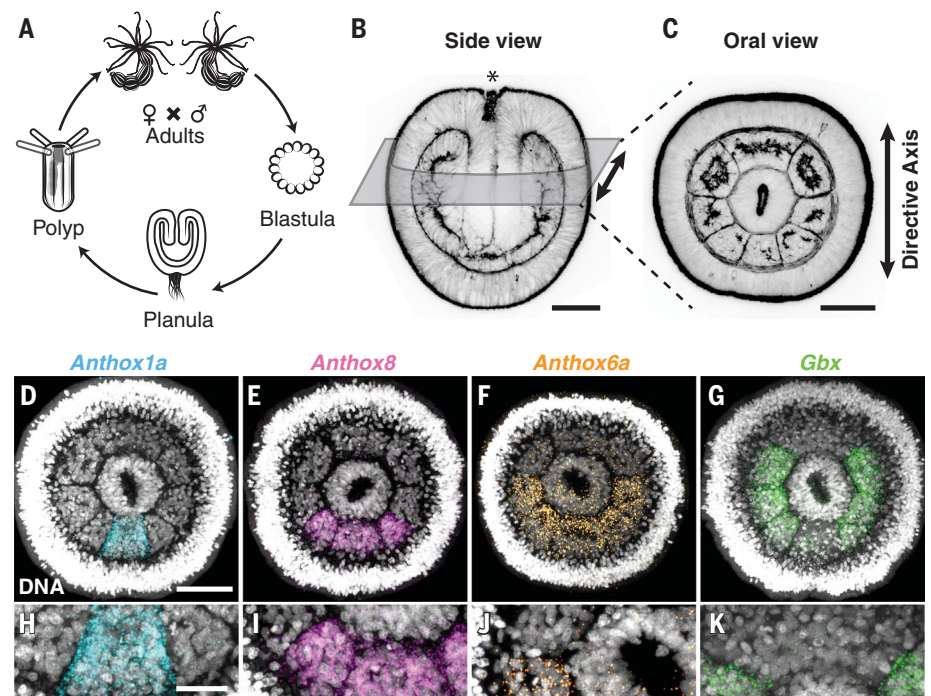
in staggered domains, setting up a molecular code that determines body segment identity and directs the formation of distinct appendages (8, 9). Among early-branching phyla, true Hox genes are found only within Cnidaria, indicating an ancient evolutionary origin predating the bilaterian-cnidarian split approximately 600 million years ago (1, 10, 11). Nevertheless, functional requirements for cnidarian Hox genes are unclear, leaving the possible ancestral role of this crucial developmental gene cluster poorly understood (12).

The sea anemone *Nematostella vectensis* is a model anthozoan cnidarian that has multiple Hox genes (13, 14). Under the control of bone mor-

phogenetic protein (BMP) signaling, *Anthox1a*, *Anthox6a*, and *Anthox8*, together with *Gastrulation brain homeobox (Gbx)* (a Hox-linked subfamily gene), exhibit partially overlapping endodermal expression patterns in planula larvae (10, 14, 15). During this stage, the developing endoderm undergoes morphogenetic segmentation into eight sectors along the directive axis (Fig. 1, A to C), generating internal anatomical subdivisions that further correlate with the positioning of the first four tentacles in metamorphosed polyps (fig. S1). To determine whether the larval Hox-Gbx expression domains specifically define endodermal segment boundaries, we performed fluorescence in situ hybridization (FISH) on mid-planula stage larvae. *Anthox1a*, *Anthox8*, *Anthox6a*, and *Gbx* exhibited sharp expression territories, cell-autonomously defining segment boundaries (Fig. 1, D to K, and fig. S2). Further, each nested Hox expression domain formed in concert with the stepwise sequence of endodermal boundary morphogenesis (fig. S3), indicating a temporal correlation between Hox-Gbx expression and tissue segmentation. These observations suggest that a Hox-Gbx-dependent code controls the formation of the eight endodermal segments (segments s1 to s8) and specifies distinct positional identities along the directive axis (Fig. 2A).

To test the developmental requirements for the Hox-Gbx code during segmentation of the larval endoderm, we took advantage of the distinctive specificity of the cnidarian microRNA

**Fig. 1. *Nematostella* Hox-Gbx expression patterns cell-autonomously correlate with endodermal segment boundaries in planula larvae.** (A) *Nematostella* life cycle. Embryos enter a free-swimming planula larva stage at approximately 48 hours postfertilization. (B and C) Wild-type mid-planula larvae stained to label F-actin. (B) Side view with oral pole (asterisk) facing upward, showing the focal plane for oral view images. (C) Oral view depicting the formation of eight endodermal segments. Scale bars, 50  $\mu$ m. (D to G) Fluorescent in situ hybridizations, stained with Hoechst (DNA) to demonstrate the expression patterns of *Nematostella* *Anthox1a*, *Anthox8*, *Anthox6a*, and *Gbx* in planula larvae. Scale bar, 50  $\mu$ m. (H to K) Magnified images from (D) to (G) illustrating the sharp expression boundaries for each gene, corresponding to distinct endodermal segment boundaries. Scale bar, 25  $\mu$ m.



<sup>1</sup>Stowers Institute for Medical Research, Kansas City, MO 64110, USA.

<sup>2</sup>Developmental Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany.

<sup>3</sup>Department of Anatomy and Cell Biology, The University of Kansas School of Medicine, Kansas City, KS 66160, USA.

\*Corresponding author. Email: mg2@stowers.org



pathway (16) to establish a robust short hairpin RNA (shRNA)-based gene knockdown (KD) technique (figs. S4 to S6 and movies S1 and S2). KD of *Anthox1a*, *Anthox8*, *Anthox6a*, and *Gbx* elicited clear segmentation defects that directly correlated with the genes' endogenous expression domains (Fig. 2, B to F). In *Anthox1a* KD larvae, the boundaries flanking segment s5 were abolished, resulting in the fusion of s4, s5, and s6 into a single large segment (s4-6) (Fig. 2C). In *Anthox8* KD larvae, segment boundaries between s3 and s4 and between s6 and s7 were lost, resulting in two enlarged endodermal segments flanking s5 (s3-4 and s6-7) (Fig. 2D). In *Anthox6a* KD animals, the boundaries between s2 and s3 and between s7 and s8 were lost, generating fusion segments s2-3 and s7-8 flanking s1 (Fig. 2E). Lastly, in *Gbx* KD larvae, the boundaries flanking s1 were abolished, resulting in the fusion of s8, s1, and s2 into a single large segment (s8-2) (Fig. 2F). Similar phenotypes were observed for at least two independent shRNAs targeting each gene (Fig. 2 and materials and methods). Quantitative polymerase chain reaction (qPCR) validation further confirmed substantial reduction of the target mRNA level for each shRNA (fig. S7). These results demonstrate that a Hox-Gbx-dependent code drives morphogenetic tissue segmentation in the endoderm of developing *Nematostella* larvae.

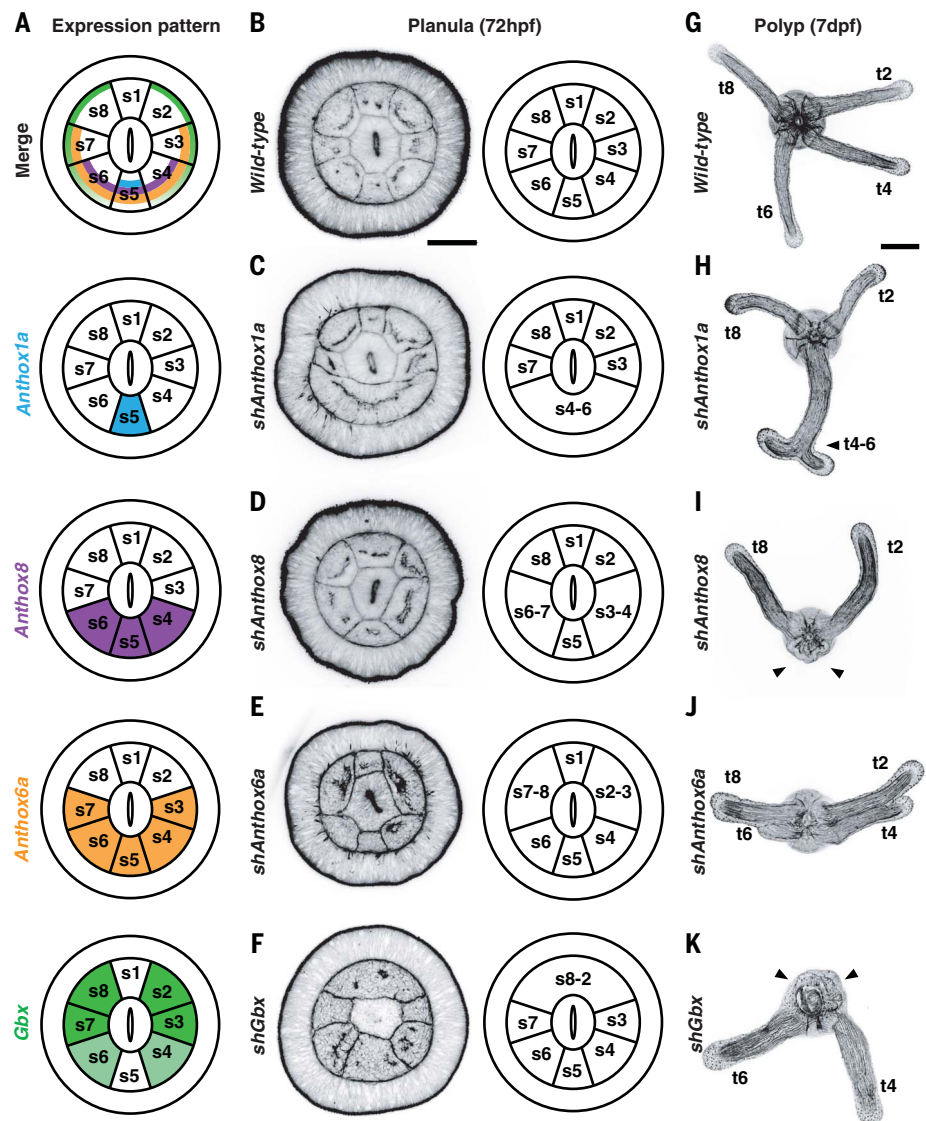
To characterize later roles of *Anthox1a*, *Anthox8*, *Anthox6a*, and *Gbx* genes in body patterning, experimental larvae were reared through metamorphosis to the polyp stage. KD of each gene elicited marked and highly penetrant tentacle-patterning defects, each of which corresponded to the position of larval segmentation abnormalities (Fig. 2, G to K). In wild-type controls, four tentacles of equal size developed in stereotyped radial positions corresponding to the endodermal segments s2, s4, s6, and s8 (tentacles t2, t4, t6, and t8) (Fig. 2G and figs. S1C and S8A). In contrast, *Anthox1a* KD polyps exhibited a single large tentacle replacing t4 and t6, resulting in animals with three tentacles. The tip of this enlarged tentacle was frequently bifurcated, suggesting a possible fusion of t4 and t6 (t4-6) (Fig. 2H and fig. S8B). *Anthox8* KD polyps specifically lost tentacles t4 and t6 (Fig. 2I and fig. S8C), resulting in animals with only two tentacles. *Anthox6a* KD polyps maintained four tentacles, although t2 and t8 were enlarged and partially fused with the adjacent tentacles t4 and t6, respectively (Fig. 2J, fig. S8D). Lastly, *Gbx* KD polyps consistently lost tentacles t2 and t8, resulting in two-tentacled animals with a mirror-image phenotype to *Anthox8* KD animals (Fig. 2K and fig. S8E). Similar developmental requirements were not observed for the other *Nematostella* homeodomain-containing genes, *Anthox6*, *Anthox7*, and *Cdx* (fig. S9). Collectively, these experiments demonstrate key roles for *Anthox1a*, *Anthox8*, *Anthox6a*, and *Gbx* in controlling tentacle patterning, revealing a link between the process of endodermal segmentation and the specification of tentacle primordia.

Previous studies have demonstrated a role for BMP signaling in regulating both larval Hox expression and the development of mesenteries, di-

gestive and reproductive organs that form during larval development at the boundaries between endodermal segments (fig. S6, A to C) (15). To explore additional roles for Hox genes in axial patterning, we developed a method to image *Anthox1a*, *Anthox8*, and *Anthox6a* KD animals by using selective plane illumination microscopy (SPIM) (17). We observed only six mesenteries in each KD condition (movies S1 and S3 to S5), consistent with the earlier defects in boundary formation. Taken together with earlier results, these findings confirm that Hox-dependent segmentation

of the larval endoderm establishes key elements of the polyp bauplan, including the positioning of mesenteries and the patterning of tentacle primordia.

To validate the developmental function of the *Nematostella* Hox-Gbx code with a technically independent approach, we next mutated *Anthox1a*, *Anthox6a*, *Anthox8*, and *Gbx* by using CRISPR-Cas9-mediated genome editing (18–20). After the injection of Cas9-guide RNA (gRNA) complexes into embryos (with either single or paired gRNAs), F<sub>0</sub> indel mutations were recovered in all four



**Fig. 2. *Nematostella* Hox-Gbx genes control tissue segmentation and tentacle patterning.**

(A) Color-coded expression pattern of individual genes in wild-type larvae. (B to F) Oral views of wild-type versus Hox-Gbx KD planula-stage larvae. (B) Wild type [ $n = 24$  larvae (image is representative of 24 of 26 wild-type larvae)]; (C) *Anthox1a* KD ( $n = 31/32$ ); (D) *Anthox8* KD ( $n = 26/28$ ); (E) *Anthox6a* KD ( $n = 30/36$ ); (F) *Gbx* KD ( $n = 26/33$ ). Scale bar, 50  $\mu$ m. hpf, hours postfertilization. (G to K) Oral views of wild-type versus Hox-Gbx KD polyps. (G) Wild type ( $n = 197$  of 264 polyps); (H) *Anthox1a* KD polyps ( $shRNA1$ ,  $n = 214/253$ ;  $shRNA2$ ,  $n = 165/192$ ); (I) *Anthox8* KD polyps ( $shRNA1$ ,  $n = 144/162$ ;  $shRNA2$ ,  $n = 112/132$ ); (J) *Anthox6a* KD polyps ( $shRNA1$ ,  $n = 92/126$ ;  $shRNA2$ ,  $n = 123/162$ ); (K) *Gbx* KD polyps ( $shRNA1$ ,  $n = 84/181$ ;  $shRNA2$ ,  $n = 50/70$ ;  $shRNA3$ ,  $n = 24/38$ ). Arrowheads indicate missing tentacles. Scale bar, 100  $\mu$ m. dpf, days postfertilization.



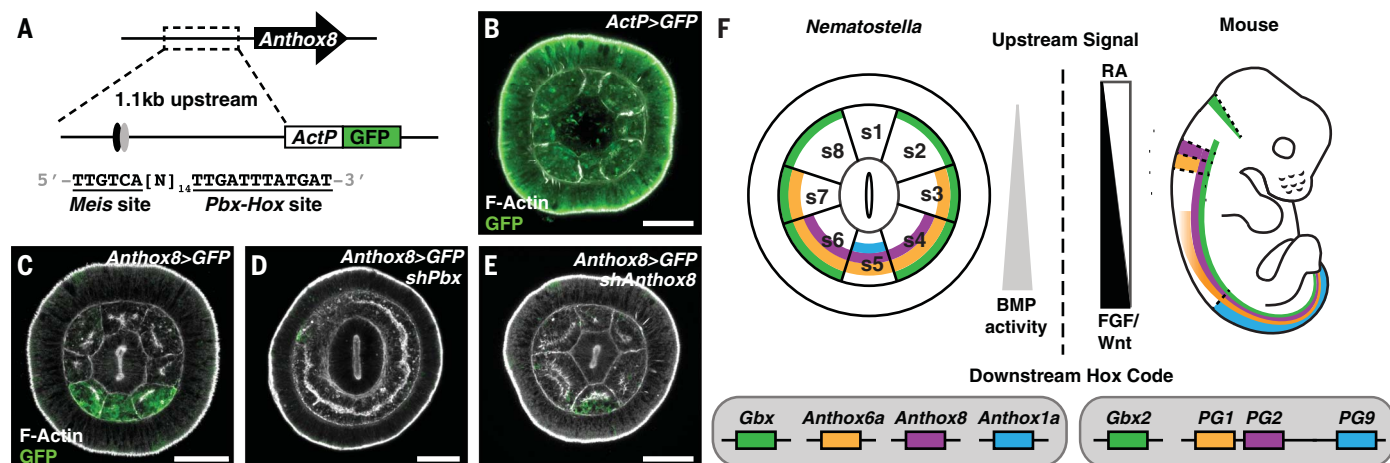
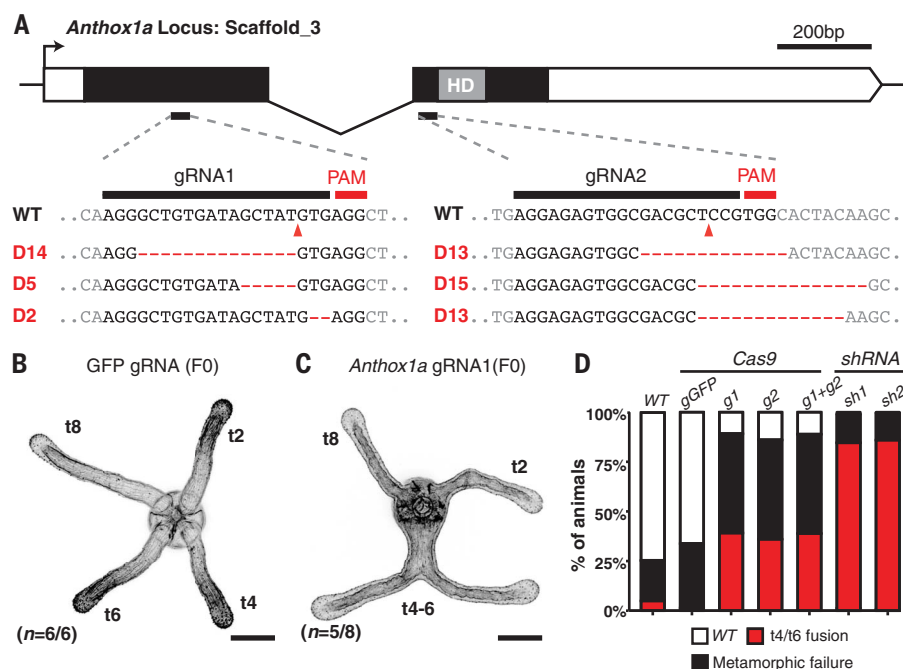
loci, along with low frequencies of the expected tentacle phenotypes (Fig. 3 and fig. S10). Through subsequent controlled crosses, we obtained genetically null mutants for *Anthox1a*, *Anthox8*, and *Anthox6a*. In each case, these animals exhibited tentacle-patterning defects identical to those of the cognate shRNA KD animals (figs. S11 to S14). Putative *Gbx* mutants showed severe growth defects, and only heterozygous animals were recovered from F<sub>0</sub> founder crosses, suggesting that *Gbx* has additional roles in later developmental

processes. Taken together, these complementary CRISPR-Cas9- and shRNA-based methods illuminate clear developmental requirements for a Hox-Gbx network during cnidarian development (Figs. 2 and 3 and figs. S7 to S14).

In bilaterian systems, Hox-dependent patterns typically arise through extensive cross-regulatory interactions (21–24). We therefore performed a series of double-KD experiments to determine whether *Nematostella* Hox genes exhibit genetic interactions during early development. Consistent

with independent requirements for each locus, strictly additive phenotypes were observed in *Anthox1a*-*Anthox8* (loss of t2 and t4), *Anthox1a*-*Gbx* (t4–6 fusion and loss of t2 and t8), and *Anthox8*-*Gbx* (no tentacles) double-KD animals (fig. S8, F to J). To further interrogate the molecular regulation of Hox-Gbx gene expression, we took advantage of a transgenic *Anthox8*-GFP reporter line that faithfully recapitulates endogenous *Anthox8* expression in endodermal segments s4, s5, and s6 (Fig. 4, A to C). Notably, the

**Fig. 3. CRISPR-Cas9-mediated mutagenesis of *Anthox1a* confirms its function in tentacle patterning.** (A) gRNA strategy for the *Anthox1a* locus. F<sub>0</sub> animals injected with either gRNA carried both frameshift and non-frameshift indel mutations (arrowheads indicate mutation positions). HD, homeodomain; PAM, proto-spacer adjacent motif; D, deletion of the specified number of bases. (B and C) In contrast to GFP gRNA-injected controls, putative *Anthox1a* F<sub>0</sub> founders displayed the characteristic t4–6 tentacle fusion phenotype observed in KD experiments. Scale bars, 100  $\mu$ m. (D) Quantification of t4–6 fusion phenotypes observed in uninjected controls [wild type (WT)], GFP gRNA controls (F<sub>0</sub>) (gGFP), *Anthox1a* gRNA-injected animals (F<sub>0</sub>) (g1, g2, and g1+g2), and two independent shRNA KD groups (sh1 and sh2).



**Fig. 4. *Nematostella* Hox-Gbx genes comprise a spatial code that directs axial patterning.** (A) Design of the *Anthox8*-GFP reporter construct. (B and C) Although the *Actin* promoter alone drives ubiquitous GFP expression, addition of the *Anthox8* upstream enhancer region restricts GFP expression to segments s4, s5, and s6 at the mid-planula stage ( $n = 10$  of 10 larvae). (D) *Pbx* shRNA-injected transgenic animals lost all endodermal segmentation and failed to activate the *Anthox8* reporter ( $n = 6$  of 6 animals). (E) shRNAs targeting *Anthox8* substantially decreased the GFP signal ( $n = 5$  of 6 animals), restricting

it to segment s5. Scale bars in (B) to (E), 50  $\mu$ m. (F) Comparison between the functional Hox codes in a representative cnidarian (*Nematostella*) and a representative vertebrate (mouse). Although the Hox codes respond to different upstream signaling pathways along distinct axes [BMP in *Nematostella* and retinoic acid and fibroblast growth factor (RA and FGF) in mouse], similar downstream molecular programs drive segmental patterning. Data on potential homologies between *Nematostella* and bilaterian Hox genes are from previous publications (12, 14).

enhancer region in this construct contains a previously identified *Meis/Pbx-Hox* binding site (25). Because Pbx is a key binding partner for bilaterian Hox genes that biochemically interacts with several Hox proteins in *Nematostella* (25–28), we tested whether *Pbx* regulates *Anthox8* expression. Larval *Anthox8* transcription was undetectable after *Pbx* shRNA KD, which caused a severe and uniform loss of endodermal segmentation and subsequent metamorphic failure (Fig. 4D and fig. S8, K and L). To identify the putative Hox cofactor responsible for Pbx-dependent activation of *Anthox8*, we knocked down all Hox-Gbx genes and assayed *Anthox8>GFP* expression. Only shRNAs targeting *Anthox8* substantially reduced green fluorescent protein (GFP) intensity (Fig. 4E and fig. S15). Further validated by FISH experiments (fig. S16), these results demonstrate Pbx-dependent autoregulation of *Anthox8*. In parallel experiments to explore cross-regulatory interactions, we also found that *Anthox1a* unidirectionally repressed *Gbx* expression in developing endodermal segment s5 (fig. S17). Combined, these findings hint at the possibility of similar regulatory mechanisms between cnidarian and bilaterian Hox networks.

In summary, this work leverages both classical genetics and a robust gene KD methodology to demonstrate the existence of a functional Hox code in a developing cnidarian. Reminiscent of the sophisticated Hox networks that operate in arthropods and chordates, *Nematostella* Hox-Gbx genes encode axial identities, thus governing the patterning of secondary structures such as tentacles and mesenteries (Fig. 4F). *Hox-Gbx*-dependent endodermal segments may be established in a manner analogous to bilaterian posterior prevalence, whereby posteriorly expressed Hox genes generally override the effects of genes that are more anterior (29–31). According to this logic, *Anthox1a* expression in *Nematostella* segment

s5 would reflect a dominant pole of the anthozoan directive axis, with *Anthox8*, *Anthox6a*, and *Gbx* operating to define successive segment boundaries toward the opposite end (fig. S18). Although understanding the direct or indirect nature of these Hox-Gbx interactions will be an important area for future studies, our findings further demonstrate the existence of a Pbx-dependent functional network and provide initial evidence for both Hox auto- and cross-regulation (Fig. 4). Despite limited data regarding other cnidarian species, the phylogenetically basal position of Anthozoa (32) permits speculation that the *Nematostella* Hox code reflects a conserved gene regulatory module, one that could have been co-opted to direct A-P patterning in the ancient urbilaterian.

## REFERENCES AND NOTES

1. N. H. Putnam et al., *Science* **317**, 86–94 (2007).
2. C. Nielsen, *Animal evolution: Interrelationships of the living phyla* (Oxford University Press, ed. 3, ebook, 2012), p. 1.
3. B. Galliot, V. Schmid, *Int. J. Dev. Biol.* **46**, 39–48 (2002).
4. U. Technau, R. E. Steele, *Development* **138**, 1447–1458 (2011).
5. L. H. Hyman, *The invertebrates* (Publications in the Zoological Sciences, McGraw-Hill, ed. 1, 1940), p. v.
6. J. Garcia-Fernández, *Nat. Rev. Genet.* **6**, 881–892 (2005).
7. D. Lemons, W. McGinnis, *Science* **313**, 1918–1922 (2006).
8. E. B. Lewis, *Nature* **276**, 565–570 (1978).
9. M. Kessel, P. Gruss, *Science* **249**, 374–379 (1990).
10. J. R. Finnerty, K. Pang, P. Burton, D. Paulson, M. Q. Martindale, *Science* **304**, 1335–1337 (2004).
11. D. Chourrout et al., *Nature* **442**, 684–687 (2006).
12. R. Chiori et al., *PLOS ONE* **4**, e4231 (2009).
13. J. R. Finnerty, M. Q. Martindale, *Biol. Bull.* **193**, 62–76 (1997).
14. J. F. Ryan et al., *PLOS ONE* **2**, e153 (2007).
15. G. Genikhovich et al., *Cell Rep.* **10**, 1646–1654 (2015).
16. Y. Moran et al., *Genome Res.* **24**, 651–663 (2014).
17. J. Huiskens, J. Swoger, F. Del Bene, J. Wittbrodt, E. H. Stelzer, *Science* **305**, 1007–1009 (2004).
18. A. Ikmi, S. A. McKinney, K. M. Delventhal, M. C. Gibson, *Nat. Commun.* **5**, 5486 (2014).
19. Y. Kraus, A. Aman, U. Technau, G. Genikhovich, *Nat. Commun.* **7**, 11694 (2016).
20. M. D. Servetnick et al., *Development* **144**, 2951–2960 (2017).
21. M. Mallo, C. R. Alonso, *Development* **140**, 3951–3963 (2013).
22. V. Zappavigna et al., *EMBO J.* **10**, 4177–4187 (1991).
23. H. Pöpperl et al., *Cell* **81**, 1031–1042 (1995).
24. H. Pöpperl, M. S. Featherstone, *EMBO J.* **11**, 3673–3680 (1992).
25. B. Hudry et al., *eLife* **3**, e01939 (2014).
26. H. Asahara, S. Dutta, H. Y. Kao, R. M. Evans, M. Montminy, *Mol. Cell. Biol.* **19**, 8219–8225 (1999).
27. R. S. Mann, S.-K. Chan, *Trends Genet.* **12**, 258–262 (1996).
28. M. Slattery et al., *Cell* **147**, 1270–1282 (2011).
29. A. González-Reyes, G. Morata, *Cell* **61**, 515–522 (1990).
30. D. Duboule, *Curr. Opin. Genet. Dev.* **1**, 211–216 (1991).
31. D. Duboule, G. Morata, *Trends Genet.* **10**, 358–364 (1994).
32. E. Kayal, B. Roue, H. Philippe, A. G. Collins, D. V. Lavrov, *BMC Evol. Biol.* **13**, 5 (2013).

## ACKNOWLEDGMENTS

We thank R. Krumlauf (Stowers Institute), P. Cartwright (University of Kansas), and D. Lambert (University of Rochester) for suggestions and critical reading of the manuscript. We also thank M. Kirkman and K. Delventhal for genotyping assistance and the Stowers Institute Aquatics Core facility for animal husbandry. **Funding:** This study was supported by the Stowers Institute for Medical Research. **Author contributions:** S.H. and M.C.G. designed and analyzed the experiments. S.H. developed the shRNA approach and performed all RNA interference experiments. A.I. generated the *Actin>GFP* transgenic line and tested shRNA perdurance. S.H. generated the *Anthox8>GFP* transgenic line. S.H. and F.D.V. performed CRISPR-Cas9 genome editing. C.-Y.C. optimized the FISH protocol and performed time-course reverse transcription-qPCR analysis. S.H. and C.-Y.C. performed the FISH experiments. A.E.K., A.I., and S.H. performed SPIM imaging and SPIM data analysis. S.H. and M.C.G. wrote the manuscript. All authors discussed the experiments and read and approved the manuscript. **Competing interests:** None declared. **Data and materials availability:** Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at <http://www.stowers.org/research/publications/116pb-1247>. 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/361/6409/1377/suppl/DC1](http://www.sciencemag.org/content/361/6409/1377/suppl/DC1)  
Materials and Methods  
Figs. S1 to S18  
Tables S1 and S2  
References (33–46)  
Movies S1 to S5

21 December 2017; accepted 9 August 2018  
10.1126/science.aar8384

## SINGLE-CELL GENOMICS

# Joint profiling of chromatin accessibility and gene expression in thousands of single cells

Junyue Cao<sup>1,2</sup>, Darren A. Cusanovich<sup>1\*†</sup>, Vijay Ramani<sup>1\*</sup>, Delasa Aghamirzaie<sup>1</sup>, Hannah A. Pliner<sup>1</sup>, Andrew J. Hill<sup>1</sup>, Riza M. Daza<sup>1</sup>, Jose L. McFaline-Figueroa<sup>1</sup>, Jonathan S. Packer<sup>1</sup>, Lena Christiansen<sup>3</sup>, Frank J. Steemers<sup>3</sup>, Andrew C. Adey<sup>4,5</sup>, Cole Trapnell<sup>1,6,7‡</sup>, Jay Shendure<sup>1,6,7,8‡</sup>

Although we can increasingly measure transcription, chromatin, methylation, and other aspects of molecular biology at single-cell resolution, most assays survey only one aspect of cellular biology. Here we describe sci-CAR, a combinatorial indexing–based coassay that jointly profiles chromatin accessibility and mRNA (CAR) in each of thousands of single cells. As a proof of concept, we apply sci-CAR to 4825 cells, including a time series of dexamethasone treatment, as well as to 11,296 cells from the adult mouse kidney. With the resulting data, we compare the pseudotemporal dynamics of chromatin accessibility and gene expression, reconstruct the chromatin accessibility profiles of cell types defined by RNA profiles, and link cis-regulatory sites to their target genes on the basis of the covariance of chromatin accessibility and transcription across large numbers of single cells.

**T**he concurrent profiling of multiple classes of molecules—for example, RNA and DNA—within single cells has the potential to reveal causal regulatory relationships and to enrich the utility of organism-scale single-cell atlases. However, to date, nucleic acid “coassays” rely on physically isolating each cell, limiting their throughput to a few cells per study (fig. S1A and table S1) (1–6).

Single-cell combinatorial indexing (sci) methods use split-pool barcoding to uniquely label the nucleic acid contents of single cells or nuclei (7–13). Here we describe sci-CAR, which jointly profiles single-cell chromatin accessibility and mRNA (CAR) in a scalable fashion. sci-CAR effectively combines sci-ATAC sequencing (sci-ATAC-seq) and sci-RNA-seq into a single protocol (Fig. 1) by the following steps: (i) Nuclei are extracted, with or without fixation, and distributed to wells. (ii) A first RNA-seq “index” is introduced by in situ reverse transcription (RT) with a polythymidine [poly(T)] primer that bears a well-specific barcode and a unique molecular identifier (UMI). (iii) A first ATAC-seq index is

introduced by in situ tagmentation with Tn5 transposase that bears a well-specific barcode. (iv) All nuclei are pooled and redistributed by fluorescence-activated cell sorting to multiple plates. (v) After second-strand synthesis of cDNA, nuclei in each well are lysed, and the lysate is split into RNA- and ATAC-dedicated portions. (vi) To provide a second priming site for amplification of 3′ cDNA tags, the RNA-dedicated lysate is subjected to transposition with undigested Tn5 transposase. 3′ cDNA tags are amplified with primers corresponding to the Tn5 adaptor and RT primer. These primers also bear a well-specific barcode that is the second RNA-seq index. (vii) The ATAC-seq–dedicated lysate is amplified with primers specific to the barcoded Tn5 adaptors from (iii). These primers also bear a well-specific barcode that is the second ATAC-seq index. (viii) Amplicons from RNA-seq- and ATAC-seq–dedicated lysates are respectively pooled and sequenced. Each sequence read is associated with two barcodes that correspond to each round of indexing. As with other sci protocols, most nuclei pass through a unique combination of wells, thereby receiving a unique combination of barcodes that can be used to group reads derived from the same cell. Because the barcodes introduced to RNA-seq and ATAC-seq libraries correspond to specific wells, we can link the mRNA and chromatin accessibility profiles of individual cells.

We applied sci-CAR to a cell-culture model of cortisol response, wherein dexamethasone (DEX), a synthetic mimic of cortisol, activates glucocorticoid receptor (GR), which binds to thousands of locations across the genome, altering the expression of hundreds of genes (14–17). We collected human lung adenocarcinoma–derived

A549 cells after 0, 1, or 3 hours of 100 nM DEX treatment and performed a 96 well (first round indexing)–by–576 well (second round indexing) sci-CAR experiment. The three time points were each represented by 24 wells during the first round of indexing, whereas the remaining 24 wells contained a mixture of human embryonic kidney (HEK) 293T and NIH/3T3 (mouse) cells (fig. S1B).

We obtained sci-RNA-seq profiles for 6093 cells (median 3809 UMIs) and sci-ATAC-seq profiles for 6085 cells (median 1456 unique reads) (fig. S1, C to E). For both data types, reads assigned to the same cell overwhelmingly mapped to one species (fig. S1, F and G). We obtained roughly equivalent UMIs per cell from “RNA-only” plates processed in parallel, albeit at a lower sequencing depth per cell. Aggregated transcriptomes of coassayed versus RNA-only plates were well correlated [Pearson correlation coefficient ( $r$ ) = 0.97 to 0.98; fig. S2]. By contrast, although coassayed versus “ATAC-only” plates were similar in data quality and well correlated in aggregate (fig. S3), ATAC-only plates had ~10-fold higher molecular complexity. The lower efficiency of the coassay for ATAC is likely explained by factors including buffer modifications and our use of only half the lysate.

There were 4825 cells (70% of either set) for which we recovered both transcriptome and chromatin accessibility data. To confirm that paired profiles truly derived from the same cells, we asked whether cells from mixed human-mouse wells were consistently assigned as human or mouse. Indeed, 1423/1425 (99%) of coassayed cells from those wells were assigned the same species label from both sci-RNA-seq and sci-ATAC-seq profiles (Fig. 2A).

We next examined the time course of GR activation. DEX treatment of A549 cells increased both transcription and promoter accessibility of markers of GR activation, including genes *NFKB1A*, *SCNN1A*, *CKB*, *PER1*, and *CDH16* (14, 16) (fig. S4, A and B). Unsupervised clustering or *t*-distributed stochastic neighbor embedding (t-SNE) visualization of either sci-RNA-seq or sci-ATAC-seq profiles readily separated clusters corresponding to untreated and DEX-treated cells (Fig. 2, B and C). Reassuringly, cells from coassay plates and single-assay plates of either type were intermixed (fig. S4C).

Of coassayed cells in clusters 1 and 2 of sci-ATAC-seq data, 88 and 93% were found in corresponding sci-RNA-seq clusters (fig. S4, D and E). Cells with concordant versus discordant assignments did not significantly differ in read depth ( $P > 0.1$ , Welch two-sample *t* test) but notably fell on the border between clusters 1 and 2 in either t-SNE (Fig. 2D and fig. S4F). Whereas most discordant cells (70%) were from 0 hours, the remainder tended to derive from 1 hour rather than 3 hours (5% of 1-hour cells versus 1% of 3-hour cells,  $P = 2.2 \times 10^{-16}$ , Fisher’s exact test). Although we cannot rule out that this is due to imperfect clustering, these discordantly assigned cells potentially reflect transitional states in GR activation.

Differential expression (DE) analysis of sci-RNA-seq data revealed significant changes in

<sup>1</sup>Department of Genome Sciences, University of Washington, Seattle, WA, USA. <sup>2</sup>Molecular and Cellular Biology Program, University of Washington, Seattle, WA, USA. <sup>3</sup>Illumina Inc., San Diego, CA, USA. <sup>4</sup>Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA. <sup>5</sup>Knight Cardiovascular Institute, Portland, OR, USA. <sup>6</sup>Allen Discovery Center for Cell Lineage Tracing, Seattle, WA, USA. <sup>7</sup>Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. <sup>8</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.

\*These authors contributed equally to this work.

†Present address: Asthma & Airway Disease Research Center and Department of Cellular and Molecular Medicine, The University of Arizona, Tucson, AZ, USA.

‡Corresponding author. Email: coletrap@uw.edu (C.T.); shendure@uw.edu (J.S.)



2613 genes [5% false discovery rate (FDR)] (table S2). For comparison, a similar analysis with bulk RNA-seq data of DEX treatment in A549 cells at 0 versus 3 hours (18) identified 870 DE genes, 536 of which were also DE here. Log<sub>2</sub> fold changes were well correlated between the datasets for DE genes ( $r = 0.86$ , fig. S4G).

Differential accessibility (DA) analysis of sci-ATAC-seq profiles identified significant changes at 4763 sites (5% FDR) (table S3). For comparison, a similar analysis of bulk deoxyribonuclease (DNase)-seq data from DEX-treated A549 cells at 0 versus 3 hours (18) identified 672 DA sites, 544 of which were also DA here. Log<sub>2</sub> fold changes were well correlated between the datasets for DA sites [Spearman's rank correlation coefficient ( $\rho$ ) = 0.68, fig. S4H].

Of our DA sites, 701 (15%) were promoters, of which 175 overlapped with DE transcripts. Transcripts for genes with DA promoters that were not DE were detected in significantly fewer cells than genes with DA promoters that were DE (median 10 versus 25%,  $P < 5 \times 10^{-5}$ , unpaired two-sample permutation test based on 20,000 simulations), suggesting that we may be insufficiently powered to detect DE at many genes with DA promoters. For the 175 genes that are both DA and DE, the log<sub>2</sub> fold changes were modestly correlated ( $\rho = 0.63$ , fig. S4I), and 130/175 (74%) exhibited directional concordance (exact two-sided binomial test,  $P = 9 \times 10^{-11}$ ).

We ordered cells along a pseudotime trajectory with Monocle (19) based on the top 1000 DE genes (fig. S5A). Cells were ordered consistently with the time course (Fig. 2E). Of note, the aforementioned cells from 1 hour whose cluster assignments were discordant (Fig. 2D and fig. S4F) occurred significantly earlier in pseudotime than cells with concordant assignments ( $P = 3 \times 10^{-5}$ , Wilcoxon rank sum test, fig. S5B). Of the 2613 DE genes, 979 (37%) increased and 1111 (43%) decreased in expression along pseudotime, whereas 523 (20%) exhibited transient changes (fig. S5, C and D, and tables S2 and S4). We exploited the coassay to examine the dynamics of chromatin accessibility across RNA-defined pseudotime, identifying opening (47%), closing (32%), and transient (21%) DA sites (fig. S5E and tables S3 and S5). There were 11 genes that showed significant changes in both gene expression and promoter accessibility along pseudotime (5% FDR for both), with well-correlated dynamics (Fig. 2F and fig. S5, F to H).

We converted the ATAC-seq (cell-by-site) matrix to a [cell-by-transcription factor (TF) motif] matrix, simply by counting occurrences of each motif in all accessible sites for each cell (20). The motifs of 91/399 (23%) of expressed TFs were DA across the treatment conditions (5% FDR) (tables S6 and S7). Where chromatin immunoprecipitation (ChIP) sequencing data was available for the same time course (18), we observed consistent dynamics of increasing motif-associated accessibility (fig. S6A) and TF binding to accessible sites (fig. S6B). Motif-accessibility dynamics across expression-defined pseudotime are summarized in fig. S6C. The motif of the canonical GR *NR3C1* was the most

activated, even though its expression decreased (Fig. 2G), consistent with its activation by recruitment from the cytosol rather than by increased expression. By contrast, *KLF9* is a direct target of GR activation via a feed-forward loop (21). Consistent with this, we observed that both its expression and its motif accessibility increase along pseudotime (Fig. 2G and fig. S6, D and E).

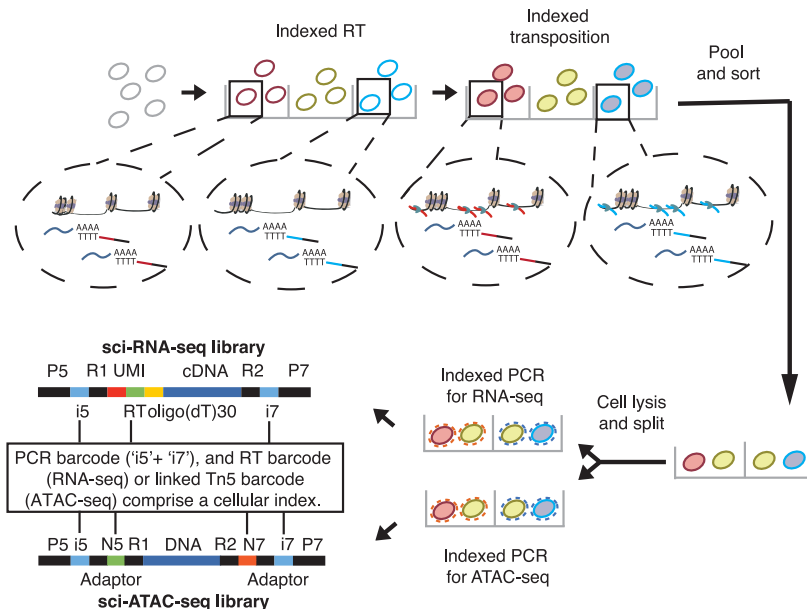
Single-cell RNA-seq studies have recently characterized the transcriptomes of diverse cell types represented in the mammalian kidney (22–24). However, little is known about the epigenetic landscapes that underlie these cell type-specific gene expression programs. To investigate this, we isolated and fixed nuclei from whole kidneys of two 8-week-old male mice (fig. S7A). From one sci-CAR experiment, we obtained sci-RNA-seq profiles for 13,893 nuclei (median 1011 UMIs; fig. S7B) and sci-ATAC-seq profiles for 13,395 nuclei (median 7987 unique reads; fig. S7C). There were 11,296 cells for which we recovered both transcriptome and chromatin accessibility profiles.

We compared sci-CAR transcriptomes with a recently published single-cell RNA-seq dataset of the same tissue generated by Drop-seq (24). After correcting for gene-length biases (Drop-seq is biased toward shorter transcripts), and sci-RNA-seq toward longer transcripts), aggregated transcriptomes were reasonably well correlated ( $r = 0.73$ , fig. S7D). Semisupervised clustering of 10,727 sci-CAR transcriptomes (>500 UMIs) identified 14 groups, ranging in size from 74 (0.7%) to 2358 (22.0%) cells (Fig. 3A and fig. S7, E and F). Es-

tablished markers identified nearly all expected cell types (fig. S8, A and B). The expression profiles of proximal tubule cells separate them into three subtypes, including S1/S2 cells (*Slc5a12*+, *Gatm*+, *Alpl*+, and *Slc34a1*+, S3 type 1 cells (*Slc34a1*+ and *Atp11a*+), and S3 type 2 cells (*Atp11a*+ and *Rnf24*+) (fig. S8C) (25, 26). The smallest cluster is positive for cell-cycle progression markers (*Mki67* and *Cenpp*) and may represent an actively proliferating subpopulation (fig. S8D) (25, 26). Cell-type proportions were well correlated between replicate kidneys, with the exception of paranephric body adipocytes (1.2 versus 0.4%), likely owing to technical variation in kidney dissection because these reside superficial to the renal fascia (fig. S7E).

We identified 8774 genes that were DE across the 14 cell types (5% FDR), including 1771 with more than twofold greater expression in the highest versus second-highest cell type (fig. S9, A and B, and tables S8 and S9). New marker genes were identified, such as *Daam2* for renal pericytes and *Calcr* for collecting duct intercalated cell B (fig. S9, C and D) (25, 26). We examined the expression of solute carrier transporters, because these are essential to a principal function of the kidney. Of these, 208/345 (60%) were DE in subsets of renal tubule cell types, many corresponding to known and potentially as yet uncharacterized reabsorption specificities (Fig. 3B, fig. S9E, and table S10).

We compared aggregated sci-CAR chromatin accessibility profiles with published bulk ATAC-seq data on adult mouse kidney (18) and found them to be reasonably well correlated ( $r = 0.75$ ;



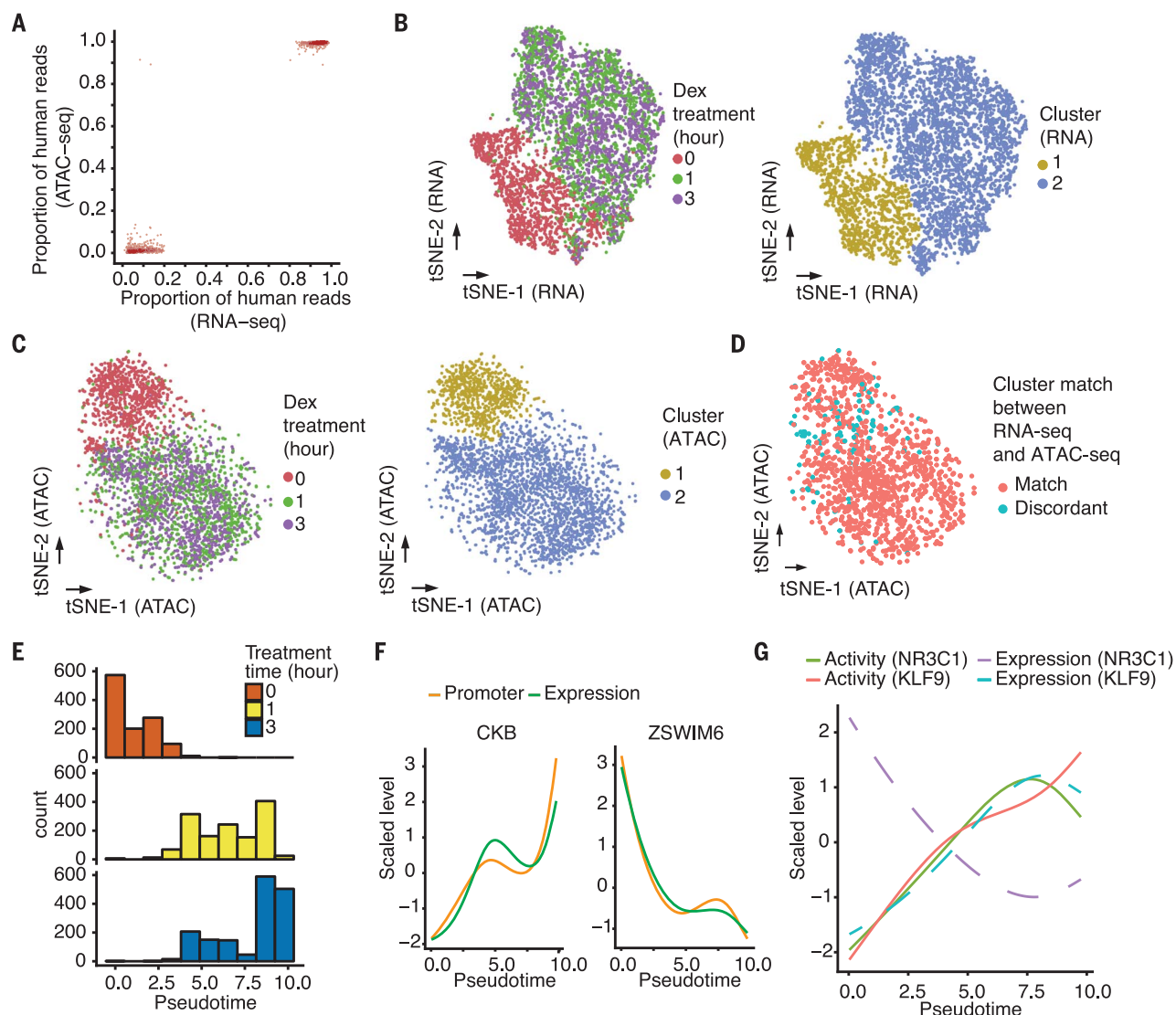
**Fig. 1. sci-CAR workflow.** For a more detailed explanation of key steps, see the text. From extracted nuclei, a first RNA-seq index is introduced by RT and a first ATAC-seq index by transposition. The nuclei are pooled and sorted, cDNA is synthesized, and nuclei are lysed. The lysate is split into portions for RNA-seq and ATAC-seq. For RNA-seq, index2 and read1 cover the i5 index, UMI, and RT barcode, and index1 and read2 cover the i7 index and cDNA fragment. For ATAC-seq, read1 and read2 cover the genomic DNA sequence, and index 1 and index 2 cover the Tn5 and polymerase chain reaction (PCR) barcodes. P, Illumina P5 or P7 adaptor sequence; R, annealing sites for Illumina sequencing primers; i, Illumina sequencing index; N, Tn5 transposase index.

fig. S10, A and B). Across all genes, aggregate promoter accessibility correlated with aggregate gene expression ( $\rho = 0.26$ ; fig. S10C). Nonetheless, a considerable challenge for single-cell ATAC-seq data, relative to single-cell RNA-seq data, is the sparsity of the resulting matrices (8). Thus, our initial efforts to cluster coassayed cells solely on the basis of their ATAC-seq profiles failed to discover the expected diversity of cell types. We therefore sought to leverage the coassay aspect of these data to recover the chromatin landscapes of individual cell types.

As a first approach, we simply annotated cell types from transcriptional profiles for ~96% of the 11,296 cells that were successfully coassayed. We then aggregated ATAC-seq signals for each cell type separately, followed by peak calling (27). As a second approach, we also developed an algorithm to combine the ATAC-seq profiles of cells with highly similar RNA-seq profiles before clustering (fig. S7A). For cells from each RNA-seq-defined cell type, we identified subsets of cells with highly similar expression profiles (a mean of 50 cells assigned to each of 222

“pseudocells”). We then aggregated the ATAC-seq profiles of each pseudocell and performed t-SNE on these. In contrast with single-cell ATAC-seq data, pseudocell chromatin accessibility profiles corresponding to the same cell types clustered together (Fig. 3C). Overall, these analyses illustrate how coassay data can be leveraged to overcome the relative sparsity of single-cell ATAC-seq data and define chromatin accessibility profiles even for closely related cell types.

We identified 22,026 DA sites across the 14 mouse kidney cell types, including 2096 promoters



**Fig. 2. Joint profiling of chromatin accessibility and transcription in DEX-treated A549 cells.** (A) Scatter plot showing the proportion of human reads, out of all reads that map uniquely to the human or mouse reference genomes, for cells in which both RNA-seq profiles and ATAC-seq profiles were obtained. Only HEK293T (human) and NIH/3T3 (mouse) cells are plotted. (B) t-SNE visualization of A549 cells (RNA-seq) including cells from both sci-CAR and sci-RNA-seq-only plates, colored by DEX treatment time (left) or unsupervised clustering ID (right). (C) t-SNE visualization of A549 cells (ATAC-seq) including cells from both sci-CAR and sci-ATAC-seq-only plates, colored by DEX treatment time (left) or unsupervised clustering ID (right). (D) t-SNE

visualization of A549 cells (ATAC-seq) with linked RNA-seq profiles. If the cell is in cluster 1 (or cluster 2) in both RNA-seq and ATAC-seq, then it is labeled as “Match,” otherwise it is labeled “Discordant.” (E) Distribution of cells from different DEX-treatment time points in gene expression pseudotime inferred by trajectory analysis. Pseudotime units are arbitrary. (F) Smoothed line plot showing scaled (with the R function scale) gene expression and promoter accessibility of *CKB* and *ZSWIM6* across pseudotime. The unscaled, unsmoothed data are shown in fig. S5, F and G. (G) Smoothed line plot showing the scaled mRNA level and activity change of transcription factors NR3C1 and KLF9 across pseudotime. The unscaled, unsmoothed data are shown in fig. S6, D and E.

and 19,930 distal sites (5% FDR; Fig. 3D; fig. S10, D and E; and tables S11 and S12). In some cases, DA at a gene's promoter was concordant with DE (fig. S11, A and B), but this was the exception rather than the rule. Out of 2096 genes with a DA promoter in at least one cell type, 132 genes were also DE (1% FDR) with a greater than twofold difference between the first- and second-ranked cell type. Although promoter accessibility and expression of these genes across cell types are positively correlated (median  $\rho = 0.17$ ), most (112/132 or 85%) exhibited maximal promoter accessibility and gene expression in different cell types (fig. S11C). The relatively weaker correlation, compared with what we observed in the A549 DEX time series ( $\rho = 0.63$ ; fig. S4I), is potentially a consequence of the fact that, in the A549 cells, we were comparing changes in

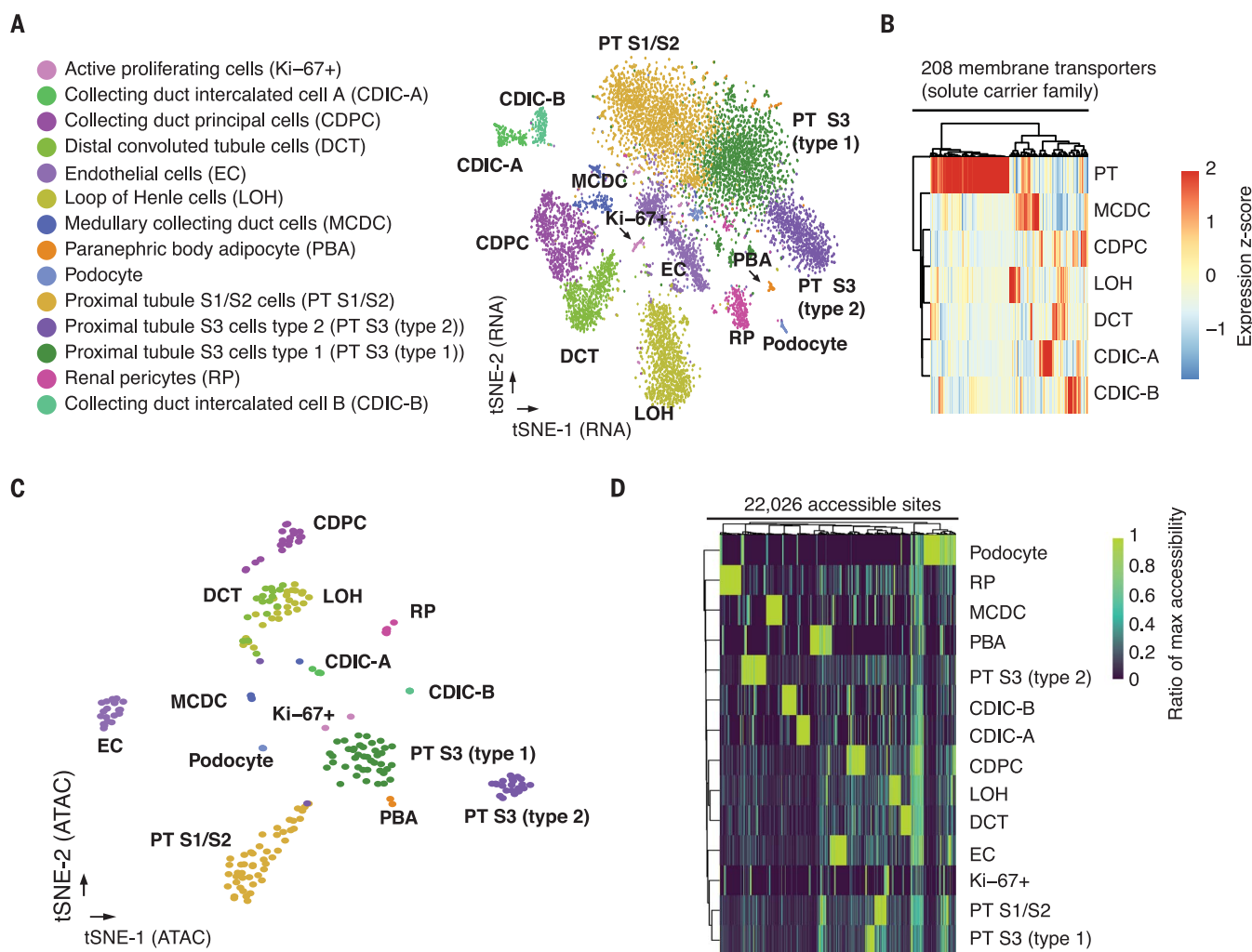
promoter accessibility versus expression, whereas here we are comparing absolute enrichment of accessibility at promoters versus expression.

We sought to link distal cis-regulatory elements to their target genes on the basis of the covariance of chromatin accessibility and gene expression across large numbers of coassayed cells. As the sparsity of our single-cell profiles makes this challenging, we worked with the previously described 222 pseudocells (fig. S12A). For each gene, we computed correlations between its expression and the adjusted accessibility of all sites within 100 kb of its transcriptional start site (TSS) using LASSO (least absolute shrinkage and selection operator).

Within the top 2000 DE genes (ranked by  $q$  value), we linked 1260 distal sites to 321 genes (median three sites per gene, out of a median

19 sites within 100 kb of the TSS tested; fig. S12, B and C, and table S13). Of the sites, 44% were linked to the nearest TSS and 21% to the second-nearest TSS (fig. S12D). Distal site-gene linkages were significantly closer than all possible pairs tested (mean 41 kb for links versus 48 kb for all pairs tested;  $P < 5 \times 10^{-5}$ , unpaired permutation test based on 20,000 simulations; fig. S12E).

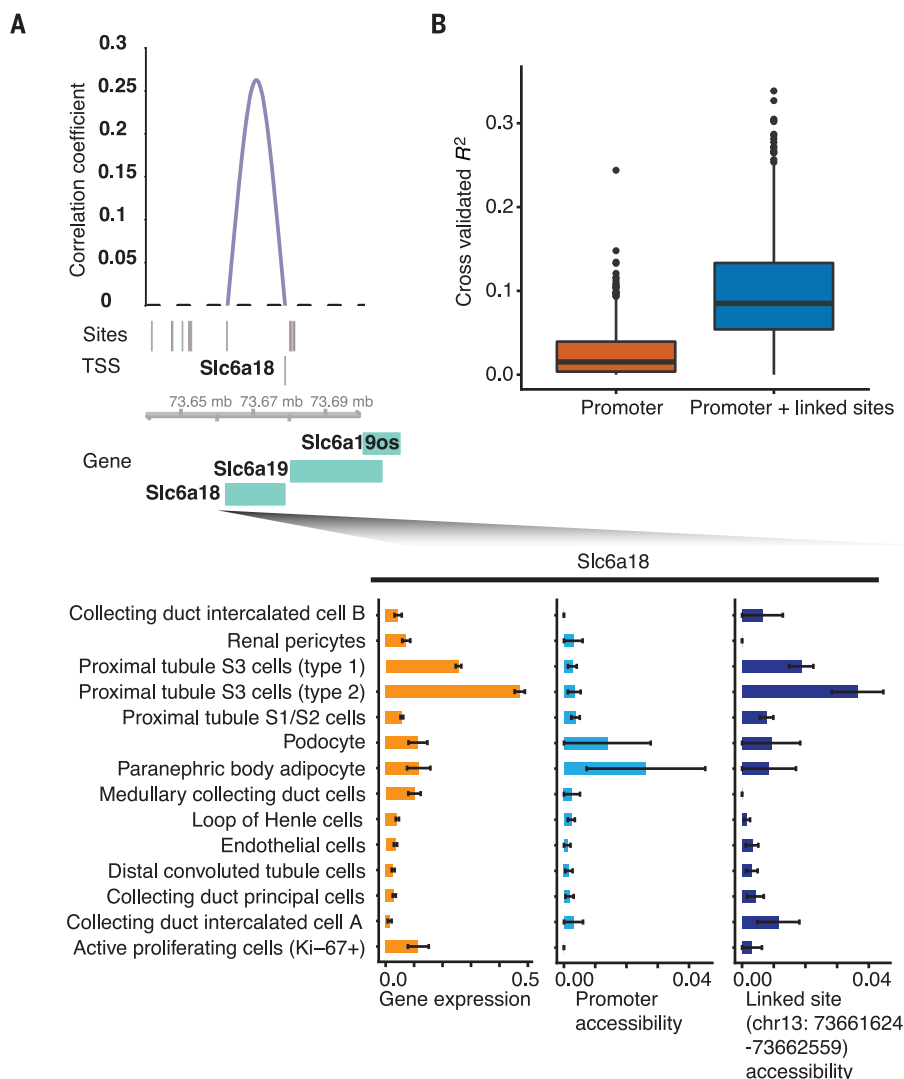
To evaluate the possibility that the links were artifacts of regularized regression, we permuted the sample IDs of the chromatin accessibility matrix and performed the same analysis. After this permutation, only four links were identified (fig. S12B). To control for correlations between closely located accessible sites in the genome, we separately permuted the peak IDs. This yielded 216 links, or just 17% as many links as without permutation (fig. S12B).



**Fig. 3. sci-CAR enables joint profiling of chromatin accessibility and transcription in mouse kidney.** (A) t-SNE visualization of mouse kidney nuclei (RNA-seq). Cell types are assigned on the basis of established marker genes. (B) Heatmap showing the relative expression of genes from the solute carrier group of membrane transport proteins in consensus transcriptomes of each cell type estimated by RNA-seq data from the coassay. The raw expression data (UMI count matrix) was log-transformed, column-centered, and scaled (by using the R function scale), and the

resulting values clamped to  $(-2, 2)$ . (C) t-SNE visualization of mouse kidney nuclei (ATAC-seq) after aggregating cells with highly similar transcriptomes (pseudocells), colored by cell types identified from RNA-seq. (D) Heatmap showing the relative chromatin accessibility of cell type-specific sites for each cell type estimated by ATAC-seq data from the coassay. The raw aggregated ATAC-seq data (read count matrix) was normalized first by the total number of reads for each cell type and then by the maximum accessibility score across all cell types.





**Fig. 4. Linking cis-regulatory elements to regulated genes on the basis of covariance in single-cell coassay data.**

(A) Top: Genome browser plot showing links between accessible distal regulatory sites and the gene *Slc6a18*. The height corresponds to the correlation coefficient. Bottom: Bar plots showing the average expression, promoter accessibility, and linked site accessibility for cell type-specific marker gene *Slc6a18* across different cell types. Gene expression values for each cell were calculated by dividing the raw UMI count by cell-specific size factors. Site accessibilities for each cell were calculated by dividing the raw read count by cell-specific size factors. Error bars represent SEM. chr13, chromosome 13. (B) Two linear regression models were built to predict gene expression differences between cell types. The first model predicts changes on the basis of promoter accessibility alone. The second model predicts changes on the basis of the chromatin accessibility of the promoter and distal sites that are linked to it. The boxplot shows the cross-validated coefficient of determination ( $R^2$ ) calculated for each gene from the two models.

The 321 genes with linked distal sites were specifically expressed in a variety of cell types (fig. S12F). For example, the link with the highest correlation is between distal convoluted tubule cell marker gene *Slc12a3* and a site 36-kb downstream of its TSS that overlaps its last exon (fig. S13). The accessibility of this linked site was modestly more specific to distal convoluted tubule cells than the *Slc12a3* promoter. By contrast, the accessible site closest to the *Slc12a3* promoter (only 216 base pairs away) was not linked to the *Slc12a3* promoter by our approach nor is its accessibility specific to distal convoluted tubule cells. Similarly, a marker gene for loop of Henle cells, *Slc12a1*, is linked to two distal sites (fig. S14), both of which exhibit accessibility specific to loop of Henle cells. By contrast, the nearest accessible site (9 kb from the TSS), which was not linked, does not exhibit this specificity.

Links between distal cis-regulatory elements and their target genes can be useful for explaining differential expression across cell types. For example, the cell type-specific expression of *Slc6a18*, a marker gene for type 2 proximal tubule S3 cells, is not mirrored by cell type-specific pro-

motor accessibility (fig. S11C). However, from our covariance approach, its TSS is linked to a site 16 kb away whose accessibility is correlated with *Slc6a18* expression (Fig. 4A). To quantify the utility of the links between distal cis-regulatory elements and their target genes identified from sci-CAR data, we constructed a linear regression model to predict gene expression differences based on chromatin accessibility at promoters only versus promoters together with linked distal sites. Including linked distal sites improved predictions by fourfold ( $P < 5 \times 10^{-5}$ , paired permutation test based on 20,000 simulations; Fig. 4B).

Our analyses illustrate the advantages of a single-cell coassay over assays that solely profile transcription or chromatin accessibility. sci-CAR is compatible with fresh or fixed nuclei and, like other sci-seq techniques, can encode multiple samples per experiment. Its throughput can potentially be increased by additional rounds of split-pool indexing (13). With 384-well-by-384-well-by-384-well sci-CAR, one could potentially coassay millions of single cells per experiment. A limitation of sci-CAR is the sparsity of the resulting

data, particularly with respect to chromatin accessibility. This can potentially be overcome in the future through protocol optimizations, particularly of cross-linking conditions. A second limitation is that, although we were able to link distal elements and target genes on the basis of covariance of accessibility and expression, these data remain correlative and involve a minority of DE genes and DA elements.

Notwithstanding these limitations, sci-CAR expands the potential of combinatorial indexing for scalably profiling single-cell molecular phenotypes and may be particularly useful in the context of organism-scale single-cell atlases. With further development, we anticipate that additional DNA and RNA coassays may be realized by simply integrating other sci-seq protocols together with sci-RNA-seq (e.g., methylation plus transcripts, chromosome conformation plus transcripts, or DNA sequence plus transcripts) (8–13). A longer-term goal is to adapt single-cell combinatorial indexing to span the central dogma, such that aspects of DNA, RNA, and protein species can be concurrently assayed from each of many single cells.

## REFERENCES AND NOTES

1. S. J. Clark *et al.*, *Nat. Commun.* **9**, 781 (2018).
2. C. Angermueller *et al.*, *Nat. Methods* **13**, 229–232 (2016).
3. Y. Hou *et al.*, *Cell Res.* **26**, 304–319 (2016).
4. Y. Hu *et al.*, *Genome Biol.* **17**, 88 (2016).
5. S. Pott, *eLife* **6**, e23203 (2017).
6. F. Guo *et al.*, *Cell Res.* **27**, 967–988 (2017).
7. S. Amini *et al.*, *Nat. Genet.* **46**, 1343–1349 (2014).
8. D. A. Cusanovich *et al.*, *Science* **348**, 910–914 (2015).
9. V. Ramani *et al.*, *Nat. Methods* **14**, 263–266 (2017).
10. Y. Yin *et al.*, bioRxiv 338053 [Preprint], 12 June 2018. <https://doi.org/10.1101/338053>.
11. S. A. Vitak *et al.*, *Nat. Methods* **14**, 302–308 (2017).
12. R. M. Mulqueen *et al.*, *Nat. Biotechnol.* **36**, 428–431 (2018).
13. J. Cao *et al.*, *Science* **357**, 661–667 (2017).
14. T. E. Reddy *et al.*, *Genome Res.* **19**, 2163–2171 (2009).
15. S. John *et al.*, *Nat. Genet.* **43**, 264–268 (2011).
16. T. E. Reddy, J. Gertz, G. E. Crawford, M. J. Garabedian, R. M. Myers, *Mol. Cell. Biol.* **32**, 3756–3767 (2012).
17. C. M. Vockley *et al.*, *Cell* **166**, 1269–1281.e19 (2016).
18. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
19. X. Qiu *et al.*, *Nat. Methods* **14**, 979–982 (2017).
20. J. D. Buenostro *et al.*, *Nature* **523**, 486–490 (2015).
21. Y. Chinenov, M. Coppo, R. Gupte, M. A. Sacta, I. Rogatsky, *BMC Genomics* **15**, 656 (2014).
22. L. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9989–E9998 (2017).
23. X. Han *et al.*, *Cell* **172**, 1091–1107.e17 (2018).
24. J. Park *et al.*, *Science* **360**, 758–763 (2018).
25. The Human Protein Atlas; [www.proteinatlas.org](http://www.proteinatlas.org).
26. M. Uhlen *et al.*, *Science* **357**, eaan2507 (2017).
27. Y. Zhang *et al.*, *Genome Biol.* **9**, R137 (2008).

## ACKNOWLEDGMENTS

We thank members of the Shendure and Trapnell labs for helpful discussions and feedback, particularly B. Martin, X. Qiu, A. Leith, A. Minkina, Y. Yin, Z. Duan, and R. Qiu, as well as R. Hunter and R. Rualo in the Transgenic Resources Program of University of Washington for their exceptional assistance.

**Funding:** This work was funded by the Paul G. Allen Frontiers Foundation (Allen Discovery Center grant to J.S. and C.T.), grants from the NIH (DP1HG007811 and R01HG006283 to J.S.; DP2 HD088158 to C.T.; R35GM124704 to A.C.A.), the W. M. Keck Foundation (to C.T. and J.S.), the Dale. F. Frey Award for Breakthrough Scientists (to C.T.), the Alfred P. Sloan Foundation Research Fellowship (to C.T.), and the Brotman Baty Institute for Precision Medicine. D.A.C. was supported in part by T32HL007828 from the National Heart, Lung, and Blood Institute. J.S. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** J.S. and C.T. designed and supervised the research; J.C. developed techniques and performed experiments with assistance from D.A.C., V.R., R.M.D., J.L.M.-F., L.C., F.J.S., and A.C.A.; J.C. performed computation analysis with assistance from D.A.C., V.R., D.A., H.A.P., A.J.H., and J.S.P.; J.S., C.T., and J.C. wrote the paper.

**Competing interests:** L.C. and F.J.S. declare competing financial interests in the form of stock ownership and paid employment by Illumina, Inc. One or more embodiments of one or more patents and patent applications filed by Illumina may encompass the methods, reagents, and data disclosed in this manuscript. Some work in this study may be related to technology described in the following exemplary published patent applications: WO2010/0120098 and WO2011/0287435.

**Data and materials availability:** Processed and raw data can be downloaded from NCBI GEO (GSE117089). All methods for making the transposase complexes are described in (7); however, Illumina will provide transposase complexes in response to reasonable requests from the scientific community subject to a material transfer agreement. The primary scripts for sci-CAR data processing are available at [https://github.com/JunyueC/sci-CAR\\_analysis](https://github.com/JunyueC/sci-CAR_analysis).

## SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/361/6409/1380/suppl/DC1](http://www.sciencemag.org/content/361/6409/1380/suppl/DC1)  
Materials and Methods  
Figs. S1 to S14  
Tables S1 to S13  
References (28–53)

3 May 2018; accepted 22 August 2018  
Published online 30 August 2018  
10.1126/science.aau0730

## MOLECULAR BIOLOGY

# A mechanism for preventing asymmetric histone segregation onto replicating DNA strands

Chuanhe Yu<sup>1\*</sup>, Haiyun Gan<sup>2\*</sup>, Albert Serra-Cardona<sup>2</sup>, Lin Zhang<sup>3,4</sup>, Songlin Gan<sup>3,4</sup>, Sushma Sharma<sup>5</sup>, Erik Johansson<sup>5</sup>, Andrei Chabes<sup>5</sup>, Rui-Ming Xu<sup>3,4</sup>, Zhiguo Zhang<sup>2†</sup>

How parental histone (H3-H4)<sub>2</sub> tetramers, the primary carriers of epigenetic modifications, are transferred onto leading and lagging strands of DNA replication forks for epigenetic inheritance remains elusive. Here we show that parental (H3-H4)<sub>2</sub> tetramers are assembled into nucleosomes onto both leading and lagging strands, with a slight preference for lagging strands. The lagging-strand preference increases markedly in budding yeast cells lacking Dpb3 and Dpb4, two subunits of the leading strand DNA polymerase, Pol  $\epsilon$ , owing to the impairment of parental (H3-H4)<sub>2</sub> transfer to leading strands. Dpb3-Dpb4 binds H3-H4 in vitro and participates in the inheritance of heterochromatin. These results indicate that different proteins facilitate the transfer of parental (H3-H4)<sub>2</sub> onto leading versus lagging strands and that Dpb3-Dpb4 plays an important role in this poorly understood process.

Posttranslational modifications (PTMs) on histones in eukaryotic chromatin have a profound impact on gene expression. Recently, it has been shown that at least some of these PTMs are traits that are inheritable during mitotic cell division and even through meiosis (1–4). As the “first” step of transmission of these PTMs, it was proposed that parental histone (H3-H4)<sub>2</sub> tetramers, the primary carriers of epigenetic modifications, are randomly and equally distributed to leading and lagging strands of DNA replication forks (5, 6) and serve as the “template” for copying epigenetic modifications onto newly synthesized (H3-H4)<sub>2</sub> tetramers, which do not mix with parental (H3-H4)<sub>2</sub> (7) and have distinct PTMs from parental histones. This dogmatic view has not been tested in vivo owing to challenges in monitoring histone segregation onto replicating DNA strands. Moreover, the molecular mechanisms underlying the transfer of parental histone (H3-H4)<sub>2</sub> tetramers onto replicating DNA remain largely unknown (8, 9).

We used the eSPAN (enrichment and sequencing of protein-associated nascent DNA) method, which can detect whether a protein is enriched on leading or lagging strands at a genome-wide scale (10), to monitor the segregation of newly synthesized and parental histone H3, which should represent (H3-H4)<sub>2</sub> tetramers, onto replicating DNA strands (Fig. 1, A and B, and fig. S1A).

Briefly, we released G<sub>1</sub>-arrested yeast cells into medium containing bromodeoxyuridine (BrdU) to label newly synthesized DNA and hydroxyurea (HU) to facilitate analysis (10, 11). Chromatin from G<sub>1</sub> and early S phase cells was digested with micrococcal nuclease (MNase), which cleaves DNA between nucleosomes, and was used for deep sequencing (MNase-seq). The digested chromatin was also analyzed by chromatin immunoprecipitation (ChIP)—with antibodies (fig. S1B) against acetylation of histone H3 lysine 56 (H3K56ac), a mark of newly synthesized H3 (12), and trimethylation of histone H3 lysine 4 (H3K4me3, a surrogate mark for parental H3, see below)—and subsequent ChIP—strand-specific sequencing (ChIP-ssSeq) and eSPAN analysis.

Analysis of MNase-seq data from G<sub>1</sub> and early S phase chromatin revealed well-positioned nucleosomes surrounding early replication origins (fig. S2, A to C), consistent with published results (13). Moreover, H3K56ac ChIP-ssSeq peaks colocalized with BrdU-IP-ssSeq (BrdU immunoprecipitation and strand-specific sequencing) peaks (fig. S2, A, D, and E). We chose H3K4me3 as the surrogate mark for parental H3 for the following reasons: First, H3K4me3 was not detected on newly synthesized H3 (14). Second, we observed that the H3K4me3 level at newly replicated regions, but not its total level, was reduced compared to that observed during the G<sub>1</sub> phase of the cell cycle (fig. S1, C and D), likely reflecting the dilution of parental H3K4me3 during the short time frame of our experiments. Finally, although at reduced occupancy, the positioning of H3K4me3-containing nucleosomes on newly replicated chromatin, as detected by H3K4me3 ChIP-ssSeq, was similar to that of G<sub>1</sub> phase (fig. S2, F and G).

We calculated the log<sub>2</sub> ratio of H3K56ac and H3K4me3 eSPAN sequence reads of the Watson (5' → 3') strand to those of the Crick (3' → 5') strand at 20 individual nucleosomes, which, on

average, span the replicated region surrounding each of the 134 early replication origins. If H3K56ac and H3K4me3 were equally distributed onto leading and lagging strands, one would expect that the ratio would be close to zero (Fig. 1B). Instead, we observed that H3K56ac eSPAN peaks exhibited a small, but consistent, leading-strand bias at most nucleosomes except the –1 and +1 nucleosomes (Fig. 1, C and D, and fig. S3A), whereas H3K4me3 eSPAN peaks showed a lagging-strand bias (Fig. 1, E and F, and fig. S3B). These bias patterns were more consistent at group 1 origins than at less-efficient group 2 origins (Fig. 1, C and E, and fig. S3, C and D). We estimated that about 8% more H3K56ac-H4 tetramers were deposited onto leading strands than lagging strands, whereas at least 23% more parental H3K4me3-H4 tetramers were transferred to lagging strands than leading strands. The differential enrichment of parental and newly synthesized (H3-H4)<sub>2</sub> tetramers at lagging and leading strands (23 versus 8%) observed here and in Figs. 2 and 3 likely suggests that nucleosomes that are formed from parental (H3-H4)<sub>2</sub> tetramers are more stable and resistant to MNase digestion than those formed from newly synthesized tetramers. Thus, parental (H3-H4)<sub>2</sub> tetramers are transferred onto both leading and lagging strands, with a slight preference for lagging strands.

Unexpectedly, we discovered that the deletion of genes *DPB3* and *DPB4* (*dpb3Δ* and *dpb4Δ* cells), which encode two nonessential subunits of the leading-strand DNA polymerase, Pol  $\epsilon$  (15), significantly increased the bias ratio of H3K56ac and H3K4me3 eSPAN peaks compared with wild-type cells (figs. S4 and S5) but had no apparent effect either on the levels of H3K56ac and H3K4me3 (fig. S1B) or on the overall nucleosome occupancy and positioning of G<sub>1</sub> and early S phase chromatin (fig. S6). We estimated that ~41% more H3K56ac-H4 was deposited onto leading strands than lagging strands, whereas 120% more parental H3K4me3-H4 was transferred to the lagging strand than to the leading strand in *dpb3Δ* and *dpb4Δ* cells.

We used the recombination-induced tag exchange (RITE) system (16), which marks newly synthesized and parental histone H3 with a T7 or hemagglutinin (HA) tag, respectively (Fig. 2A and fig. S7A), as an independent approach to analyze the impact of *DPB3* and *DPB4* deletion on histone segregation. The H3-T7 and H3K56ac eSPAN peaks showed a strong leading-strand bias (Fig. 2, B and C, and fig. S7, B and C) in *dpb3Δ* cells, whereas the H3-HA and H3K4me3 eSPAN peaks exhibited a lagging-strand bias (Fig. 2, D and E, and fig. S7, D and E). We also analyzed the distribution of H3K56ac and H3K4me3 at replicating chromatin in *dpb3Δ* cells during normal S phase without HU and observed the same effect (Fig. 3 and fig. S8). These results demonstrate that deletion of *DPB3* and *DPB4* alters the distribution pattern of new and parental (H3-H4)<sub>2</sub> tetramers at replicating DNA strands.

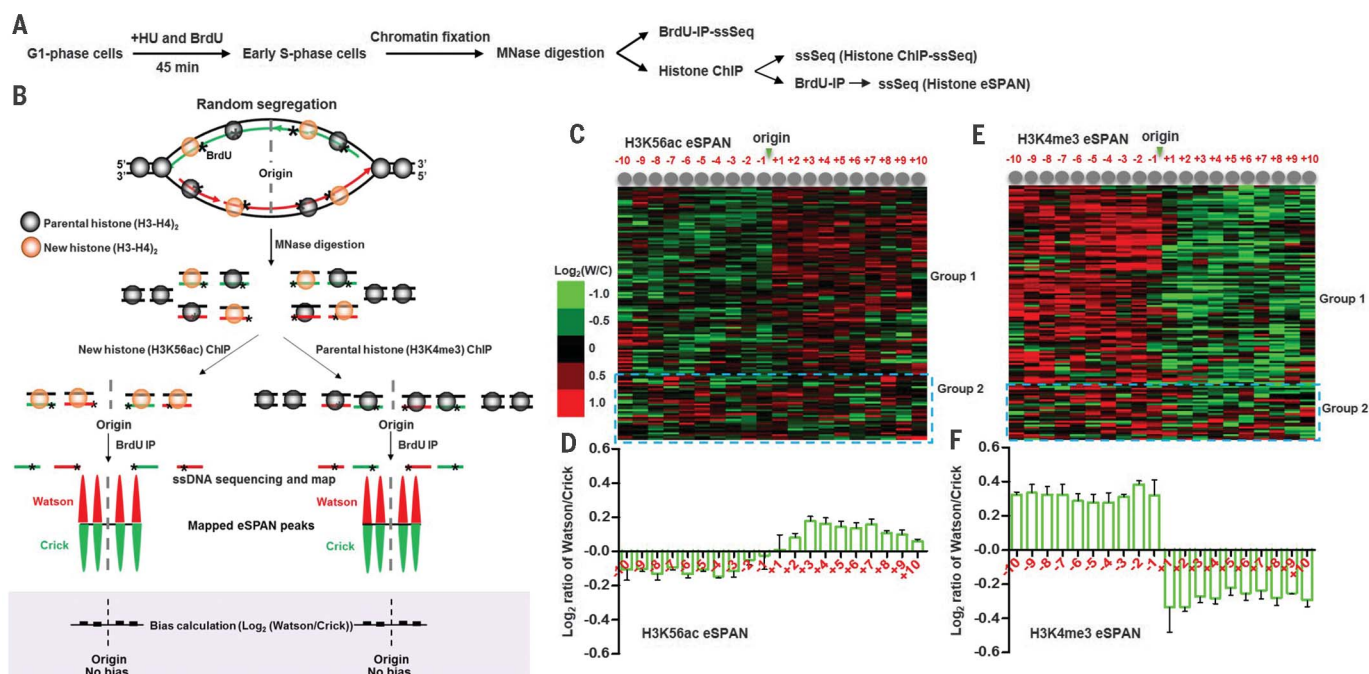
To determine how depletion of *DPB3* and *DPB4* affects the distribution pattern of H3K4me3 and H3K56ac at replicating DNA, we calculated the

<sup>1</sup>Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN 55905, USA. <sup>2</sup>Institute for Cancer Genetics, Department of Pediatrics and Genetics and Development, Columbia University, New York, NY 10032, USA. <sup>3</sup>National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. <sup>4</sup>University of Chinese Academy of Sciences, Beijing 100049, China. <sup>5</sup>Department of Medical Biochemistry and Biophysics, Umeå University, SE 90187 Umeå, Sweden.

\*These authors contributed equally to this work.

†Corresponding author. Email: zz2401@cumc.columbia.edu

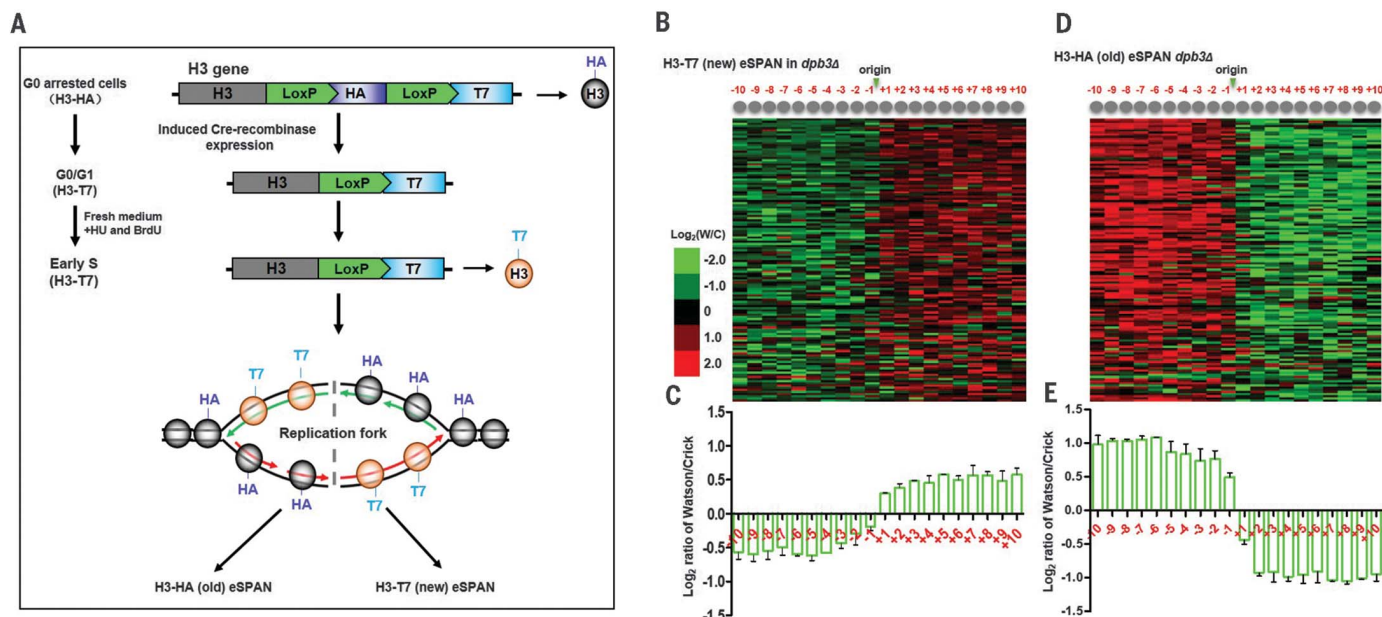




**Fig. 1. Newly synthesized (H3K56ac) and parental histone (H3K4me3) show a slight preference for leading and lagging strands, respectively.**

(A and B) An outline of experimental procedures (A) and a diagram for the hypothetical eSPAN outcome assuming that parental and new (H3-H4)<sub>2</sub> tetramers are equally distributed to leading and lagging strands (B). Black lines, parental DNA; red and green lines, newly synthesized Watson and Crick strands, respectively; ssDNA, single-stranded DNA; asterisks, BrdU. (C) Heatmap showing the bias ratio of H3K56ac eSPAN peaks at each of the 20 individual nucleosomes surrounding 134 early DNA replication origins. The individual nucleosome positions are numbered from -10 to +10 and

are represented by circles. Each row represents the average log<sub>2</sub> ratio Watson/Crick (W/C) of H3K56ac eSPAN sequence reads at one origin and is clustered on the basis of hierarchical clustering analysis. The blue dashed outline indicates separation of two groups of origins, with group 1 origins showing a more consistent bias pattern. (D) The average bias ratio of H3K56ac eSPAN peaks at each of the 20 individual nucleosomes of the 134 early replication origins (*n* = 3 independent repeats). (E and F) H3K4me3 eSPAN peaks at newly replicated chromatin exhibit a slight lagging-strand bias. Error bars in (D) and (F) indicate standard error of three repeats.



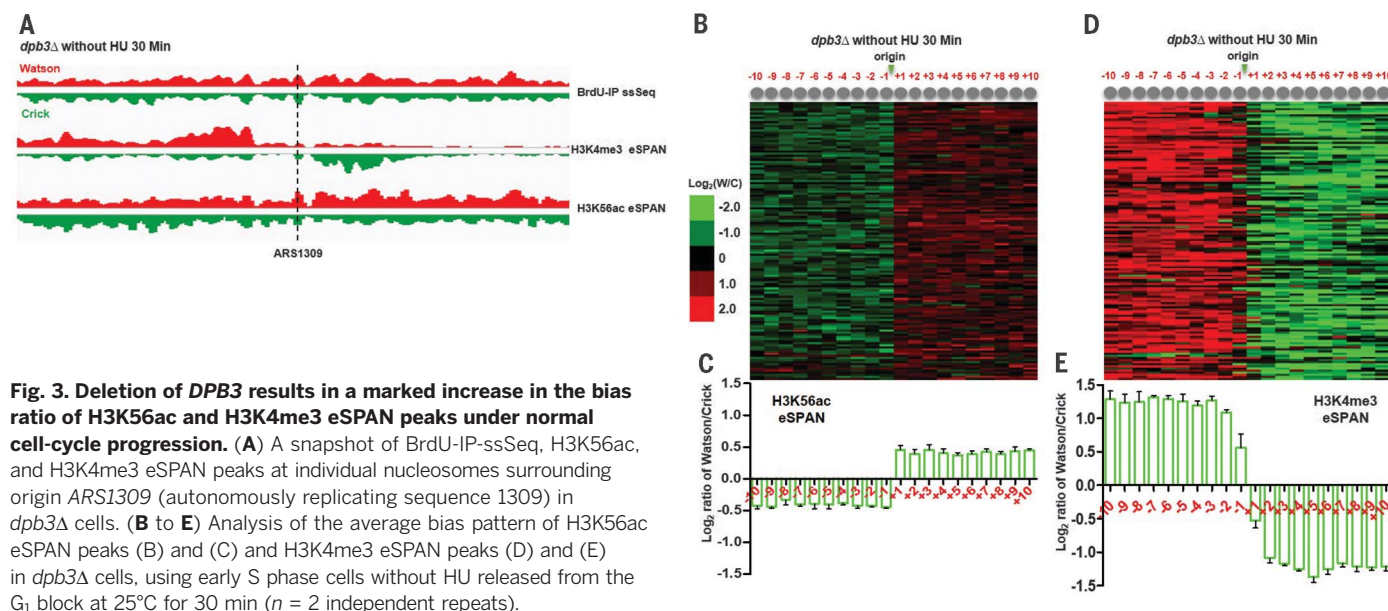
**Fig. 2. Analyzing segregation of newly synthesized and parental histone H3 in *dpb3Δ* cells with the RITE system.** (A) Schematic outline for marking the parental and newly synthesized H3 with the HA epitope (H3-HA) and T7 (H3-T7), respectively. The HA epitope was flanked by two LoxP sites, and upon induction of Cre recombinase, the HA epitope is replaced by the T7 tag. (B) Heatmap showing

the H3-T7 eSPAN bias pattern in *dpb3Δ* cells at each of the 134 individual origins ranked, from top to bottom, on the basis of the replication efficiency. (C) The average bias pattern of H3-T7 eSPAN peaks at 134 early replication origins. (D and E) H3-HA eSPAN peaks in *dpb3Δ* cells show a strong lagging-strand bias. Error bars in (C) and (E) indicate standard error of two repeats.

relative amount of these histones at leading and lagging strands in *dpb3Δ* or *dpb4Δ* cells normalized to that in wild-type cells (Fig. 4A). Compared with that in wild-type cells, the amount of H3K4me3 at leading strands in *dpb3Δ* and *dpb4Δ* cells was reduced significantly (Fig. 4B and

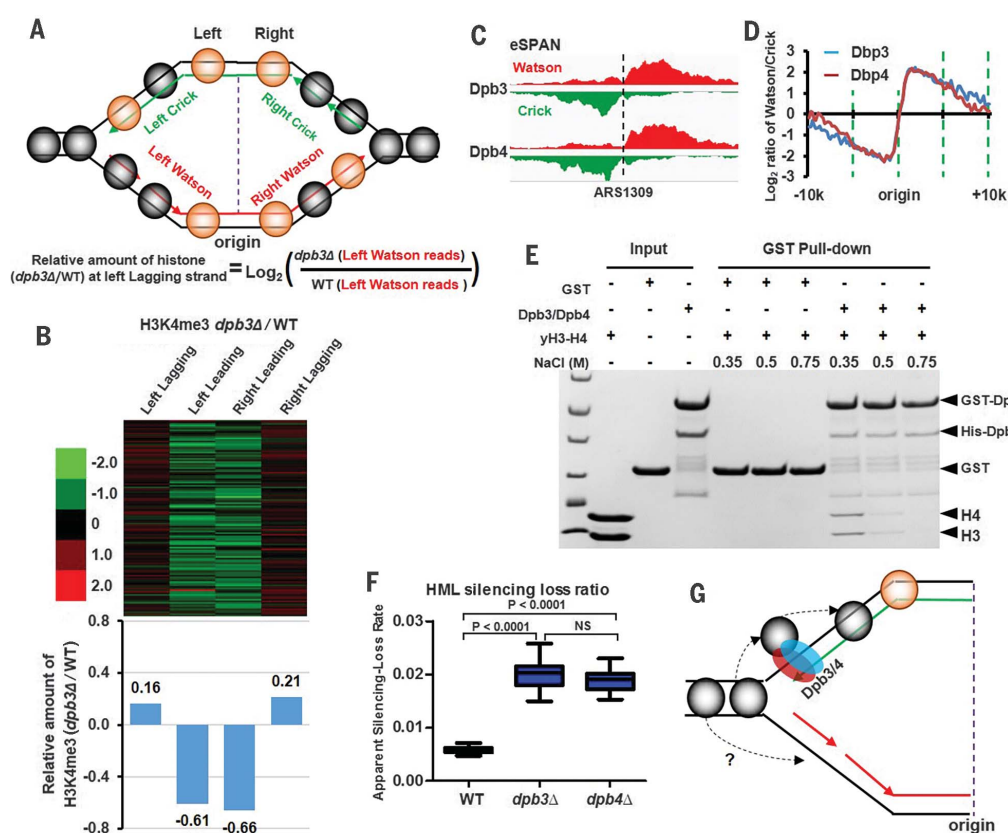
fig. S9A). By contrast, the amount of H3K4me3 at lagging strands in these two mutants increased slightly compared with that in wild type (Fig. 4B and fig. S9A). Moreover, deletion of *DPB3* and *DPB4* had minor effects on H3K56ac deposition onto replicating DNA strands compared

with wild type (fig. S9, B and C). These results indicate that the transfer of H3K4me3 to leading strands is impaired the most in *dpb3Δ* and *dpb4Δ* cells (fig. S9, D and E). Thus, Dpb3 and Dpb4 facilitate the transfer of parental (H3-H4)<sub>2</sub> onto leading strands. Supporting this idea, Dpb3 and



#### Fig. 4. Deletion of *DPB3* and *DPB4* compromises the parental (H3-H4)<sub>2</sub> transfer to leading strands.

(A) A diagram for calculating the relative amount of histone (H3-H4)<sub>2</sub> at leading and lagging strands in *dpb3Δ* cells normalized to that in wild-type (WT) cells using the eSPAN datasets. Gray circles, parental (H3-H4)<sub>2</sub>; orange circles, new (H3-H4)<sub>2</sub>; black lines, parental DNA; red and green lines, newly synthesized Watson and Crick strands, respectively. (B) Top: The relative levels, coded by color, of H3K4me3 in *dpb3Δ* cells compared with those in wild-type cells at each of the 134 individual origins. Bottom: The log<sub>2</sub> ratio of the average amount of H3K4me3 at lagging and leading strands of 134 replication origins in *dpb3Δ* cells divided by that in wild-type cells. (C) Dpb3 and Dpb4 eSPAN peaks at the ARS1309 origin. (D) The average bias of Dpb3 and Dpb4 eSPAN peaks at 134 early replication origins. The x axis is the distance, in bases, from the origin. (E) GST-Dpb3-Dpb4 pull-down assays show that Dpb3-Dpb4 dimers bind recombinant (H3-H4)<sub>2</sub> in vitro. GST, glutathione S-transferase; His, histidine tag; yH3-H4, yeast histone H3-H4. (F) *dpb3Δ* and *dpb4Δ* cells exhibit an increased loss of silencing at the *HML* locus. NS, not significant. (G) A role for Dpb3-Dpb4 in the transfer of parental histone (H3-H4)<sub>2</sub> onto leading strands. The “?” indicates an unknown protein complex mediating the transfer of parental H3-H4 to lagging strands.



Dpb4 were enriched at leading strands (Fig. 4, C and D), similar to the catalytic subunit of Pol  $\epsilon$ , which is also enriched at leading strands (10). Moreover, in vitro, Dpb3-Dpb4 dimers, which are structurally similar to H2A-H2B (17) and copurify with H4 in cells (18), interacted with (H3-H4)<sub>2</sub> tetramers (Fig. 4E). These results support the model that once nucleosomes ahead of DNA replication forks are disassembled, Dpb3-Dpb4 molecules serve as “receptors” or “chaperones” for parental (H3-H4)<sub>2</sub> tetramers and facilitate their assembly onto leading strands (Fig. 4G).

Cells lacking Dpb3 or Dpb4 are defective in heterochromatin silencing (17, 19). Using an assay that monitors transient losses of heterochromatin silencing at the mating type loci (20), we observed a significant increase in the switching of the heterochromatin state at the *HML* locus in *dpb3Δ* and *dpb4Δ* cells (Fig. 4F), suggesting that a defect in parental (H3-H4)<sub>2</sub> transfer in *dpb3Δ* and *dpb4Δ* compromises epigenetic inheritance.

Parental (H3-H4)<sub>2</sub> tetramers are distributed onto both leading and lagging strands, with a slight preference for lagging strands. Dpb3-Dpb4 facilitates their assembly onto leading strands, thereby providing a mechanism for preventing a large asymmetric distribution of parental (H3-H4)<sub>2</sub> at replicating DNA strands. Our results suggest that different proteins likely promote nucleosome assembly of parental (H3-H4)<sub>2</sub> tetramers onto leading versus lagging strands. It is known that Mcm2, a subunit of the MCM helicase, contains a histone-binding motif that binds (H3-H4)<sub>2</sub>. Moreover, it was proposed that Mcm2 is involved in nucleosome assembly of parental (H3-H4)<sub>2</sub> tet-

ramers (21). In addition, DNA polymerase  $\alpha$  and DNA polymerase clamp (proliferating cell nuclear antigen, or PCNA) also have chromatin-related functions (22, 23). It would be interesting to determine whether any of these DNA replication proteins have a direct role in reassembly of parental histones. The utilization of different proteins for the parental histone transfer also hints at a strategy used by eukaryotic cells to alter histone distribution patterns and possibly chromatin states in response to developmental stimuli, thereby endowing eukaryotic cells with the ability both to maintain epigenetic states and to generate epigenetic plasticity such as the generation of neuronal bilateral asymmetry in *Caenorhabditis elegans* during development (24).

## REFERENCES AND NOTES

1. R. T. Coleman, G. Struhl, *Science* **356**, eaai8236 (2017).
2. K. Ragunathan, G. Jih, D. Moazed, *Science* **348**, 1258699 (2015).
3. F. Laprell, K. Finkl, J. Müller, *Science* **356**, 85–88 (2017).
4. X. Wang, D. Moazed, *Science* **356**, 88–91 (2017).
5. E. I. Campos, J. M. Stafford, D. Reinberg, *Trends Cell Biol.* **24**, 664–674 (2014).
6. M. Ransom, B. K. Dennehey, J. K. Tyler, *Cell* **140**, 183–195 (2010).
7. M. Xu et al., *Science* **328**, 94–98 (2010).
8. A. Groth, W. Rocha, A. Verreault, G. Almouzni, *Cell* **128**, 721–733 (2007).
9. A. Serra-Cardona, Z. Zhang, *Trends Biochem. Sci.* **43**, 136–148 (2018).
10. C. Yu et al., *Mol. Cell* **56**, 551–563 (2014).
11. S. Liu et al., *Science* **355**, 415–420 (2017).
12. H. Masumoto, D. Hawke, R. Kobayashi, A. Verreault, *Nature* **436**, 294–298 (2005).
13. M. L. Eaton, K. Galani, S. Kang, S. P. Bell, D. M. MacAlpine, *Genes Dev.* **24**, 748–753 (2010).
14. H. Zhou, B. J. Madden, D. C. Muddiman, Z. Zhang, *Biochemistry* **45**, 2852–2861 (2006).
15. P. M. J. Burgers, T. A. Kunkel, *Annu. Rev. Biochem.* **86**, 417–438 (2017).
16. K. F. Verzijlbergen et al., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 64–68 (2010).
17. H. He et al., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12524–12529 (2017).
18. A. J. Tackett et al., *J. Cell Biol.* **169**, 35–47 (2005).
19. S. Makovets, I. Herskowitz, E. H. Blackburn, *Mol. Cell. Biol.* **24**, 4019–4031 (2004).
20. R. Janke, G. A. King, M. Kupiec, J. Rine, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2030–E2039 (2018).
21. M. Foltman et al., *Cell Reports* **3**, 892–904 (2013).
22. Z. Zhang, K. Shibahara, B. Stillman, *Nature* **408**, 221–225 (2000).
23. Nakayama Ji, R. C. Allshire, A. J. Klar, S. I. Grewal, *EMBO J.* **20**, 2857–2866 (2001).
24. S. Nakano, B. Stillman, H. R. Horvitz, *Cell* **147**, 1525–1536 (2011).

## ACKNOWLEDGMENTS

We thank S. Jia, G. Struhl, and B. Stillman for comments on this manuscript and F. van Leeuwen, H. D. Ulrich, and J. Rine for yeast strains. **Funding:** This study was supported by grant NIH R35GM118015 (Z.Z.); grants NSFC 31521002, MOST 2017YFA0103304, and CAS XDB08010100 (R.-M.X.); and Cancerfonden and the Swedish Research Council (E.J. and A.C.). **Author contributions:** C.Y., H.G., and Z.Z. conceived the project. C.Y., A.S.-C., L.Z., S.G., and S.S. performed experiments indicated online. H.G. performed the data analysis. Z.Z., R.-M.X., E.J., and A.C. supervised the corresponding study. C.Y., H.G., and Z.Z. wrote the manuscript with comments from all authors. **Competing interests:** The authors declare no conflicts of interest. **Data and materials availability:** The deep sequencing datasets have been deposited in the Gene Expression Omnibus (GEO) database (GSE112522).

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6409/1386/suppl/DC1  
Materials and Methods  
Figs. S1 to S9  
Table S1  
References (25–39)

18 April 2018; accepted 30 July 2018  
Published online 16 August 2018  
10.1126/science.aat8849



## MOLECULAR BIOLOGY

# MCM2 promotes symmetric inheritance of modified histones during DNA replication

Nataliya Petryk<sup>1,2\*</sup>, Maria Dalby<sup>3\*</sup>, Alice Wenger<sup>1,2</sup>, Caroline B. Stromme<sup>1†</sup>, Anne Strandsby<sup>1‡</sup>, Robin Andersson<sup>3§</sup>, Anja Groth<sup>1,2§</sup>

During genome replication, parental histones are recycled to newly replicated DNA with their posttranslational modifications (PTMs). Whether sister chromatids inherit modified histones evenly remains unknown. We measured histone PTM partition to sister chromatids in embryonic stem cells. We found that parental histones H3-H4 segregate to both daughter DNA strands with a weak leading-strand bias, skewing partition at topologically associating domain (TAD) borders and enhancers proximal to replication initiation zones. Segregation of parental histones to the leading strand increased markedly in cells with histone-binding mutations in MCM2, part of the replicative helicase, exacerbating histone PTM sister chromatid asymmetry. This work reveals how histones are inherited to sister chromatids and identifies a mechanism by which the replication machinery ensures symmetric cell division.

**H**istone posttranslational modifications (PTMs) contribute to the establishment and maintenance of epigenetic chromatin states that regulate transcriptional programs during development (1, 2), but the mechanisms that ensure transmission of histone PTM patterns to daughter cells remain unclear. Chromatin is disrupted upon replication fork passage, and nucleosomes are rapidly reassembled on newly synthesized DNA through recycling of evicted parental histones and de novo deposition of new histones (3, 4). The recycling of modified parental histones is a critical step in histone PTM transmission (5), and early studies suggested that parental histones segregate randomly to both daughter DNA strands (6, 7). However, whether histone PTM inheritance is truly symmetric and how parental histones are segregated to the leading and lagging strands of the replication fork remain open questions. Multiple replication origins are used to replicate large metazoan chromosomes, and replication fork directionality (RFD) and leading- and lagging-strand replication therefore alternate along chromosomes (8, 9). Potential biases in segregation of modified parental histones during replication will thus result in a specific pattern of sister chromatid asymmetry. We investigated

the distribution of parental and new histones on sister chromatids and linked this distribution to RFD to understand histone segregation.

We developed SCAR-seq (sister chromatids after replication by DNA sequencing) to track histone recycling and de novo deposition genome-wide (Fig. 1A and methods). We differentiated old and new histones H4 by dimethylation at lysine 20 (H4K20me2) (fig. S1A), present exclusively on >80% of old H4 in nascent chromatin (5, 10), and acetylation at lysine 5 (H4K5ac), present on >95% of new H4 (3, 11). Mouse embryonic stem cells were labeled with 5-ethynyl-2'-deoxyuridine (EdU), and nascent mononucleosomes carrying H4K20me2 or H4K5ac were purified sequentially by chromatin immunoprecipitation (ChIP) and streptavidin capture of biotinylated EdU-labeled DNA. The new and parental DNA strands were separated (fig. S1B) and sequenced in a strand-specific manner to score genome-wide sister chromatid histone partition (Fig. 1A and fig. S1C).

To determine locally which sister chromatid was replicated preferentially by the leading strand, we measured RFD by Okazaki fragment sequencing (OK-seq) (methods) (8). Replication initiation zones ( $n = 2,844$ ) (fig. S2A) were comparable to those in humans (8) and *Caenorhabditis elegans* (12), ranging in size and efficiency (fig. S2, B and C), and were mostly intergenic (fig. S2D), enriched in enhancer-associated features [H3 acetylation at lysine 27 (H3K27ac), H3 monomethylation at lysine 4 (H3K4me1), p300 occupancy, and deoxyribonuclease I-hypersensitive sites], and flanked by active genes [marked by H3 trimethylation at lysine 36 (H3K36me3) and lysine 4 (H3K4me3)] (fig. S2E). Around initiation zones, the partition of old and new H4 showed a weak reciprocal shift, with H4K20me2 and H4K5ac skewed toward leading- and lagging-strand replication, respectively (Fig. 1, B to D, and fig. S3, A to C). The partition amplitude was considerably

lower than RFD, suggesting that old histones segregate to both strands but not entirely symmetrically. Analysis of the parental DNA strands showed the complementary partition shift (fig. S3, D and E), excluding an effect of EdU on partition measurements. The partition skew was most pronounced around highly efficient initiation zones (fig. S3F), indicating that DNA replication drives the observed sister chromatid asymmetries. Histone partition skew also tracked with RFD at higher genomic scales (fig. S4)—for example, across replication units with early-replicating borders and late-replicating centers, termed U-domains (8, 9, 13). Together, these results demonstrate that parental histones segregate to both arms of the replication fork with a slight preference for the leading strand, whereas de novo deposition has a comparable bias toward the lagging strand.

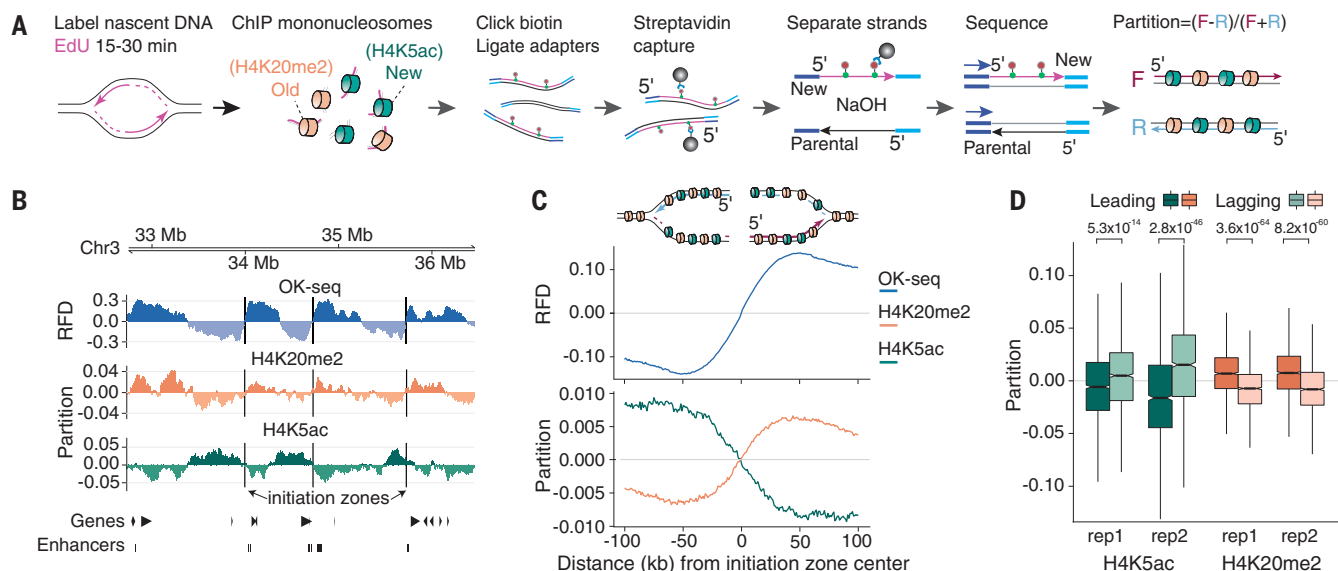
Replication timing is related to chromosome organization in topologically associating domains (TADs) (9, 14, 15). TAD borders (16) are enriched in initiation zones (fig. S5A) (8, 13) and showed a reciprocal histone partition skew (Fig. 2, A and B, and fig. S5B). B-compartment TADs (transcriptionally inactive) displayed stronger RFD and partition shifts than transcriptionally active A-compartment TADs (17) (Fig. 2A), possibly because of increased internal initiation within active TADs (Fisher's exact test, odds ratio 2.2,  $P < 2.2 \times 10^{-16}$ ) or an effect of transcription. To investigate partition asymmetries over genes, we tracked H3K36me3 (fig. S5C) present on parental histones in gene bodies (5, 18, 19). H3K36me3 partition skewed moderately toward leading-strand replication, consistent with H4K20me2 partition (Fig. 2A and fig. S5, C to E), and was stronger over active genes and codirectional with transcription (Fig. 2C) (8). Further, the correlation of chromatin interaction directionality with histone partition was weaker than the correlation between chromatin interaction directionality and RFD (Fig. 2B), suggesting that although histone partitioning is driven by RFD it can be affected by transcription. Active enhancers often coincided with initiation zone centers, and promoters tended to be flanking (20) (fig. S5F), suggesting that enhancer activity affects partitioning in neighboring regions. Consistently, both RFD and histone partition asymmetries were greater around active enhancers (21) and super-enhancers that control cell type-specific genes (22) (Fig. 2D) (Mann-Whitney U test,  $P < 1.1 \times 10^{-16}$ ).

MCM2, part of the replicative helicase, is proposed to recycle parental histone H3-H4 via its N-terminal histone-binding domain (HBD) (11, 23–25). Using genome editing, we mutated two critical residues in the HBD [Tyr<sup>81</sup>→Ala (Y81A) and Tyr<sup>90</sup>→Ala (Y90A), yielding MCM2-2A] (24, 25) (fig. S6A) that disrupt histone binding (fig. S6B) (25) without affecting cell cycle progression (fig. S6C). Notably, in MCM2-2A mutants, partition of old and new histones was strongly skewed toward leading and lagging strands, respectively, generating partition ratios similar to RFD in amplitude and pattern (Fig. 3, A to C, and figs. S6D and S7). Moreover, partition of new and

<sup>1</sup>Biotech Research and Innovation Centre (BRIC), Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark. <sup>2</sup>Novo Nordisk Foundation Center for Protein Research (CPR), Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark. <sup>3</sup>The Bioinformatics Centre, Department of Biology, Faculty of Science, University of Copenhagen, 2200 Copenhagen, Denmark.

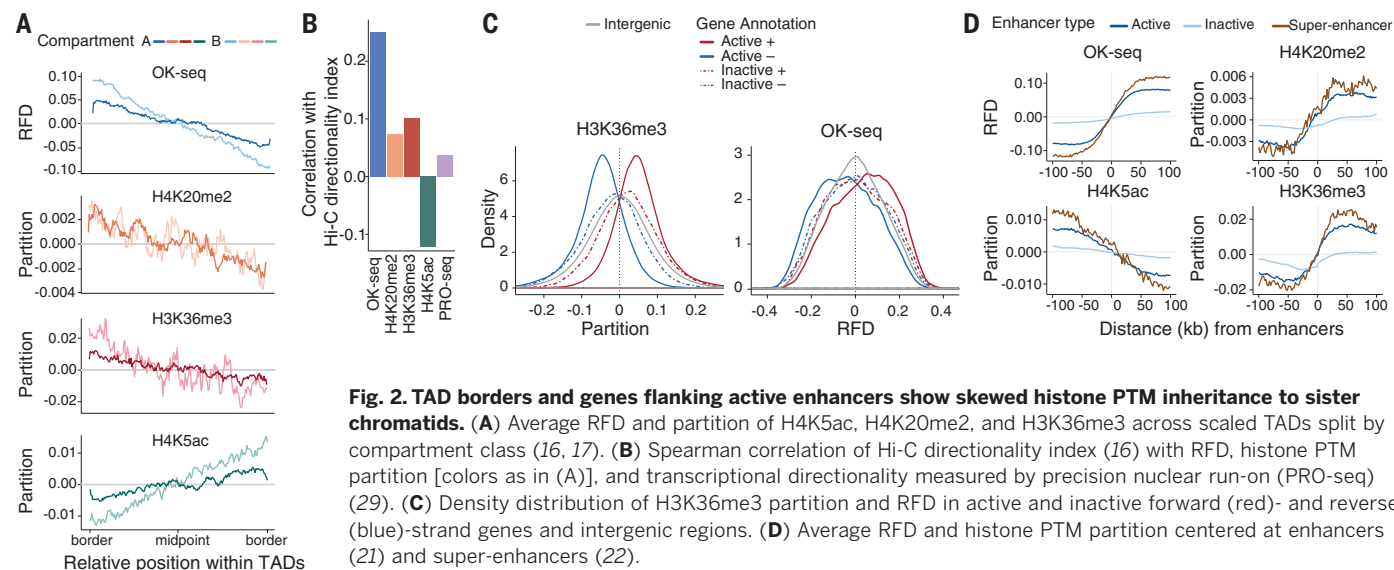
\*These authors contributed equally to this work. †Present address: Novo Nordisk, 2860 Soborg, Denmark. ‡Present address: Virus and Microbiological Special Diagnostics, Statens Serum Institut, 2300 Copenhagen S, Denmark.

§Corresponding author. Email: robin@binf.ku.dk (R.A.); anja.groth@bric.ku.dk (A.G.)



**Fig. 1. Parental histones segregate to both sister chromatids with a weak bias toward the leading strand.** (A) SCAR-seq technique. Partition of old and new histones is calculated as the proportion of forward (F) (red) and reverse (R) (blue) counts in genomic windows [the range is between -1 (100% reverse strand) and 1 (100% forward strand)]. (B) RFD and partition of H4K20me2 and H4K5ac at a genomic region. Initiation zone centers (lines), active gene orientation (arrowheads), and active enhancers

(bars) are shown. Chr3, chromosome 3. RFD is calculated as the proportion between right- and left-moving replication forks [the range is between -1 (100% left moving) and 1 (100% right moving)]. (C) Average RFD (blue) and partition of old (H4K20me2) and new (H4K5ac) histones around initiation zones. (D) Partition at downstream (leading strand) and upstream (lagging strand) edges of initiation zones, with significant partition difference in each replicate (rep) (paired Wilcoxon signed-rank test,  $P < 5.3 \times 10^{-14}$ ).



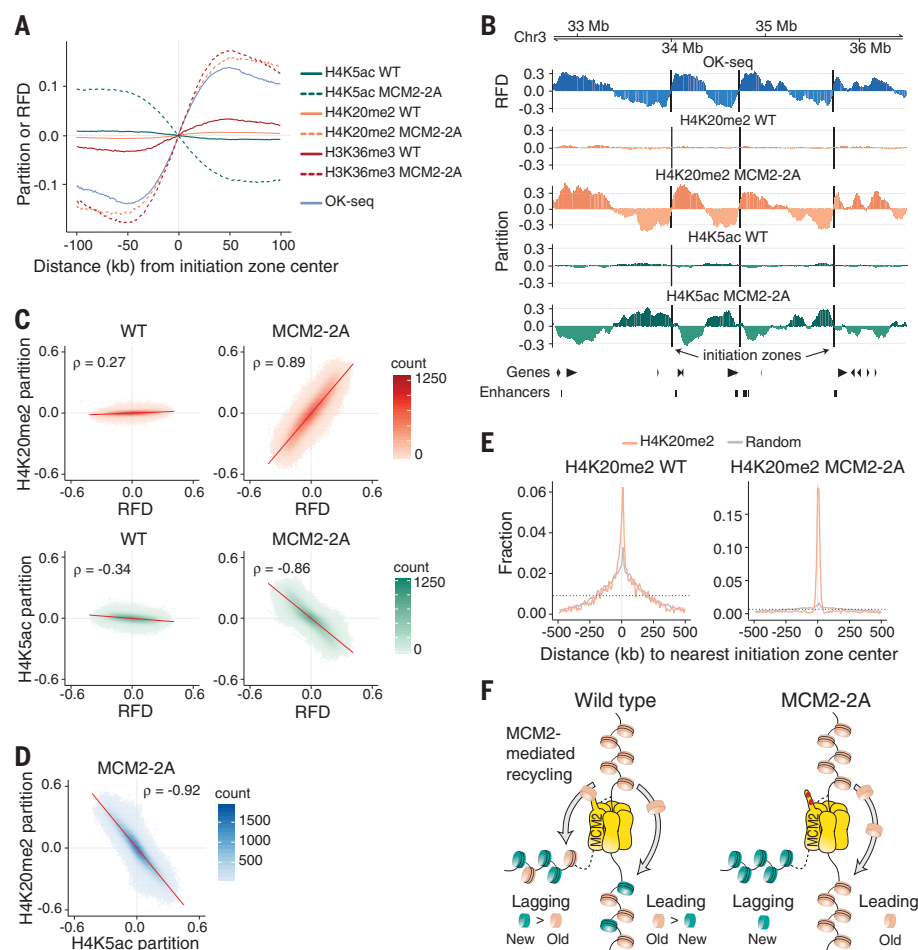
**Fig. 2. TAD borders and genes flanking active enhancers show skewed histone PTM inheritance to sister chromatids.** (A) Average RFD and partition of H4K5ac, H4K20me2, and H3K36me3 across scaled TADs split by compartment class (16, 17). (B) Spearman correlation of Hi-C directionality index (16) with RFD, histone PTM partition [colors as in (A)], and transcriptional directionality measured by precision nuclear run-on (PRO-seq) (29). (C) Density distribution of H3K36me3 partition and RFD in active and inactive forward (red)- and reverse (blue)-strand genes and intergenic regions. (D) Average RFD and histone PTM partition centered at enhancers (21) and super-enhancers (22).

old histones showed strong anticorrelation in MCM2-2A (Fig. 3D and fig. S7D), indicating segregation to opposite strands. H3K36me3 occupancy was not altered in MCM2-2A cells (fig. S8, A and B), indicating that histone partitioning rather than recycling was perturbed. The association between H3K36me3 partition and transcriptional directionality was reduced in MCM2-2A cells (fig. S8, C to E), further indicating increased replication-driven sister chromatid asymmetry in parental and new histones. Sister chromatid asymmetry was also strongly increased at TAD borders, around enhancers (fig. S8, F and G),

and across important developmental loci (e.g., the *Hox* clusters) (fig. S9) in MCM2-2A cells, which thus provides a model to address histone PTM inheritance in development. The high correlation between H4K20me2 partition in MCM2-2A cells and RFD prompted us to map H4K20me2 partition break points. They showed strong colocalization with initiation zones mapped by OK-seq (Fig. 3E and methods), suggesting H4K20me2 SCAR-seq in MCM2-2A as a method to map replication dynamics.

In summary, SCAR-seq revealed that parental histone segregation is almost symmetrical, with

a weak inherent preference for the leading strand (Fig. 3F, left), creating modest sister chromatid asymmetries that might be mitotically transmitted as new histones acquire PTMs with slow kinetics (5). Importantly, MCM2 histone chaperone activity promotes balanced segregation of old histones to leading and lagging strands, thereby ensuring inheritance of histone-based information to both sister chromatids. This is consistent with chaperoning of old histones by MCM2 (11, 23) and cryo-electron microscopy data placing the MCM2 HBD in front of the fork (26, 27). We envisage that MCM2 recycles



**Fig. 3. MCM2 histone binding is required for parental histone recycling to the lagging strand.** (A) Average RFD and partition of old (H4K20me2 and H3K36me3) and new (H4K5ac) histones in the wild type (WT) (solid lines) and MCM2-2A (dashed lines) around initiation zones. (B) RFD and histone PTM partition at a genomic region in the WT and MCM2-2A. Initiation zone centers (lines), active gene orientation (arrowheads), and active enhancers (bars) are shown. (C) Scatterplots of RFD and histone PTM partition in the WT and MCM2-2A. Spearman's rank correlation coefficient is shown in the top left corner of each plot. (D) Scatterplot of H4K20me2 versus H4K5ac partition in MCM2-2A. Spearman's rank correlation coefficient is shown in the top right corner. (E) Fraction of initiation zones with the nearest distance to predicted H4K20me2 partition break points or random H4K20me2 bins in the WT and MCM2-2A. Horizontal dotted lines represent random mean fractions. (F) Model for segregation of parental histones H3-H4 in WT and MCM2-2A cells.

parental histones to the lagging strand (Fig. 3F, right), whereas a separate pathway deposits parental histones on the leading strand. In this vein, it is conceivable that histone segregation can be regulated during development to drive asymmetric cell fates (28).

#### REFERENCES AND NOTES

1. C. D. Allis, T. Jenuwein, *Nat. Rev. Genet.* **17**, 487–500 (2016).
2. R. P. Halley-Stott, J. B. Gurdon, *Brief. Funct. Genomics* **12**, 164–173 (2013).
3. C. Alabert, A. Groth, *Nat. Rev. Mol. Cell Biol.* **13**, 153–167 (2012).
4. E. I. Campos, J. M. Stafford, D. Reinberg, *Trends Cell Biol.* **24**, 664–674 (2014).

5. C. Alabert *et al.*, *Genes Dev.* **29**, 585–590 (2015).
6. V. Pospelov, G. Russev, L. Vassilev, R. Tsanev, *J. Mol. Biol.* **156**, 79–91 (1982).
7. A. T. Annunziato, *J. Biol. Chem.* **280**, 12065–12068 (2005).
8. N. Petryk *et al.*, *Nat. Commun.* **7**, 10208 (2016).
9. O. Hyrien, in *The Initiation of DNA Replication in Eukaryotes*, D. Kaplan, Ed. (Springer, 2016), pp. 65–85.
10. G. Saredi *et al.*, *Nature* **534**, 714–718 (2016).
11. Z. Jasencakova *et al.*, *Mol. Cell* **37**, 736–743 (2010).
12. E. Pourkarimi, J. M. Bellush, I. Whitehouse, *eLife* **5**, e21728 (2016).
13. A. Baker *et al.*, *PLOS Comput. Biol.* **8**, e1002443 (2012).
14. B. D. Pope *et al.*, *Nature* **515**, 402–405 (2014).
15. J. R. Dixon *et al.*, *Nature* **485**, 376–380 (2012).
16. B. Bonev *et al.*, *Cell* **171**, 557–572.e24 (2017).
17. E. Lieberman-Aiden *et al.*, *Science* **326**, 289–293 (2009).
18. A. Loyola, T. Bonaldi, D. Roche, A. Imhof, G. Almouzni, *Mol. Cell* **24**, 309–316 (2006).
19. J. C. Black, C. Van Rechem, J. R. Whetstone, *Mol. Cell* **48**, 491–507 (2012).
20. C. Cayrou *et al.*, *Genome Res.* **25**, 1873–1885 (2015).
21. R. Andersson *et al.*, *Nature* **507**, 455–461 (2014).
22. W. A. Whyte *et al.*, *Cell* **153**, 307–319 (2013).
23. A. Groth *et al.*, *Science* **318**, 1928–1931 (2007).
24. M. Foltman *et al.*, *Cell Rep.* **3**, 892–904 (2013).
25. H. Huang *et al.*, *Nat. Struct. Mol. Biol.* **22**, 618–626 (2015).
26. R. Georgescu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E697–E706 (2017).
27. M. E. Douglas, F. A. Ali, A. Costa, J. F. X. Diffley, *Nature* **555**, 265–268 (2018).
28. V. Tran, C. Lim, J. Xie, X. Chen, *Science* **338**, 679–682 (2012).
29. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).

#### ACKNOWLEDGMENTS

We thank P. Lansdorp for initial discussions on SCAR-seq; K. Stewart-Morgan and C. Hammond for comments on the manuscript; A. Fossom for help with fluorescence-activated cell sorting; and the Groth, Andersson, and Brakebusch laboratories for discussions. **Funding:** Research in the Groth lab was supported by the Independent Research Fund Denmark (4092-00404), the European Research Council (CoG no. 724436), the Novo Nordisk Foundation (NNF140C0012839), the Lundbeck Foundation (R198-2015-269), and the Danish Cancer Society. Research in the Andersson lab was supported by the Independent Research Fund Denmark (6108-000388) and the European Research Council (StG no. 638173). **Author contributions:** N.P. and A.G. conceived of the project and designed experiments. N.P. developed and performed SCAR-seq. M.D. and R.A. developed computational methods. M.D. performed computational analyses with support from N.P. and R.A. C.B.S. and A.S. performed genome editing. A.W. characterized MCM2-2A cells and assisted with SCAR-seq. A.G. and R.A. supervised the project. N.P. and M.D. wrote the manuscript with input from A.G., R.A., and A.W. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data have been deposited to NCBI GEO under accession number GSE117274.

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6409/1389/suppl/DC1  
 Materials and Methods  
 Figs. S1 to S9  
 Tables S1 to S3  
 References (30–42)  
 Data S1

30 April 2018; accepted 6 August 2018  
 Published online 16 August 2018  
 10.1126/science.aau0294

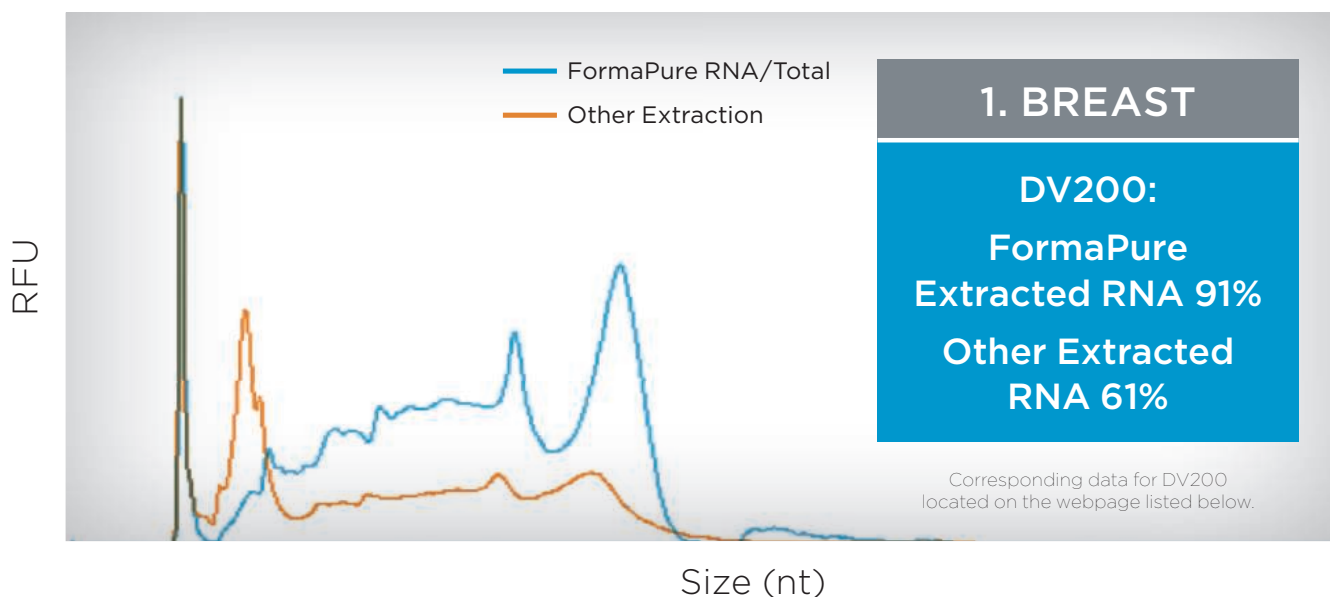


EMPOWERING  
*real discoveries.*

Receive a  
**FREE 10-PREP  
KIT at ASHG  
Booth 1214**  
Oct. 16-20, 2018  
San Diego, CA

# FFPE EXTRACTION IS A CHALLENGE. **FORMAPURE AND CHILL.**

FormaPure Total | FormaPure DNA | FormaPure RNA



Not attending ASHG2018? Request your  
free FormaPure Total 10-prep kit online now!

[beckman.com/EvaluateFormaPure](http://beckman.com/EvaluateFormaPure)

For research use only, not intended or validated for use in the diagnosis of disease or other conditions. © 2018 Beckman Coulter, Inc. All rights reserved. Beckman Coulter, the stylized logo, and the Beckman Coulter product and service marks mentioned herein are trademarks or registered trademarks of Beckman Coulter, Inc. in the United States and other countries.

 **BECKMAN  
COULTER**  
*Life Sciences*



**CALL FOR**

# Genome Sciences Papers

PNAS is committed to bringing you high-impact, cutting-edge research in genome sciences. We encourage you to submit your noteworthy genome sciences papers, particularly those that will create discussion among our broad readership. The increasingly interdisciplinary nature of scientific research underscores the value of publishing in PNAS and is reflected in our high citation rates.

We invite you to submit your outstanding, original research articles in genome sciences at [www.pnascentral.org](http://www.pnascentral.org).

**PNAS**

[www.pnas.org](http://www.pnas.org)

## **GENOME SCIENCES EDITORIAL BOARD MEMBERS**

David Baker  
Stephen M. Beverley  
John Carlson  
C. Thomas Caskey  
Joseph R. Ecker  
Douglas J. Futuyma

Fred H. Gage  
Joseph L. Goldstein  
Daniel L. Hartl  
Rakesh K. Jain  
Mary-Claire King  
Douglas E. Koshland

Richard E. Lenski  
Jeremy Nathans  
Jasper Rine  
Gene E. Robinson  
Allan C. Spradling  
Stephen T. Warren



# Flexible Precision.

## NEBNext Direct<sup>®</sup> Custom Ready Panels for NGS target enrichment

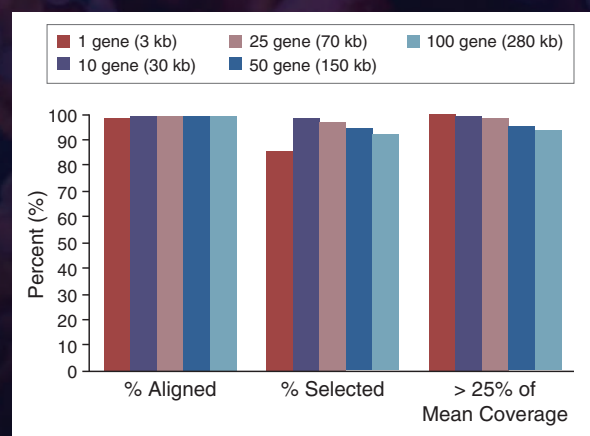
Employing the unique NEBNext Direct hybridization-based enrichment method, NEBNext Direct Custom Ready Panels allow rapid customization of targeted gene panels for Illumina<sup>®</sup> sequencing. Select from a list of genes for which baits have been carefully designed and optimized to give complete coverage of the full coding regions. High quality panels can be designed by you and rapidly delivered, from any combination of genes. NEBNext Direct Custom Ready Panels provide the content you want with the performance you need.

- Choose from a single gene to hundreds of genes
- Experience unmatched specificity and coverage uniformity
- Eliminate synthesis and optimization steps for faster turnaround
- Improve sensitivity with our Unique Molecule Index (UMI)
- Generate results in one day with our automation-friendly workflow

For more information visit

**NEBNextDirect.com**

NEBNext Direct Custom Ready Panels demonstrate optimum performance across a wide range of panel sizes



Key target enrichment metrics demonstrate consistent performance across a range of panel sizes. 100 ng of DNA was tested against panels of 1, 10, 25, 50 and 100 genes, and sequenced using Illumina<sup>®</sup> paired-end 150 bp sequencing. Larger panels included all genes present in smaller panels.



# INTRODUCING 2-MINUTE PCR



Since PCR was introduced, DNA replication has incrementally gotten easier and faster. Now, a revolution has happened. NEXTGENPCR heats and cools instantly — “ramp rates” no longer exist. Using one thermocycler, you can run more than ten plates per hour. The math adds up quickly — 8 hours of 10 plates in a 384-well format means more than 30,000 data points in a single day.



## Ultrafast

100 base pairs and 30 cycles in 2 minutes



## Versatile

Open chemistry



## Flexible

96-well or 384-well format without changing blocks



## Precise

Well uniformity of less than 0.1°C



NG  
PCR

GO FASTER  
[www.canon-biomedical.com/pcr](http://www.canon-biomedical.com/pcr)

[contactus@canon-biomedical.com](mailto:contactus@canon-biomedical.com) | [www.canon-biomedical.com](http://www.canon-biomedical.com) | 844-CANONBIO

For research use only. Not for use in diagnostic procedures.

The NEXTGENPCR products are manufactured by Molecular Biology Systems, B.V.

All referenced product names, and other marks, are trademarks of their respective owners. © 2018 Canon BioMedical, Inc. All rights reserved.

**Canon**  
CANON BIOMEDICAL



# Preemptive pharmacogenomics

*Working toward bringing timely, individually tailored, precision health care to patients*

**W**hen it comes to health care, the right prescriptions not only improve patients' lives—they sometimes save them. So, when a patient requires a medication, a clinician prefers to prescribe it immediately. But that's not always possible. About 200 of the drugs approved by the U.S. Food and Drug Administration (FDA) note pharmacogenomic (PGx) information—interactions between a patient's genes and the medicine they need—that should be considered before prescribing that medication. An example is the cancer drug tamoxifen, the efficacy of which depends on an individual's liver enzyme activity as determined by their particular genetic makeup. Using preemptive pharmacogenomics (PPGx), clinicians can obtain a patient's genetic information—on all potential markers, or only those that are informative when prescribing medications—in preparation for future pharmaceutical needs. More important, early genetic testing can support potentially life-threatening decisions about the best drug treatment for a patient. The Clinical Pharmacogenetics Implementation Consortium, in providing guidelines to clinicians on using genetic test results in prescribing, assumes that PPGx implementation will grow. Collecting and analyzing that genetic information requires technology designed for the job, and Thermo Fisher Scientific (Waltham, Massachusetts) and some of its partners are looking forward to fulfilling those needs.

## Better and safer

"Genetics can cause a person to respond to a drug in a way that was not intended," says Scott Megill, president and CEO at Coriell Life Sciences, in Philadelphia, Pennsylvania—a company that partners with Thermo Fisher to provide reporting and interpretation of molecular diagnostics tests. Depending on a patient's genetic makeup, a drug could provide no therapeutic effect at all or even be unsafe.


Ulrich Broeckel, founder and CEO of RPRD Diagnostics, in Wauwatosa, Wisconsin—another Thermo Fisher partner—explains that PGx testing can be done essentially two ways: when requested, on an as-needed basis; or preemptively, so that if a need arises, the patient's genetic information is already on file. His company focuses on making PPGx testing more widely available so that physicians have information readily at hand to make the best decision for each patient.

PPGx could benefit many—maybe most—patients. About 50 human genes show a high degree of variation that is correlated with responses to medications, but just knowing that a gene varies does not provide meaningful information. "You can't say that someone who has a variant will always have a problem with a medicine," says Philip Empey, associate director for pharmacogenomics at the Institute of Precision Medicine, a collaboration between the University of Pittsburgh and the University of Pittsburgh Medical Center. And it's more complicated than simply identifying whether a gene varies or not. "With some targets, the number of the genetic variant—the number of copies of that gene—is important," adds Empey. With the cholesterol drug simvastatin, for example, having more copies of a specific allele increases the odds of an adverse event (1).

As in any area of medicine, more data is better. Making more PPGx data available will improve its overall value to health care. For PPGx purposes, genomic data collected from the individual is stored as an electronic health record. When enough data is collected, it can be utilized to improve health outcomes. For this reason, the University of Pittsburgh plans to do PPGx testing on more than 100,000 patients and track the results. "Only with a large study of preemptive genotyping can we determine its clinical and economic value," says Empey, who is leading this study.

## Easier data acquisition

PPGx data collection and analysis starts with a sample, such as blood, buccal swab, or saliva. Then, the genes from that sample must be analyzed. "When focusing on analyzing a targeted set of pharmacogenomic markers, where a rapid turnaround is important, qPCR—real-time [quantitative] polymerase chain reaction—is the better choice," says Peter Norster, associate director, global market development, applied healthcare solutions at Thermo Fisher. For example, Megill says his colleagues "build specialized panels, such as for cardiac medications, mood disorders, and even a presurgery test" based on qPCR. This technology also makes it easy to add a gene or a section of a gene to a panel, he notes. Thermo Fisher provides a qPCR research solution for PPGx in their OpenArray platform, although other companies are also in that market space.



“Only with a large study of preemptive genotyping can we determine its clinical and economic value.”

### Addressing hurdles to PPGx

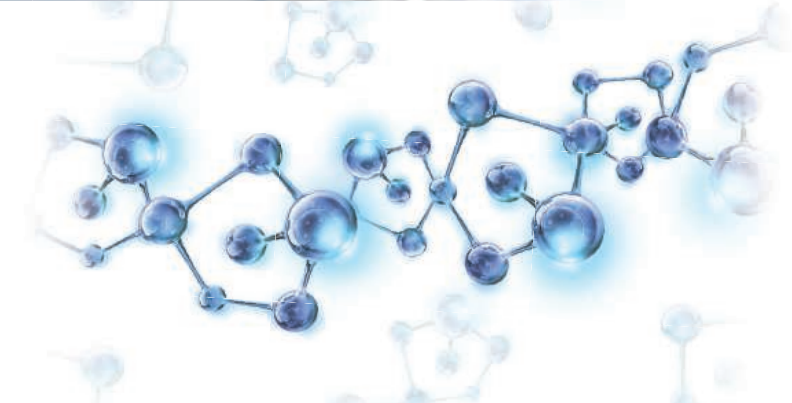
Given the value of PPGx, it seems likely that it would be frequently applied if more widely available. When asked how often PPGx is used, however, Shannon Manzi, director of the Clinical Pharmacogenomics Service at Boston Children's Hospital, says, “not often enough.” In fact, preemptive testing is mostly used in very limited circumstances. “Currently, our experience is that outside of a research study or a well-defined standard of care, the vast majority of testing is done after a patient either experiences adverse effects or has a history of nonresponse [to a drug],” says Manzi. So, the first hurdle is that the standard of care isn't well defined.

The second challenge is that clinicians can't always access the information as soon as it's needed. “If the prescriber cannot access the data at the time of prescribing—or in less than 30 minutes—it is not useful,” Manzi explains. “This also applies to the pharmacist at the time of dispensing.” Manzi sees two possible solutions: “Either universal preemptive testing, such as newborn screening—or something real-time—is needed.” Real-time or near-real-time PGx testing is unfortunately not yet possible.

Furthermore, to get the full benefits of PPGx, says Norster, “the information must go into a person's electronic health record, and that data must be accessible and follow you around,” and be integrated into clinician workflows.

The last challenge is that the entire health care system—including clinicians, patients, and payers—must see the value of PPGx. Experts in the field already see its benefits and envision more in the future. “The key is to raise awareness about the value of preemptive pharmacogenomics,” says Broeckel. In addition, the methods of testing must be easy to use, widely available, and covered by payers. Once all of that is accomplished, every patient's health record could include the crucial genetic information that would allow faster access to many medications—some of which could be needed at a moment's notice.

1. Search Collaborative Group, *N. Engl. J. Med.* **359**, 789–799 (2008).



However, qPCR is not the best method for analyzing large numbers of genes. “When covering all markers—ones that are relevant today and ones that might be useful in the future—a microarray method works better,” says Norster. Working together, RPRD Diagnostics and Thermo Fisher developed the PharmacoScan Solution, which includes the microarray plates and reagents needed for PPGx genotyping. “It covers all of the genes that should be tested,” says Broeckel. PharmacoScan is fully verified and includes highly predictive markers in critical genes such as *CYP2D6*, *CYP2C9*, and *CYP4F2* as well as copy number variation calling. It allows clinical researchers to gain valuable insight into an individual's ability to process drugs based upon high, moderate, low, or preliminary scientific evidence.

### Usable reporting

Collecting the genomic data is only the first step—interpreting the genetic data in a relevant way is also crucial. “You're looking at targets and placing patients in different categories, such as their ability to metabolize a drug,” says Empey. Most physicians were never taught to interpret a patient's genetics to predict a drug response. To get the most from existing medicines, genomic data must be combined with medical information, and that need is likely to increase with precision medicine, which aims to find drugs for specific gene targets.

This information can save lives. Megill relates the example of the blood-thinner Plavix (clopidogrel), which is often prescribed after a heart attack or stroke, and is even incorporated into some stents. Nonetheless, 27% of Caucasians and about 50% of Asians cannot metabolize this drug, which makes it worthless to them—all based on variation in one gene that encodes cytochrome P450, a liver enzyme involved in breaking down molecules in the body. With global sales of USD 1.7 billion in 2017, that would be about USD 450 million wasted without PGx testing—and that's assuming it would fail for only 27% of the general population.

For more information:

[www.thermofisher.com/pgxwhitepaper](http://www.thermofisher.com/pgxwhitepaper)

Sponsored by

For research use only.  
Not for use in diagnostic  
procedures.

**ThermoFisher**  
SCIENTIFIC



# MICHELSON PRIZES

---

## EXPANDED CALL FOR ENTRIES

The Michelson Prizes are scientific awards that support young investigators from across the globe applying disruptive research concepts to significantly advance the development of vaccines and immunotherapies for major global diseases.

2019 Prizes will be awarded in three focus areas, all related to vaccine discovery:

- *Human Immunology*
- *Computational Biology and Protein Engineering*
- *Neglected Parasitic Diseases*

Proposals will be reviewed by a distinguished committee of internationally-recognized scientists.

Researchers under the age of 35 are invited to apply.



## APPLY TODAY

[www.humanvaccinesproject.org/  
MichelsonPrizes](http://www.humanvaccinesproject.org/MichelsonPrizes)

**DEADLINE OCTOBER 29, 2018**

---

**Prize Money of \$150,000**

**Opportunity to present  
at the annual conference on  
'The Future of Vaccine Discovery'**



# The Institute for Genomic Medicine: Creating the Future of Cancer Treatment

“We’re working hard to blur the line between basic research and clinical genomics,” says Richard Wilson, PhD, executive director, Institute for Genomic Medicine at Nationwide Children’s Hospital. The institute’s clinical and genomic services laboratories, computational genomics, technology development group and translational research programs fall under the same umbrella, collaborating to focus on more than just DNA.

“Our Cancer Protocol is aimed at moving genomics into direct patient care. We designed it to better understand the genomics of a patient’s tumor and to make some tests available to patients prior to clinical availability,” says Julie Gastier-Foster, PhD, senior director, IGM Clinical Laboratory. “It’s not a pure research or a pure clinical protocol — it’s a blend.”

The Cancer Protocol reviews patient nominations from hospital oncologists for genomic profiling. In most cases, the nominees have tumors that did not respond to prior rounds of therapy or have recurred. Selected patients are invited to have a full genomic analysis performed.

“We study for tumor and normal DNA and compare the sequences of all known genes. We look for cancer susceptibility genes by testing the normal cells’ DNA, and we examine both the cancer DNA and RNA for clues to what is driving their cancer and might help guide therapy,” says Elaine Mardis, PhD, co-executive director, IGM and president-elect of the American Association for Cancer Research. “That level of comprehensive study is quite uncommon.”

**Learn more about the Institute for Genomic Medicine at [NationwideChildrens.org/IGM](http://NationwideChildrens.org/IGM)**



**NATIONWIDE CHILDREN'S**  
*When your child needs a hospital, everything matters.<sup>SM</sup>*

# NEXT LEVEL CleanTag® Library Preparation for NEXT-GENERATION SEQUENCING

## Increase Single Cell Sensitivity in NGS Libraries with NEW TriLink CleanTag® Technology

Achieve exceptional assay sensitivity,  
resulting in unmatched data quality

Small RNA sequencing from single cells is a powerful analytical tool. However, library prep is dominated by adapter dimer formation. This limits the amount of quality data retrieved from each sample. TriLink has developed a unique solution employing CleanTag® modified adapters that significantly reduce adapter dimer formation. This technology allows for full automation of the library preparation process by using magnetic bead-based methods instead of labor intensive PAGE purification.

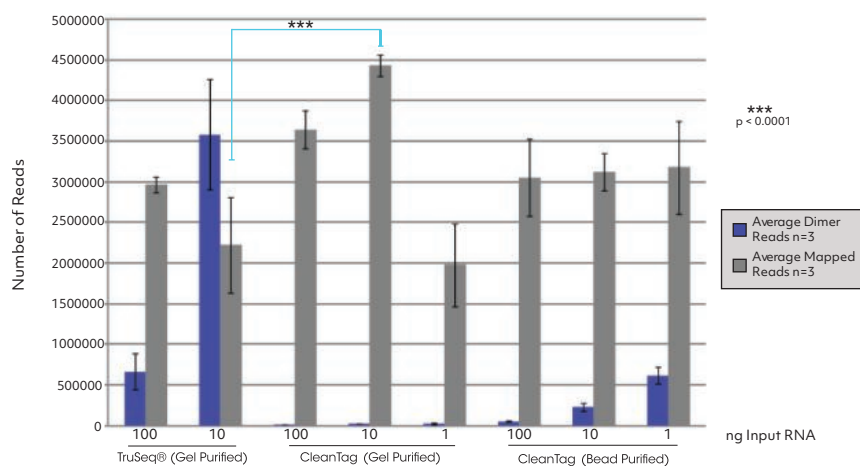
- ✓ Increase mappable reads for inputs as low as 1 ng
- ✓ Efficiently tag samples with low miRNA abundance
- ✓ Reduce sample loss & enable higher NGS sensitivity
- ✓ Decrease library prep labor & overall cost
- ✓ Compatible with Illumina® & Thermo Fisher platforms

### CleanTag® Technology

Blocks adapter-adapter  
ligation with our modified  
adapters

Improves the ratio of adapter  
dimer to tagged library

Generates higher quality  
sequencing data



TruSeq® and CleanTag® miRNA sequencing from gel or bead purified libraries

Illumina® and TruSeq® are properties of Illumina, Inc.



# WILL YOUR RESEARCH LEAD TO BETTER LIVES FOR PATIENTS?



Gopinath Sutendra and Evangelos D. Michelakis, "Pulmonary Arterial Hypertension: Challenges in Translational Research and a Vision for Change", *Sci. Transl. Med.* 5, 208sr5 (2013) Credit: Science Source

**ScienceTranslational Medicine** | AAAS  
INTEGRATING SCIENCE, ENGINEERING, AND MEDICINE

Find out more about the scope of the journal and submit your research today. [ScienceTranslationalMedicine.org](http://ScienceTranslationalMedicine.org)

## NEW!

### Low-Noise Ultra-Fast Digital Patch Clamp Amplifier System

- High bandwidth acquisition of fastest signals
- Single-channel and whole-cell patch clamp recordings
- Digital compensation circuitry for precision and signal fidelity
- Quick and easy setup
- Bundled SutterPatch® software built on Igor Pro platform

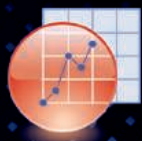


## dPatch®

The next generation Digital Patch Clamp Amplifier System. Combining high-speed, high-resolution digital processing, precision A/D circuitry, integrated data acquisition and bundled SutterPatch® software, the dPatch system provides capabilities previously out of reach for the electrophysiologist. Available in either a single- or double-headstage configuration, the dPatch meets the requirements of today's research and anticipates the demands of tomorrow's.

**SUTTER INSTRUMENT®**

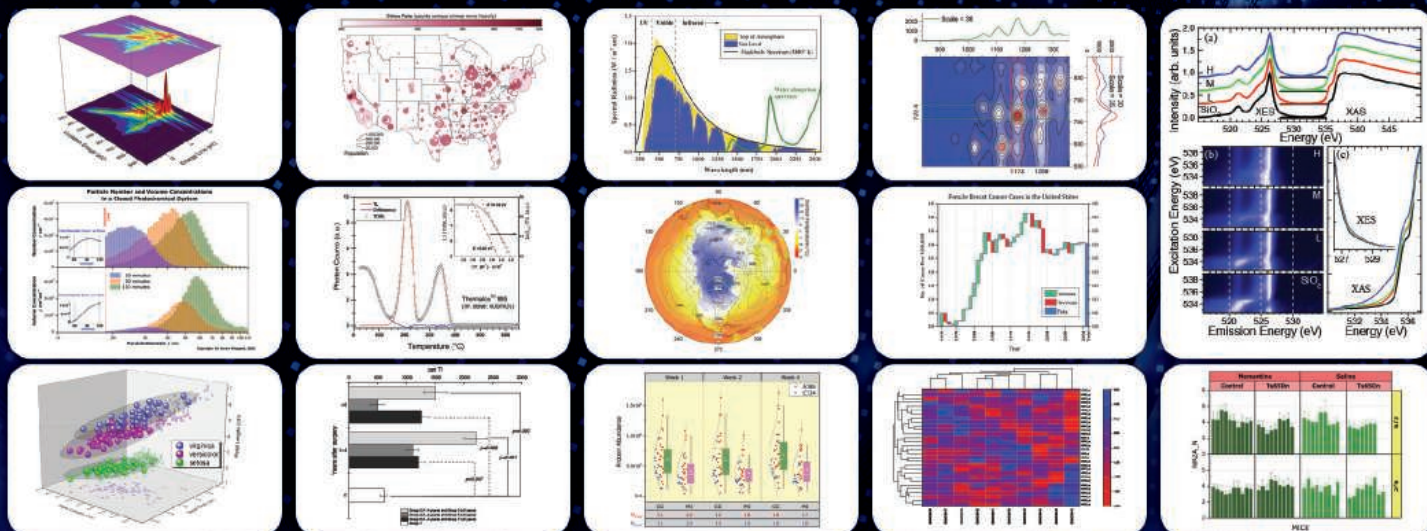
PHONE: +1.415.883.0128 | FAX: +1.415.883.0572  
EMAIL: [INFO@SUTTER.COM](mailto:INFO@SUTTER.COM) | [WWW.SUTTER.COM](http://WWW.SUTTER.COM)



# ORIGIN® 2018

Graphing & Analysis

**New Version!**



Over 75 New Features & Apps in Origin 2018!

Over 500,000 registered users worldwide in:

- 6,000+ Companies including 20+ Fortune Global 500
- 6,500+ Colleges & Universities
- 3,000+ Government Agencies & Research Labs

For a **FREE** 60-day evaluation, go to [OriginLab.Com/demo](http://OriginLab.Com/demo) and enter code: 7564

**OriginLab®**

25+ years serving the scientific & engineering community



2019 **MRS**® SPRING MEETING & EXHIBIT

April 22–26, 2019 | Phoenix, Arizona

# CALL FOR PAPERS

**Abstract Submission Opens**  
September 28, 2018

**Abstract Submission Closes**  
October 31, 2018

Spring Meeting registrations include MRS Membership July 1, 2019 – June 30, 2020

## BROADER IMPACT

- BI01 High Impact Practice—Increasing Ethnic and Gender Diversification in Engineering Education

## CHARACTERIZATION, PROCESSING AND THEORY

- CP01 Advances in *In Situ* Experimentation Techniques Enabling Novel and Extreme Materials/Nanocomposite Design
- CP02 Design and *In Situ* TEM Characterization of Self-Assembling Colloidal Nanosystems
- CP03 Advances in *In Situ* Techniques for Diagnostics and Synthetic Design of Energy Materials
- CP04 Interfacial Science and Engineering—Mechanics, Thermodynamics, Kinetics and Chemistry
- CP05 Materials Evolution in Dry Friction—Microstructural, Chemical and Environmental Effects
- CP06 Smart Materials for Multifunctional Devices and Interfaces
- CP07 From Mechanical Metamaterials to Programmable Materials
- CP08 Additive Manufacturing of Metals
- CP09 Mathematical Aspects of Materials Science—Modeling, Analysis and Computations

## ELECTRONICS AND PHOTONICS

### *Soft Organic and Bimolecular Electronics*

- EP01 Liquid Crystalline Properties, Self-Assembly and Molecular Order in Organic Semiconductors
- EP02 Photonic Materials and Devices for Biointerfaces
- EP03 Materials Strategies and Device Fabrication for Biofriendly Electronics
- EP04 Soft and Stretchable Electronics—From Fundamentals to Applications
- EP05 Engineered Functional Multicellular Circuits, Devices and Systems
- EP06 Organic Electronics—Materials and Devices

### *Semiconductor Devices, Interconnects, Plasmonic and Thermoelectric Materials*

- EP07 Next-Generation Interconnects—Materials, Processes and Integration
- EP08 Phase-Change Materials for Memories, Photonics, Neuromorphic and Emerging Application
- EP09 Devices and Materials to Extend the CMOS Roadmap for Logic and Memory Applications
- EP10 Heterovalent Integration of Semiconductors and Applications to Optical Devices
- EP11 Hybrid Materials and Devices for Enhanced Light-Matter Interactions
- EP12 Emerging Materials for Plasmonics, Metamaterials and Metasurfaces
- EP13 Thermoelectrics—Materials, Methods and Devices

## ENERGY AND SUSTAINABILITY

### *Energy Storage*

- ES01 Organic Materials in Electrochemical Energy Storage
- ES02 Next-Generation Intercalation Batteries
- ES03 Electrochemical Energy Materials Under Extreme Conditions
- ES04 Solid-State Electrochemical Energy Storage

### *Catalysis, Alternative Energy and Fuels*

- ES05 Cooperative Catalysis for Energy and Environmental Applications
- ES06 Atomic-Level Understanding of Materials in Fuel Cells and Electrolyzers
- ES07 New Carbon for Energy—Materials, Chemistry and Applications
- ES08 Materials Challenges in Surfaces and Coatings for Solar Thermal Technologies
- ES10 Rational Designed Hierarchical Nanostructures for Photocatalytic System
- ES11 Advanced Low Temperature Water-Splitting for Renewable Hydrogen Production via Electrochemical and Photoelectrochemical Processes

- ES12 Redox-Active Oxides for Creating Renewable and Sustainable Energy Carriers

### *Water-Energy Materials and Sustainability*

- ES09 Advanced Materials for the Water-Energy Nexus
- ES13 Materials Selection and Design—A Tool to Enable Sustainable Materials Development and a Reduced Materials Footprint

- ES14 Materials Circular Economy for Urban Sustainability

### *Photovoltaics and Energy Harvesting*

- ES15 Fundamental Understanding of the Multifaceted Optoelectronic Properties of Halide Perovskites
- ES16 Perovskite Photovoltaics and Optoelectronics
- ES17 Perovskite-Based Light-Emission and Frontier Phenomena—Single Crystals, Thin Films and Nanocrystals
- ES18 Frontiers in Organic Photovoltaics
- ES19 Excitonic Materials and Quantum Dots for Energy Conversion
- ES20 Thin-Film Chalcogenide Semiconductor Photovoltaics
- ES21 Nanogenerators and Piezotronics

## QUANTUM AND NANOMATERIALS

- QN01 2D Layered Materials Beyond Graphene—Theory, Discovery and Design
- QN02 Defects, Electronic and Magnetic Properties in Advanced 2D Materials Beyond Graphene
- QN03 2D Materials—Tunable Physical Properties, Heterostructures and Device Applications
- QN04 Nanoscale Heat Transport—Fundamentals
- QN05 Emerging Thermal Materials—From Nanoscale to Multiscale Thermal Transport, Energy Conversion, Storage and Thermal Management
- QN06 Emerging Materials for Quantum Information
- QN07 Emergent Phenomena in Oxide Quantum Materials
- QN08 Colloidal Nanoparticles—From Synthesis to Applications

## SOFT MATERIALS AND BIOMATERIALS

- SM01 Materials for Biological and Medical Applications
- SM02 Progress in Supramolecular Nanotheranostics
- SM03 Growing Next-Generation Materials with Synthetic Biology
- SM04 Translational Materials in Medicine—Prosthetics, Sensors and Smart Scaffolds
- SM05 Supramolecular Biomaterials for Regenerative Medicine and Drug Delivery
- SM06 Nano- and Microgels
- SM07 Bioinspired Materials—From Basic Discovery to Biomimicry

[www.mrs.org/spring2019](http://www.mrs.org/spring2019)

### Meeting Chairs

**Yuping Bao** The University of Alabama

**Bruce Dunn** University of California, Los Angeles

**Subodh Mhaikar** Nanyang Technological University

**Ruth Schwaiger** Karlsruhe Institute of Technology—  
Institute for Applied Materials

**Subhash L. Shinde** University of Notre Dame

### Don't Miss These Future MRS Meetings!

**2019 MRS Fall Meeting & Exhibit**

December 1–6, 2019, Boston, Massachusetts

**2020 MRS Spring Meeting & Exhibit**

April 13–17, 2020, Phoenix, Arizona



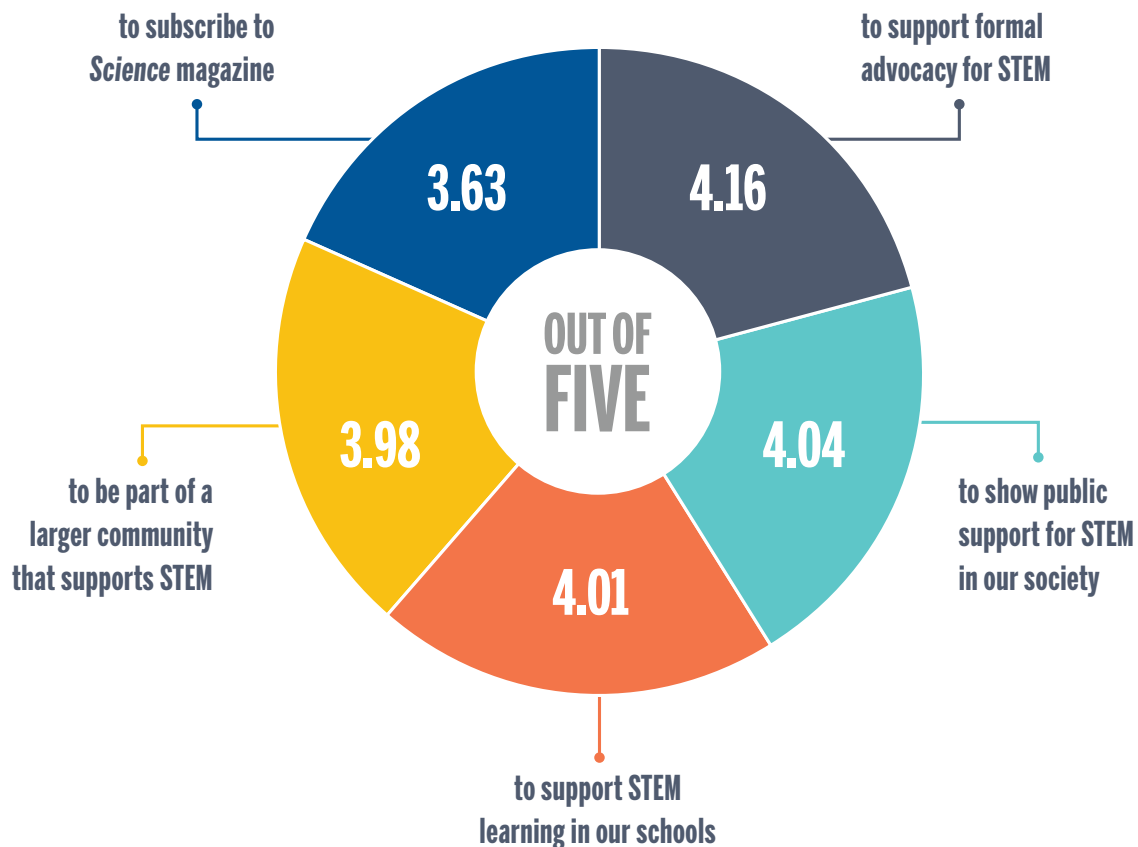
**MATERIALS RESEARCH SOCIETY®**  
*Advancing materials. Improving the quality of life.*

506 Keystone Drive • Warrendale, PA 15086-7573

Tel 724.779.3003 • Fax 724.779.8313 • [info@mrs.org](mailto:info@mrs.org) • [www.mrs.org](http://www.mrs.org)

# AAAS IS THE FORCE FOR SCIENCE

According to the 2017 Member Survey, you joined AAAS ...

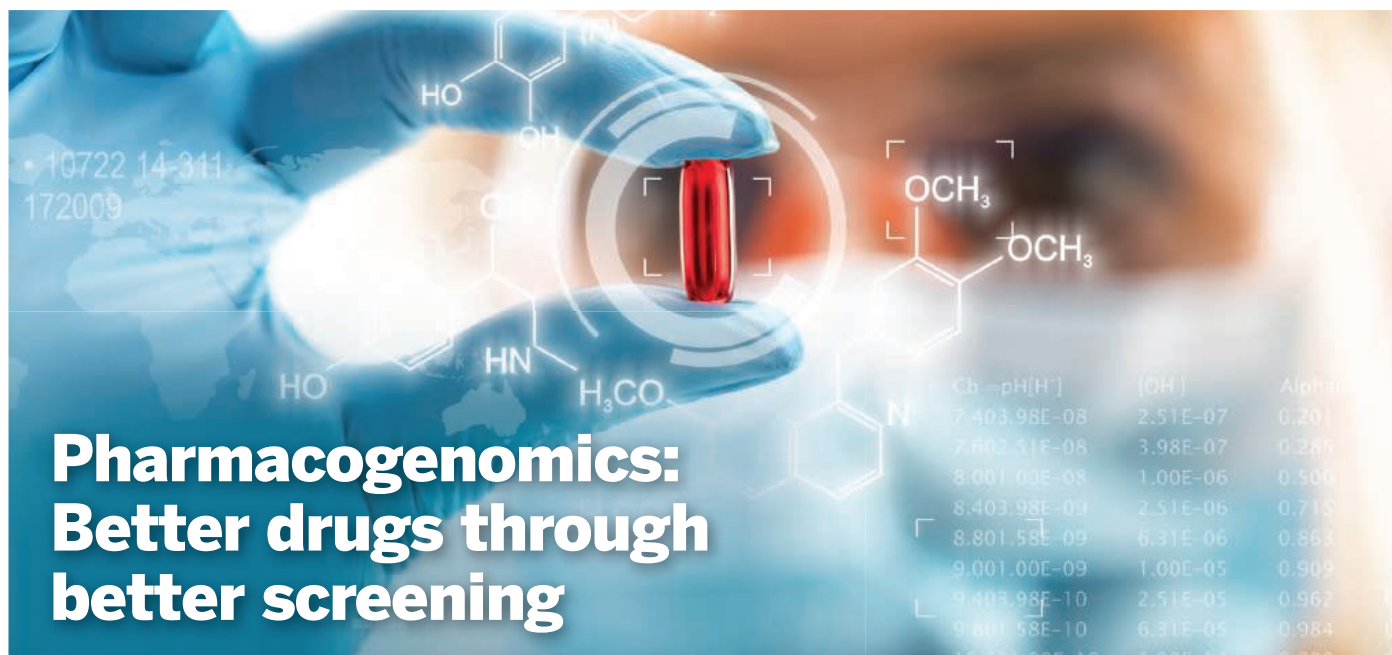


## TELL US WHAT'S IMPORTANT TO YOU!

The 2018 Member Survey is launching in September. Look in your inbox for a link.

Your responses help us to better serve science, scientists, and the global community.  
Don't miss your chance to tell us what's most important to you!





# Pharmacogenomics: Better drugs through better screening

While it's commonplace to test for liver and kidney function before prescribing certain drugs, that's not so much the case when it comes to testing for how drugs might interact with our genes—even though a variant gene can pose as much risk as compromised hepatic or renal function.

By Josh P. Roberts

**D**ata can be dangerous, especially when it comes to drugs, and the numbers do not always tell the whole story. Consider the cytochrome P450 superfamily of enzymes that metabolizes a wide range of drugs. The PharmVar database lists nearly 200 missense variants for the gene coding for a single enzyme, CYP2C19—CYP2C19 is one of 13 genes in the CYP2 family. However, only a relatively small number of these variants have been functionally validated.

While enzymes such as CYPs play important roles in breaking down and inactivating drugs, they are also responsible for activating others. For example, clopidogrel is an antiplatelet medication that is bioactivated through cleavage of the precursor prodrug by CYP2C19. A patient with a loss-of-function mutation in CYP2C19 won't get "the antiplatelet effect you want in order to prevent in-stent thrombosis and other catastrophic events after somebody has a coronary stent," says Dan Roden, Sam Clark Professor of Experimental Therapeutics at **Vanderbilt University** in Nashville, Tennessee. "It's a manifestation of what I'm fond of saying: The most common adverse drug event is failure of the drug to do what you want it to do."

Just as the genome can influence height or eye color, it can influence drug response phenotypes as well. The result can be altered drug efficacy, an adverse reaction, death, or a new pharmaceutical target. Finding the connections between genetic variants and drugs, understanding their

mechanisms of action, and learning how to mitigate, manipulate, circumvent, and perhaps harness those differences is what pharmacogenetics and its successor pharmacogenomics (PGx) are all about. Despite a few clear, well-established examples, the need to demonstrate these connections by establishing the evidence base, and to educate stakeholders such as patients, providers, regulatory agencies, and payers, remains an obstacle to PGx's progress.

## Paving the way to better drug labeling

Most of the early and best-known examples of pharmacogenetics involved the way genetic variation altered a drug's pharmacokinetics (PK). PK factors—absorption, distribution, metabolism, and excretion, (ADME)—influence how much active drug reaches its target.

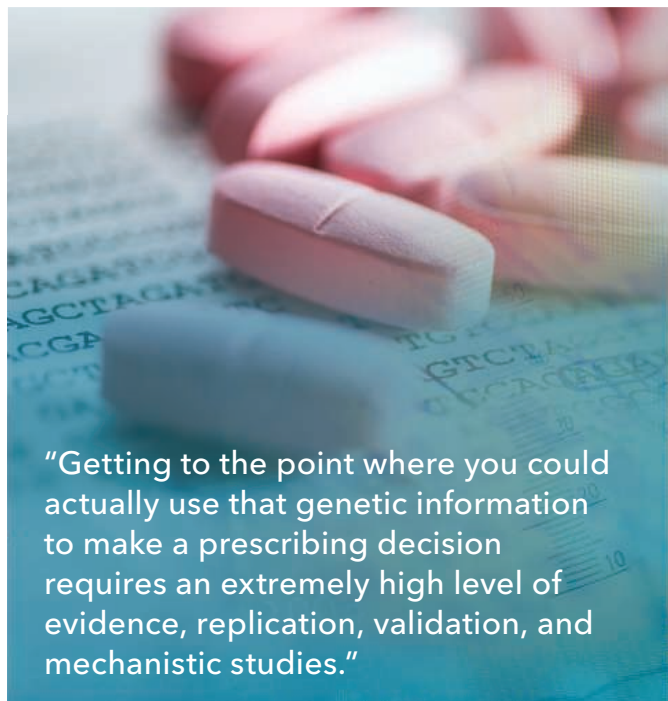
Take the case of 6-mercaptopurine (6-MP), a drug that revolutionized both transplant medicine and the treatment of childhood acute lymphoblastic leukemia (ALL). It acts by interfering with rapid cell turnover and DNA replication; but in about one of 300 children, 6-MP wipes out the bone marrow, with potentially fatal consequences. "So my lab said, 'let's look at the enzyme that catalyzes a major step in the biotransformation of 6-MP, thiopurine methyltransferase (TPMT),' " says Richard Weinshilboum, professor of medicine at **Mayo Clinic** in Rochester, Minnesota.

Two TPMT variants were found that caused the proteins to be rapidly degraded, preventing them from adequately metabolizing 6-MP, "which means that a standard dose was, in their case, about 10 to 15 times too much, because their body couldn't get rid of it," explains Weinshilboum.

The work was published around 1980. By 1990, it was standard practice at Mayo Clinic to test for the variants before administering chemotherapy. Around 2002, the U.S. **cont.>**

## Upcoming features

Animal Models: Better Mouse Models—October 12 ■ Tissue/Cell Culture: Expression Systems—November 16 ■ DNA/RNA: Improving ChIP Assays—December 7



"Getting to the point where you could actually use that genetic information to make a prescribing decision requires an extremely high level of evidence, replication, validation, and mechanistic studies."

Food and Drug Administration (FDA) held first-of-their-kind public hearings on incorporating that information into drug labeling, recalls Weinshilbourn.

### Casting a wider net

With the Human Genome Project, scientists had a reference map for the entire human genetic sequence and could cast a wider net in their search for genetic variants that impact drug metabolism. They could look beyond known drug metabolizing genes to "genes having to do with disease pathophysiology and the mechanisms of drug response, that also show genetic variation," says Weinshilbourn. Recognizing this, the National Institute of General Medical Sciences at the U.S. National Institutes of Health in Bethesda, Maryland, established the Pharmacogenomics Research Network (PGRN). "That was a key in terms of opening the door for PGx to begin to take genome-wide approaches, which allow you to agnostically identify genes that we weren't teaching about to the medical students," he says.

Many of the pharmacogenes found had to do with pharmacodynamics—defined as everything involved with a disease's biology and how a drug does its job if PK is held constant, says Roden.

One example involves the human leukocyte antigen (HLA) complex of the immune system. For instance, more than half of patients with the HLA-B\*57:01 allele who are given the anti-HIV drug abacavir will get a potentially fatal hypersensitivity reaction known as Stevens-Johnson syndrome. But this has "virtually disappeared from clinical practice because every patient who comes to an HIV clinic who was going to receive abacavir has an HLA test done, and if they [test positive for HLA-B\*57:01], they get a different drug," says Munir Pirmohamed, director of the MRC Centre for Drug Safety Sciences and Wolfson Centre for Personalized Medicine at the **University of Liverpool** in the United Kingdom.

"Now we're in an era where we are rapidly identifying, through the techniques of genome-wide association studies [GWAS] and then next-generation sequencing, variants that indicate why the drugs we use today show different responses, and are potential targets into drugs of tomorrow," says Weinshilbourn.

**Thermo Fisher Scientific's** PharmacoScan GWAS platform, for example, "pretty much comprehensively covers all known [pharmacogenes], including those where the evidence is not as strong yet," says Ulrich Broeckel, founder and CEO of **RPRD Diagnostics** and professor of pediatrics at the **Medical College of Wisconsin**, Milwaukee, who uses this technology. "It is a clinical and diagnostic platform, but it's also a research platform."

### Sorting out the evidence

While some of the variation being picked up in the postgenome sequence era is shared by appreciable portions of the population, 30% to 40% of all variability in drug response is due to rare mutations found in less than 1% of the population, says Magnus Ingelman-Sundberg, vice chairman and section head of the Department of Physiology and Pharmacology at **Karolinska Institutet** in Stockholm.

Pharmacogenomic findings far outpace the field's ability to make sense of them, whether they involve rare variants, common variants that don't always lead to a phenotype, or variants of unknown significance. In addition, compliance, drug-drug interactions, and other environmental factors may play a role as well, leading Ingelman-Sundberg to estimate that only "25%-35% of the variability in drug response is caused by genetic factors in the clinic."

But even if it's all driven by genetics, for some drugs a single common genetic variant has a large effect size, while for others there can be "100 different variants, and the combination of variants together will determine whether you respond to a drug or not. And if you have 100 different variants in the same or different gene, proving that it's a certain combination of variants that drive responses in one direction or the other is much, much more difficult—I would even say impossible," says Roden.

Thus, discerning the importance of a pharmacogenetic finding and putting it into action is not a trivial task.

"There are thousands of scientific articles linking a gene with a drug," says Mary Relling, chair of pharmaceutical sciences at **St. Jude Children's Research Hospital** in Memphis, Tennessee. "But getting to the point where you could actually use that genetic information to make a prescribing decision requires an extremely high level of evidence, replication, validation, and mechanistic studies—and importantly, all of that information needs to be present for the alternative therapeutic recommendations one is going to make."

There are criticisms that many studies have failed to demonstrate efficacy in a clinical setting. "That is correct to some extent, because some studies haven't been big enough," says Pirmohamed. But the problem is compounded by the fact that many critics use randomized control trials as their standard. He encourages the use of other types of study design, perhaps taking advantage of

## Featured participants

**Karolinska Institutet**  
ki.se

**Mayo Clinic**  
www.mayoclinic.org

**Medical College of Wisconsin**  
www.mcw.edu

**Myriad Genetics**  
myriad.com

**RPRD Diagnostics**  
www.rprdx.com

**St. Jude Children's Research Hospital**  
www.stjude.org

**University of Florida College of Pharmacy**  
pharmacy.ufl.edu/education

**University of Liverpool**  
www.liverpool.ac.uk

**Vanderbilt University**  
www.vanderbilt.edu

## Additional resources

**Clinical Pharmacogenetics Implementation Consortium (CPIC)**  
cpicpgx.org

**Pharmacogenomics Research Network (PGRN)**  
www.pgrn.org

**PharmGKB**  
www.pharmgkb.org

**PharmVar**  
www.pharmvar.org

collective experience, and “using real-world evidence from that to show that it is having a clinical benefit in our populations.”

There are nearly 200 FDA-approved drugs with pharmacogenomic information on their labels, seven of which have strict warnings, says Broeckel.

“In terms of clinical application of PGx, we’re at the very early stages. I think we have examples where we have some really solid evidence that it makes a difference in the lives of patients. But we must continue to build that evidence base,” says Johnson.

The Clinical Pharmacogenetics Implementation Consortium (CPIC)—a joint effort of PGRN and the PharmGKB knowledge base—puts together a “freely available, nonprofit, evidence-based, expert-driven resource of very specific guidelines for the very, very few genes and drugs for which the evidence was already strong enough that if you had genetic variation information on a patient, it would be wrong not to use it to inform prescribing,” says Relling. To date, CPIC has 20 guidelines affecting 15 genes and 37 drugs.

Even so, says Relling, a recent survey of a major academic health care system found that only in about 1.5% of prescriptions for which the label requires or recommends a PGx test was that test performed (even for well-established drug-gene pairs), and that seems to be the norm in both academic and mainstream medicine. There are many reasons for this, from issues of education, technology, and infrastructure to the economics of health care.

## Breaking the bottleneck

In some ways, PGx faces a classic Catch-22 scenario: Evidence is needed to convince regulators, professional societies, providers, and payers alike that using genetic information makes a difference to patient outcomes and

the bottom line, but that evidence has first to be generated. And most patients and clinicians simply aren’t aware of when and where PGx testing should be used, if they are aware of it at all, says Relling.

Some major academic medical centers “are really trying to move the field forward” by actively translating discoveries into clinical practice, says Julie Johnson, dean and distinguished professor at the **University of Florida College of Pharmacy** in Gainesville. The University of Florida, she says, has published “pretty compelling evidence not only that cardiologists across a multitude of different institutions will use [genetic] information to guide therapy, but that when they do use it, those patients have better clinical outcomes.” This finding was followed by a chronic pain management study in primary care settings—some in medically underserved, minority communities—leading to genotype-guided outcomes significantly better than the standard of care, and in most cases with a reduction or even elimination of opioid use.

Several institutions and consortia are studying the efficacy and best practices of preemptively sequencing or genotyping patients for a host of pharmacogenes—most often as panels—and embedding that data into their electronic health records (EHRs). Rather than the alert firing when there is a test available, Mayo Clinic is experimenting with it firing only when genetic information in the EHR counterindicates the prescription.

Either way, sometimes you need the medication immediately and can’t wait a week or two for genetic tests to come back, says Timothy Curry, education director of the Center for Individualized Medicine at Mayo Clinic. But he cautions that “just putting it into the medical record wouldn’t have been useful if we didn’t also accompany that with education—teaching people how to use that information.”

Preemptive or reactive, PGx panels are also useful when a combination of variants is implicated in drug choice. It is becoming more commonplace to look at a panel of mutations in the somatic tumor genome to assess how to treat the cancer, for example. Similarly, for variations in the germline genome for fields such as psychiatry, “the goal is to end traditional trial-and-error prescribing, which has largely been dictated by the single-gene approach,” says Bryan Dechairo, executive vice president of **Myriad Genetics**, a molecular diagnostics company based in Salt Lake City, Utah.

There are already some drug-gene combinations for which the PGx evidence is considered incontrovertible, in terms of both patient outcome and the cost benefit of testing. Drug labels are calling for testing, and EHRs are altering prescribers. Links between other drugs and other genes are constantly being uncovered. The future of PGx will likely be built on a critical mass of stakeholders—including patients and providers informed enough to ask for it.

Josh P. Roberts is a freelance writer based in Minneapolis, Minnesota.



## New products: genomics

**Transfection Reagent**

Fuse-It-siRNA is a transfection reagent that enables rapid, efficient gene silencing with the highest biocompatibility, even in sensitive and difficult-to-transfect cells, such as primary keratinocytes. Using Fuse-It-siRNA, researchers can focus directly on the knockdown phenotype

without being distracted by side effects of the transfection method, such as cell stress and toxicity. The principle is simple: The Fuse-It liposomal carrier, which includes small interfering RNA (siRNA), fuses with the cell membrane and then releases the siRNA directly into the cytoplasm. Thus, siRNA is immediately incorporated into the RNA-induced silencing complex (RISC), leading to efficient gene knockdown without the interfering processes of endocytosis or lysosomal degradation. Because it is based on the charge of natural cell membranes, Fuse-It-siRNA can be used with most cell types. Researchers performing knockdown studies in sensitive primary cells will greatly benefit from its extremely low cytotoxicity.

**ibidi**

For info: +49-(0)-89-520-46-17-0

[www.ibidi.de](http://www.ibidi.de)

**ELISA Kit for Endotoxins**

The EndoCab ELISA has been developed for determination of endotoxin core antibodies in human plasma or serum in patients or healthy individuals. These antibodies play an important role in the clearance of pathogens in both disease and medical interventions. One hypothesis suggests that if patients' preoperative endogenous endotoxin-core antibody (EndoCAb) level is low or even moderately low, they may be unable to cope with the efflux of endotoxin, which may have mild to severe clinical consequences. The assay is of interest for various experimental conditions ranging from in vitro lipopolysaccharide neutralization by plasma components and in vivo studies on kinetics of antibodies to endotoxin levels in health and disease. The EndoCab standard IgG, IgM, and IgA median units are arbitrary and are based on medians of ranges for 1,000 healthy adults in a particular locality. Users should establish appropriate local statistical controls for their studies.

**Hycult Biotech**

For info: 855-249-2858

[www.hycultbiotech.com/hk504](http://www.hycultbiotech.com/hk504)

**Multiplex DNA Reference Standard**

Horizon Discovery's OncoSpan is a cell line-derived, multiplex DNA reference standard for validating NGS assays. It features 385 variants across 152 key cancer genes. OncoSpan uses bioinformatics and droplet digital PCR to help drive fast, easy, and more complete validation of oncology gene panels and exome sequencing assays. Horizon also offers OncoSpan customers an online companion NGS quality-control solution called OncoMatic, which enables them to upload OncoSpan NGS data after sequencing, automatically assessing the called variants, reporting on variant frequency data, and tracking several quality metrics per sample. The combination of OncoSpan and OncoMatic provides access to an accurate reference standard truth set for use during establishment

and validation of critical NGS bioinformatics pipelines, saving costs and helping to maintain compliance with standards such as those set by the College of American Pathology and EuroGentest.

**Horizon Discovery**

For info: +44-(0)-1223-976-160

[www.horizondiscovery.com](http://www.horizondiscovery.com)

**Real-Time PCR System**

The AriaMx Real-Time PCR System combines a thermal cycler, an advanced optical system with an LED excitation source, and a complete data analysis software package. The instrument can hold up to six optics modules, and the scanning optics design delivers optimal separation between the dyes and between samples. It features a closed-tube PCR detection format compatible with a variety of fluorescence detection chemistries, such as SYBR Green and EvaGreen dyes, and fluorogenic probe systems including TaqMan probes. AriaMx amplifies your productivity with its unique modular and flexible design; intuitive touchscreen interface; advanced, easy-to-use reporting; and 120-plus attributes monitored via its onboard diagnostics, which pinpoint assay or instrument issues as they arise. Use AriaMx for a variety of research applications, including gene expression analysis, genotyping/high-resolution melt capability, mRNA quantification, NGS quantification (library prep, result validation), nucleic acid monitoring, and rare-allele detection.

**Agilent Technologies**

For info: 800-227-9770

[www.genomics.agilent.com](http://www.genomics.agilent.com)

**Gene Expression Panels**

nCounter Gene Expression Panels are developed in collaboration with leading researchers to deliver the most valuable, relevant panel content across applications, providing a highly multiplexed, simple workflow—with no library prep or dilutions needed, and just 15 min total hands-on time required. Address key challenges in immuno-oncology by measuring multianalyte expression levels simultaneously within the tumor, the immune system, and the tumor microenvironment all on one platform, with NanoString's 3D Biology Technology. nCounter Immunology, Inflammation, and Myeloid Panels are all designed to evaluate biomarkers associated with autoimmune diseases and can be customized with up to 30 additional user-defined genes. Furthermore, nCounter Immunology Panels are ideal for assessment of infectious disease immune response and for pathogen identification.

**NanoString Technologies**

For info: 888-358-6266

[www.nanostring.com](http://www.nanostring.com)

**Automated DNA/RNA Extraction**

ExiPrep 16 Plus is a laboratory automation system able to extract DNA or RNA from various biological samples, including tissue, whole blood, serum, plasma, swab, bacteria, and yeast. It also features the ability to purify His-tagged recombinant proteins. The system can handle up to 16 samples at once and includes optimized extraction protocols for many sample types, giving you unsurpassed results from your nucleic acid or protein purification. ExiPrep 16 Plus has advanced features including an intuitive touchscreen, a UV sterilization cycle, and a unique contamination shield, each designed to provide an extra level of confidence in your findings.

**Bioneer**

For info: 877-264-4300

[us.bioneer.com](http://us.bioneer.com)

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.

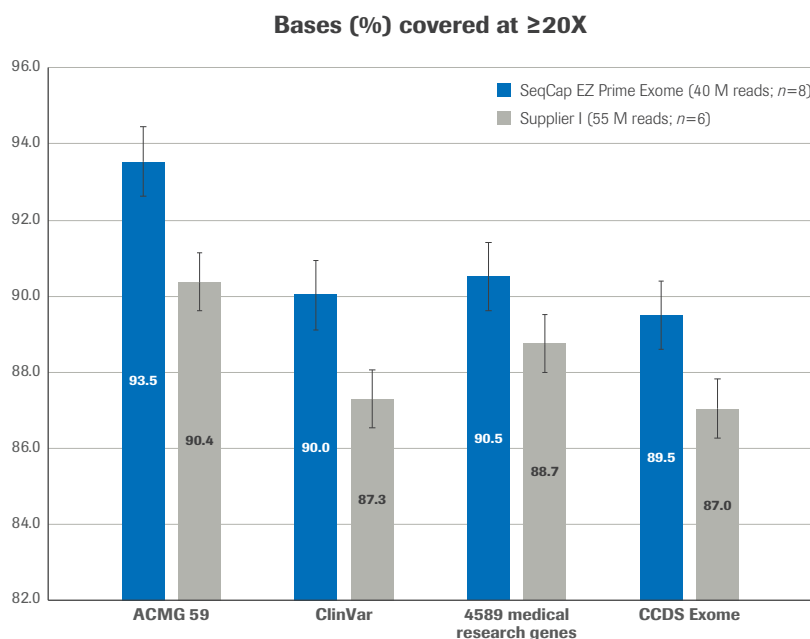
## SeqCap EZ Prime Exome

Enhance sequencing efficiency with SeqCap EZ Prime Exome. Choose a 37 Mb panel covering 98.1% of the coding regions of CCDS Rel. 20.\*

- **Greater coverage and efficiency:** Cover more target bases at  $\geq 20X$  with only 40 M reads than Supplier I's exome covers with 55 M reads (Figure 1)\*
- **Sample tracking:** Enhance sample identification by detecting 340 carefully selected SNPs\*
- **Integrated workflow:** Streamline ordering and support with a single trusted vendor from library prep through target enrichment

**27% LESS SEQUENCING**  
compared to Supplier I

**REQUEST A SAMPLE**  
[go.roche.com/DesignShareSample](http://go.roche.com/DesignShareSample)



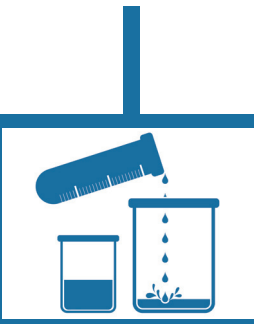
**Figure 1. SeqCap EZ Prime Exome provides better coverage of target regions with 27% less sequencing compared to Supplier I.** For the Prime Exome samples, target-enriched libraries were prepared from 100 ng of Coriell DNA using the HyperCap Workflow, with the KAPA HyperPlus Kit and the SeqCap EZ Prime Exome Panel. The same workflow was followed for the other samples, using Supplier I's probes in place of Prime Exome. Prime Exome required 4 Gb of sequencing vs 5.4 Gb required by Supplier I. Sequencing was performed using an Illumina® HiSeq® 2500 instrument with 2 x 101 read length using v4 chemistry. For each panel, 8 replicates were prepared; only 6 of the alternative supplier's samples yielded sufficient reads.



# Making the most of your **antibodies**



Antibodies are an essential part of many modern biology labs. Used correctly, they are a powerful tool ... but they need to be treated with care and respect.

**Here are some tips** for getting the most out of your precious antibody reagents.




## 1 SETUP

- confirm that your antibody has been **validated** for the intended application
- **titrate** your antibody to determine the optimal working conditions and concentrations
- **carefully store** your reagents according to manufacturers' recommendations
- limit **excessive handling** by aliquoting antibody reagents.



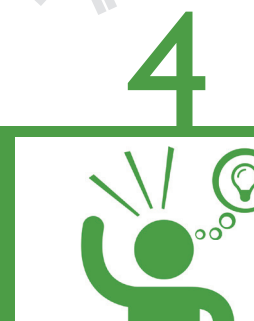
## 2 DESIGNING THE EXPERIMENT

- do all **dilution calculations** in advance (and double-check them)
- carefully select **controls** (e.g., positive, negative, and nonspecific binding)
- prepare and review your **protocol** sheet
- **set up and label** all tubes and plates in advance.



## 3 DOING THE EXPERIMENT


- **follow your protocol** carefully, checking off each step as it's done
- **handle antibody reagents with care**—don't overmix or leave at room temperature
- **pipet reagents** carefully and accurately.



## 4 TROUBLESHOOTING

**If things didn't turn out as expected:**

- carefully **check** all calculations and dilutions
- **check your protocol** against manufacturers' recommendations
- do your **controls** help you identify the source of the problem?
- confirm **compatibility** of your primary antibody with secondary antibody and other reagents
- are all of your reagents **fresh** (especially blocking agents)?
- were the proper **incubation times** used?



Sponsored by

**ThermoFisher**  
SCIENTIFIC

Produced by the *Science*/AAAS  
Custom Publishing Office

**Science**  
AAAS



# Verification of **antibody performance**

Researchers need antibodies that bind to the right target and work in their applications every time. To help ensure superior antibody results, Thermo Fisher Scientific has expanded their specificity and validation\* testing methodologies using a 2-part approach for advanced verification.

## The challenge

Antibodies are some of the most critical research reagents used in the lab. Poor specificity or application performance can lead to inconsistent results, a lack of reproducibility, and a waste of time and money.

**Invitrogen™ antibodies are currently undergoing a rigorous 2-part testing approach**

### TARGET SPECIFICITY VERIFICATION

Helps ensure the antibody will bind to the correct target. Invitrogen antibodies are being tested using at least one of the following methods:

- Immunoprecipitation/mass spectrometry
- Knockout
- Knockdown
- Independent antibody verification
- Cell treatment
- Relative expression
- Neutralization
- Peptide array
- Orthogonal

### FUNCTIONAL APPLICATION VALIDATION

These tests help ensure the antibody works in particular applications of interest, which may include (but are not limited to):

- Western blotting
- Immunofluorescence imaging
- Flow cytometry
- Chromatin immunoprecipitation
- Immunohistochemistry

## The solution

Thermo Fisher Scientific is working to redefine antibody performance with a comprehensive approach to how antibodies are evaluated and validated. By combining specificity testing with extensive application validation data, Thermo Fisher helps ensure that Invitrogen antibodies will help enable superior performance for researchers.

\*The use or any variation of the word "validation" refers only to research use antibodies that were subject to functional testing to confirm that the antibody can be used with the research techniques indicated. It does not ensure that the product(s) was validated for clinical or diagnostic uses.

For Research Use Only. Not for use in diagnostic procedures.

© 2018 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.

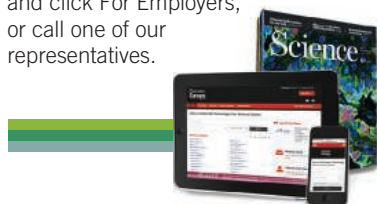
Find out more at:  
[thermofisher.com/antibodyvalidation](https://thermofisher.com/antibodyvalidation)

**ThermoFisher**  
S C I E N T I F I C

# 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)

Advance your career  
with expert advice from  
*Science Careers*.



**Download Free Career Advice Booklets!**

[ScienceCareers.org/booklets](http://ScienceCareers.org/booklets)

### Featured Topics:

- Networking
- Industry or Academia
- Job Searching
- Non-Bench Careers
- And More



**ScienceCareers**

FROM THE JOURNAL SCIENCE AAAS

## POSITIONS OPEN

### Assistant Professor in Disease Ecology and Evolution

The Department of Ecology and Evolutionary Biology at Tulane University invites applications for a tenure-track position at the Assistant Professor Level. Candidates are required to have a Ph.D. in a related field at the time of application and postdoctoral research experience is preferred. The desired area of specialization is Disease Ecology and Evolution. For details about this position and to apply, please see website: <https://apply.interfolio.com/54152>.

**ScienceCareers**

FROM THE JOURNAL SCIENCE AAAS

**Follow us for jobs,  
career advice  
and more!**



@ScienceCareers



/ScienceCareers



Science Careers

[ScienceCareers.org](http://ScienceCareers.org)

**Search more  
jobs online**

Access hundreds  
of job postings on  
[ScienceCareers.org](http://ScienceCareers.org).

**Expand your  
search today.**





## China’s Universities Adjust to Meet “Double-First Class” Initiative

**O**n August 28, 2018, Zhang Yu, one of the faculty members at a university in Northern China, was informed by his colleagues that the “Instructions on Accelerating the Process of Constructing World-Class Universities and First-Class Disciplines” were posted on the website of the Ministry of Education of the People’s Republic of China. He then rushed to read the document on his computer.

These instructions, which Yu valued so much, were released jointly by the Ministry of Education, Ministry of Finance, and National Development and Reform Commission—three Chinese ministries that take full charge of the “Double First-Class” initiative in its planning and deployment, promotion and implementation, and supervision and management. Since the publication of “The Overall Plan for Promoting the Construction of World-Class Universities and Disciplines” in 2015, and of the “Implementation Measures (Provisional) for Promoting the Construction of World-Class Universities and Disciplines” in 2017, the release of the “Instructions” is the latest development in the Double First-Class initiative,

aimed at providing more specified guidelines on construction directions, priorities, and the like for a few universities still in their infancy, and further elaborating on the status of Chinese universities as the main subject of responsibility, construction, and benefits.

To Yu’s relief, there was no mention of his university’s elimination from the initiative in the “Instructions.” This had concerned him because the university had entered into the Double First-Class list while putting on a poor performance in last year’s fourth national evaluation of subjects for China University Subject Rankings. The Double First-Class construction plan is based on subject rankings; however, the development of subjects cannot happen overnight. And what is more, impelled by a round of new policies, all universities are trying in every possible way to comply at breakneck speed or to find workable shortcuts. Outsiders cannot fully grasp the intense competitive atmosphere that this situation has created.

There are many examples of faculty members such as Yu who are concerned about their



**Shixin Wang**  
Deputy Chief Editor of  
China Education Online,  
Chief Executive Editor of  
AcaBridge



future. In fact, the prospect of entering or being eliminated from the Double First-Class list has long strained the nerves of all universities, all because of the “dynamic adjustment mechanism” that lies at the core of the crucial document. This feature underlines the major difference between the Double First-Class plan and the previous 985 Project and 211 Project, which is the introduction of a competitive element into the program.

As we can see from the timeline for the initiative, with the dynamic adjustment mechanism set up to spur improvements by colleges and universities the first grueling test for them will be carried out in 2020, highlighting the importance of the “Instructions” as a connecting document.

Back in March 2017, the Double First-Class plan did not decide the number of universities that would be included in the final list. Consequently, a burst of enormous enthusiasm for participating in the initiative was shown among all Chinese universities, with more than 20 provinces announcing new development and construction programs for local universities in the next five to 10 years. Nevertheless, when the official list was released in September 2017, the outcome was not at all what everyone expected.

Out of the total number of universities in China—nearly 3,000—the 137 selected for the list were just a small fraction. All the traditional “985” and “211” universities were included. Although no longer the driving force behind the development of China’s education, the 985 and 211 Projects still played an important part in this new initiative. Nevertheless, many formerly prestigious comprehensive universities with high hopes of being selected were overtaken by industry-based universities with characteristic subjects. From an alternative perspective, this outcome once again proved the importance of subjects to the Double First-Class initiative.

If we review the list in terms of subjects, China’s First-Class Educational Institutions have focused on building their preponderant disciplines into first-rate ones. The 137 universities are classified into three categories, as (1) Type A and (2) Type B of First-Class Educational Institutions, and (3) Double First-Class discipline universities, which are focused on building their preponderant disciplines. A further exploration into the data reveals that more than 400 subjects taught by those 137 universities have become world-class disciplines. Of those, Peking University and Tsinghua University lead with 41 and 34 world-class subjects, respectively, followed by Zhejiang University with 18, then Fudan University and Shanghai Jiao Tong University, both with 17. These five universities have accounted for 127 world-class subjects, which is 27 percent of the total.

The dominant universities in the list appear to invest more heavily in building existing research teams and at-

tracting top talents both at home and abroad than do other universities. So how will China’s other universities manage to compete with these powerhouses?

It is a question of great concern to experts in specialized fields. Large-scale institutional development has little meaning in today’s environment. On the contrary, the only road many universities can follow successfully will be to focus on building their preponderant disciplines.

According to last year’s fourth national evaluation of academic subjects, 30 universities on the Double First-Class list failed to enter the China University Subject Rankings (Type A), and some of those are even on the list of First-Class Educational Institutions (Type B). No wonder many universities are worried about their spot on the next list.

Yu’s university is no exception. In fact, for many Chinese universities, the dynamic adjustment mechanism in the Double First-Class list is a source of enormous stress. Universities not on the list are scrambling for a spot on it, while universities already on the list are struggling to maintain their position. Some observers claim that for certain provinces, only one or two universities were included on the list, in part to address the gap between different regions. For these provinces, the adjustments that will come in two years may only involve expansion or internal modifications rather than elimination. Of course, measures will be taken according to existing conditions by then.

The introduction of the adjustment mechanism into the Double First-Class initiative has no doubt swept all Chinese universities into the rising tide of educational competition like never before, whether they are on the list or not. Universities on the list, such as Zhejiang University, National University of Defense Technology, Xi’an Jiaotong University, Jinan University, Harbin Engineering University, Southwest Jiaotong University, and Tianjin Polytechnic University, are capitalizing on their early lead to further develop their competence. Other universities not on the list, including ShanghaiTech University, Southern University of Science and Technology, Xi’an University of Science and Technology, South-Central University for Nationalities, and Tianjin Normal University, are also working to increase their standing. The newly authorized Westlake University is established under China’s modern education system and will serve as an important model for future educational institutions. All these schools present a great opportunity for scholars in many different fields to showcase their talents in China.

We welcome excellent scholars interested in applying for talent programs to contact us through AcaBridge ([consultant@acabridge.edu.cn](mailto:consultant@acabridge.edu.cn)), which provides one-on-one consultations. Further information can be found at [www.edu.cn/syl](http://www.edu.cn/syl).

# Systems Science at Beijing Normal University (BNU)

## —Striving for Excellence



Systems science, as a new scientific field, is a featured top-ranking discipline of Beijing Normal University (BNU), geared to the needs of contemporary scientific progress and social development. In order to achieve the goal of the First-class Discipline Plan, BNU has decided to center around developing the fundamental theories of systems science through interdisciplinary collaboration with other disciplines, such as brain and cognitive neuroscience, global change and earth system science, and social governance, thereby deepening our knowledge of all fields concerned and discovering the underlying laws of complex systems.

### The History and Achievements of Systems Science at BNU

Since the early 1980s, there have been four BNU scholars who went to the Free University of Brussels and earned their Ph.D. degrees from Professor Ilya Prigogine, the 1977 Nobel laureate in chemistry and founder of the "Brussels school". Back in China, they introduced self-organization theory into Chinese academia. With the full support of Professor Qian Xuesen (Hsue-shen Tsien), BNU founded the undergraduate program of systems theory in 1985 and initiated the construction of systems science discipline.

In the course of more than 30 years of discipline construction, BNU, by virtue of its unique academic heritage and innovation, has greatly advanced the construction of China's systems science discipline. Having inherited the academic advantages of the "Brussels School", BNU is committed to developing the general concepts and universal methods in complex systems and helps set the direction for the research of the basic theories in complex systems in China. BNU has built the world's first complete framework of systems science talent development and formed a prominent multidisciplinary collaborative research platform, with a great reputation at home and abroad. Its research findings are highly valued by the international academic community, reported and reviewed by such international media as Nature, Science, Science Today, MIT Technology Review, and BBC.

### Opportunities and Challenges in Systems Science

The 21st century is a century of complexity. The development of science and technology has ushered us into the era of researching and regulating complexity. The importance of systems science in the future academic framework has gained the consensus of the international academic community. In the National Plan for the Development of Science and Technology, the Chinese government has made it a point to "the giant open system and complex systems" as a frontier research topic. "Complex systems, disaster formation and predictive control" are listed as basic research areas as required by the national strategies. Driven by the development of information technology, the core issues that arise in many disciplines, such as social economy, biological ecology, resources, environment, and education, tend to be systematic and complex. Social and economic development also brings about overall and complex problems. To solve these problems, systems science is urgently

needed.

### Contents and Objectives of Systems Science Discipline Construction at BNU

As the core discipline of the First-Class Discipline Plan, systems science is expected to spearhead the improvement of BNU's discipline construction, meet the demands of the new era for the development of science and technology, and achieve the goal of the national development strategy.

i. Cultivate high-quality compound talents who understand the basic concepts and master the analytical methods in systems science, so that BNU will become a crucial base for talent training of all levels in systems science.

ii. Strengthen the research on the basic theories of systems science, discover the universal laws underlying complex adaptive systems, and improve the concepts, theories and methods of systems science. To meet the needs of major national strategies, BNU will conduct interdisciplinary research to solve key issues in science, technology and the national economy. Breakthroughs are expected to be made in the areas of group decision making,



brain and cognitive neuroscience, global change, and social governance.

iii. Create a social service platform of systems science, establish a national-level consulting service center based on big data analysis, develop a social service training framework of systems science, disseminate the idea of systems science, and demonstrate the discernible value of systems science.

iv. Build a platform for international exchange and cooperation in systems science, expand the international academic influence of systems science as a discipline, establish the "BNU International Science Center for Mathematics and Complex Systems" and become an important international base for talent training and scientific research in systems science.

Building a first-class team on systems science research is not only one of the goals of discipline construction, but also the basis for achieving its many other goals. BNU will attach equal importance to cultivating the most promising talents and hiring the most qualified experts. We cordially welcome job applicants and visiting scholars with expertise in systems science and related areas.

### For more information, please contact us:

Website: <http://sss.bnu.edu.cn>

Email: [sss@bnu.edu.cn](mailto:sss@bnu.edu.cn)

Address: School of Systems Science, Beijing Normal University, No. 19, Xijiekouwai St., Haidian District, Beijing, P.R.China 100875



**Zhong Denghua,**  
President of Tianjin  
University, Academician of  
the Chinese Academy of  
Engineering



Currently, China's higher education has prioritized the "double world-class project", which brings each university new opportunities and possibilities. What plans and policies does Tianjin University have in place? What are the strong disciplines? How to build them into world-class ones in the coming years?

To construct world-class universities is of strategic importance for the development of China's higher education. Founded in 1895, Tianjin University (TJU) is the first modern university of China, with a long history and outstanding reputation. Now, the university has been included into the national scheme of building world-class universities. In light of the university capacity, a "three-step" growth strategy has been formulated with an aim to build TJU into a world-class university when it celebrates its 150th anniversary.

Disciplines are fundamental to construct world-class universities. In recent years, TJU has been restructuring and optimizing its disciplines, and the quality of disciplines has steadily improved. The latest ESI (Essential Science Indicators) data shows 8 subjects have ranked top one hundredth in the world, including the Materials Science, Engineering, Chemistry, Agricultural Science, Physics, Biology and Biochemistry, Computer Science, Pharmacology and Toxicology. Among them, Engineering, Chemistry and Materials Science enter into the world's top one thousandth. According to the 2017 academic assessment by the Ministry of Education, 14 disciplines of TJU have ranked disciplines A-Level, including Chemical Engineering ranked national No.1 for the four consecutive rounds.

As a comprehensive university with distinct strength in Engineering, TJU has refined its disciplinary layout that integrates Engineering with Sciences, Humanities and Social Sciences and Medicine, while encouraging and fostering cross-disciplinary education and research. Specifically, Tianjin University has highlighted the ten key areas, including Chemical Energy, New Materials, Management and Economics, Chemistry and Life Sciences, Sustainable Building and Environment, Construction Engineering Safety, Intelligent Manufacturing, Photoelectric Information, Data Science, and Environmental Ecology.

The "double world-class project" is inseparable from outstanding faculty members. What specific measures does Tianjin University take to attract and retain talents, especially those from overseas?

A world-class faculty team is the cornerstone of the "double world-class" construction. Focusing on Emerging Engineering Education, TJU is systematically adjusting and optimizing the

disciplinary layout. To that end, the university has promoted comprehensive and in-depth reform of the tenure system that is comparable to international practices so as to recruit and nurture top talents.

At present, through the implementation of the "Peiyang Scholar Scheme Distinguished-CORE", TJU is actively setting up a system for selecting and hiring global talents. The "Peiyang Forum for Young Scholars" will be held twice a year to gather talents around the globe on TJU campus for communication and collaboration. At the same time, the university encourages TJU faculty "going out", with delegations at the university, school, and department level traveling overseas for recruitment. International academic conferences can be an alternative, with TJU participants advertising and introducing the recruitment program to global experts. By providing internationally competitive salary, sound household insurance, advanced research equipment, and respect and attention, the university invites all talents to join TJU family with great sincerity.

Concurrently, Tianjin University is committed to building a "Peiyang Faculty Tenure System", which provides favorable conditions for the development and nurturing of overseas talents including the establishment of independent innovation research funds, the construction of multi-disciplinary research platforms, the establishment of large-scale public instruments and equipment sharing experimental centers. With adequate support and excellent services, TJU has been fostering an academic eco-system that is conducive to the attraction, growth, and clustering of overseas talents.

By actively implementing a series of recruitment and nurturing policies, Tianjin University strives to attract and retain top talents, with an aim to form a teaching team with strong expertise and appropriate structure. This sets a solid foundation for building itself into a world-class university. As the president of Tianjin University, I'd like to welcome all the talents at home and abroad to join us to work for a brighter future of the university!

#### CONTACT METHODS:

CONTACT: Zhang Yinlu, Zhang Borui, Li Dachao, Department of Human Resources

TEL: +86 22 27402079

MAIL: talent@tju.edu.cn

ADD: No. 135 Yaguan Road, Haihe Education Park, Jinan District, Tianjin, 300350





Yao Zhi, Secretary of  
CPC Committee



Yan Hua, President of  
Tianjin Medical University



## Double First-Class Initiative for Excellence in Medical Education

**Tianjin Medical University (TMU)**, one of the premier medical universities in China, aims to become a world-recognized leading medical university oriented towards research. It is also dedicated to cultivating high-level medical specialists by focusing on education.

Located in the center of Tianjin, one of the most important economic hubs in North China and an international port southeast of Beijing, half an hour's distance by train. The university was established in 1951 as the first medical university approved by the Chinese Government after the founding of the People's Republic of China. Its founding president was internationally acknowledged endocrinologist Xian Yi Zhu. TMU is now a unique and comprehensive medical university that offers programs in the full spectrum of medical disciplines. It has 21 schools and departments in Basic Medical Sciences, Clinical Medicine, Pharmacy, Public Health and Nursing among others, owns 6 hospitals, contracts with 15 hospitals for the purpose of teaching, and partners with 40 hospitals. All of these facilities provide the university with the greatest resources for patient care and medical training in Tianjin. Among them, the General Hospital of TMU is the largest comprehensive hospital in Tianjin, while the Cancer Hospital of TMU is a leading oncology research center nationwide. Also, TMU possesses a unique hospital specialized in endocrinology and metabolic diseases. The ESI shows that five disciplines of TMU, including Clinical Medicine, Biology and Biochemistry, are ranked in the top 1% globally. TMU adopts an educational mode of combining system-based curricula with practical learning at our hospitals. Every year, around 270 new undergraduate students are enrolled in the Clinical Medicine program where they can develop skills, attitudes and knowledge required by future physician practice. TMU also provides 133 M.D and Ph.D. fellows majoring in Clinical Medicine with a very supportive environment for them to be cultivated into top-level medical workers with both a solid foundation in medicine and excellent clinical and research capacity. With over 120 international students from 48 countries every year, TMU has become one of the ideal destinations for international medical education.

Through past decades to the present, TMU has continued to promote high quality medical education, research and healthcare. In 1996, it was listed as one of the key universities na-

tionwide in the "211 Program", the only university in Tianjin to gain this recognition.

In 2017, due to its strength in Clinical Medicine, TMU was recognized as a World-class Disciplinary Development University in the Double First Class Initiative, the only domestic medical university to receive this honor.

In the discipline of Clinical Medicine, they focus on the basic research and clinical practice of oncology. The strategy is depicted as "1+x", with the Basic Medical Research Center as the "1". "x" refers to such specialized research fields of oncology as epigenetics, tumor microenvironments and precision medicine in tumor treatment. They place emphasis on advancing TMU's core values, cultivating excellent medical personnel and making frontier R&D efforts in order to serve the society.

TMU encourages interdisciplinary research and international collaboration so as to attract highly qualified medical experts to work in it. Currently, more than 40 distinguished scholars from all over the world, including 3 academicians, are engaging in cutting-edge research and translational medicine at TMU. In the next 3 years, TMU plans to recruit another 100 talents and experts in medical research, pledging to provide them with abundant resources and state of the art research facilities. TMU has set up PI posts, hoping to attract high-level HR from all over the world who are leading in basic medicine and clinical medicine. They are ready to provide holistic support to them in fund, policy, and admission of postgraduate students, physical space, and other aspects. By working hard together with newcomers, TMU will be confident to build itself into a top-level medical university.

The strategy will be implemented in 3 steps. Step 1: by 2020, TMU will become the leading medical university in China; step 2: by 2030, it will become one of the first-class universities in China; step 3: by 2050, it will achieve international prestige as a high-level medical research university.

### CONTACT US:

CONTACT: Xia Muqun

TEL: 022-83336739

Email: hr@tmu.edu.cn

ADD: No.22 Qixiangtai Road, Heping district, Tianjin 300070



## 400 Positions for Talents from Home and Abroad Start Your Career in Chengdu Southwest Jiaotong University, Chengdu, China

**Southwest Jiaotong University (SWJTU)**, founded in 1896, is one of the oldest institutions of higher education in China. As the most comprehensive leading university in transportation, SWJTU is world-renowned for pioneering the Chinese railway transportation engineering and industry, and for its leading contributions to the development of Chinese high-speed rail system. For its sustained excellence and prominence, the University is not only selected in the country's construction plan of world-class universities and first-class disciplines, but also placed among the key, elite multidisciplinary "211", "985", "2011 project (Rail transportation project superior discipline innovation platform)" and Tier-1 university equipped with graduate schools which is directly administered by the Chinese Ministry of Education. We offer comprehensive education and research programs in 26 faculties and institutes/centers, covering diverse disciplines in engineering, sciences, arts, and management ranging from undergraduate to doctoral degrees. The University owns 19 doctoral, 39 master, 75 undergraduate programs and 11 post-doctoral stations among the established 26 schools. Relying on its advantages and traditional disciplines, the university has set up 12 state-level science and technology innovation platforms and 36 provincial and ministerial-level scientific research bases, including the National Laboratory of Rail Transit (in process) and Traction Power State Key Laboratory. Thus, SWJTU has built the most complete discipline system, talent system and scientific research system in the world of rail transit.

Located strategically in Chengdu, the capital of Sichuan province—the dynamically growing West of China, SWJTU is blessed with rich heritage, unparalleled vibrancy, and a beautiful campus. Leveraging on these unique advantages and the University's strengths, SWJTU is vigorously implementing its strategic plans of "Developing and Strengthening SWJTU: Attracting and Cultivating Talents" and building world-class universities and first-class disciplines. We earnestly look forward to welcome excellent talents at home and abroad to join us, contributing to the University's continuing excellence.



### Qualified Candidates

Professors and research fellows from well-known universities or research institutes at home and abroad, associate professor (associate researcher), postdoctoral, and Ph.D. are welcome to apply.

### Remuneration and Benefits

Remuneration and benefits will be highly competitive, commensurate with qualifications and experience.

### Application Methods

Candidates should send personal detailed resumes, full texts of academic works, and citations from others, and evaluations by e-mail to the school contact email address and copy the university recruitment email: [talent@swjtu.cn](mailto:talent@swjtu.cn). The type of talent to be recruited should be indicated in the e-mail.

### Contact Information

#### University contact information

For the introduction of talents, please contact: Southwest Jiaotong University Talent Affairs Office of Organisation Department

**Contact:** Liu Jizong, Zhang Changling

**Tel:** +86-28-66366202

**Email:** [talent@swjtu.cn](mailto:talent@swjtu.cn)

#### Assistant Professor (Associate Researcher)

Consultation on related issues: Teacher Management Office, Human Resources Department, Southwest Jiaotong University

**Contact:** Shi Jianhong

**Tel:** +86-28-66366234

**Email:** [szb@swjtu.cn](mailto:szb@swjtu.cn)

**Website:** <http://www.swjtu.edu.cn/>

Southwest Jiaotong University Western Park of High-Tech Zone Chengdu, Sichuan Province, China Post Code: 611756



For More Information:



# Welcome to North University of China

## 1. About us

For details, please refer to: <http://www.nuc.edu.cn>

## 2. Beauty on the Campus

North University of China(NUC) is surrounded by mountains and waters, with beautiful scenery and four distinct seasons. The campus has few smog days and thus enjoys a reputation for "the blue sky of North University of China" in the whole province.



致知于行

太行精神  
坚定正确的政治方向  
艰苦奋斗的工作作风  
实事求是的科学态度  
不畏艰难的革命意志





### 3. The Concept of Talent Introduction

1) Improve the talent introduction system. For doctors, the “Implementation Measures for the Introduction of Doctors at North University of China” was carried out to introduce these talents according to the research background, subject category and academic level, providing outstanding doctors with a series of excellent treatments such as a research start-up fund of 500,000 yuan, a settling-in allowance of 400,000 yuan and an annual salary of 120,000 yuan.

For high-level talents, the “Implementation Measures of the ‘Taihang Scholars Program’ at North University of China” was introduced to provide a research start-up fund of 15 million yuan, a settling-in allowance of 6 million yuan, an annual salary of 2 million yuan and a set of house for academician-level talents, and successfully introduced Academician Li in July this year. In addition, the university provides a very attractive annual salary of 1.2 million yuan for other top talents, as well as a research start-up fund of 10 million yuan, a settling-in allowance of 3 million yuan and a set of house. The specific treatments are available on the website.

2) Broaden the channels for overseas talent introduction. In May of this year, the “First International Scholars Forum of North University of China” was successfully held, and

more than 30 overseas scholars were invited. The university promised to provide overseas doctors with 600,000 yuan for scientific research, 500,000 yuan for settling in, a 60-square-meter house and other treatments. Finally, 17 people reached the agreements of working at NUC.

3) Optimize the environment for talent growth. A green channel has been opened for the professional title evaluation, and thus excellent talents can participate in the evaluation of senior professional and technical positions. The university also helps the talents build platforms, organize teams, and promote cooperation. There is an affiliated hospital and an affiliated school in the university, as well as a sound research platform, teaching equipment, public service facilities and logistical support, which enable the faculty to keep their minds on work and free from worries.

Our university will continue to uphold the strategy of “Talents Strengthen Schools” and recruit talents as many as possible to create the bright future jointly.

#### Recruitment Details:

<http://rlzyglc.nuc.edu.cn/index.htm>

E-mail: [rsglk@nuc.edu.cn](mailto:rsglk@nuc.edu.cn)

Tel: +86-351-3924993

# Thousand Talents Plan Professorship for Young Scholars



**Beijing Institute of Technology (BIT)** announces recruitment for outstanding overseas young scholars/professionals in the broad areas of sciences and engineering. BIT, founded in 1940, has always been a leading institution of science and technology in China. In 2016-2017, BIT was ranked among the Top 400 in QS World Universities Ranking, as well as the 15th among the Chinese universities in the above rankings. The fundamental research on engineering, material science, chemistry, physics, computer, mathematics and social science in BIT is among the top 1% in ESI ranking.

## I. Qualifications and conditions:

1. Age Limitation: 40
2. A doctoral degree obtained from a world-renowned university is required, plus over 3 years' post-doctoral research experience;
3. The applicant is taking a formal teaching/research position in a well-known research institution or the research department of a well-known enterprise, when application is filed;
4. No limitations on Nationality.

## II. Treatments

1. The applicant will be employed as a professor and doctoral supervisor and given special quotas for the admission of postdoctoral, doctoral, and master candidates;

2. An initial research fund of RMB 2-6 million Yuan; office and laboratory arrangements;
3. Annual salary no less than RMB 420,000 Yuan (excluding insurance, public accumulation and occupational pension no less than 120,000 Yuan) + Annual bonus;
4. Preferential housing purchase opportunity and RMB 1 million Yuan as settling-in allowance, or RMB 2 million Yuan as settling-in allowance if applicant give up preferential housing purchase opportunity;
5. If the applicant is Chinese citizen, he or she may apply for permanent Beijing residence regardless of the location of former household registration; BIT will help solve life issues with family members, such as job arrangement for the spouse and school or nursery access for children;
6. Allowance for international round-trip travel expenses if the applicant needs to travel back to China to audition for the "Thousand Talents Plan Professorship for Young Scholars". If the applicant fails the audition, he or she may be transferred to the "Xu Teli Program for Young Scholars" with substantial benefits.

## Contact us

If you have any other questions or concerns, please don't hesitate to contact with us.

**Contact:** Ms. Xia  
**Tel:** +8610-68914243  
**E-mail:** bitrcb@bit.edu.cn

## BIT is recruiting International Teachers!

### I. Positions

1. Teaching & research position
2. Teaching position

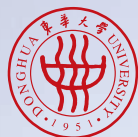
### II. Qualifications

1. Age Limitation: 65
2. Applicants for the teaching & research positions must hold a Ph.D. degree and the title of associate or assistant professorship or above.
3. Applicants for the teaching positions must hold a bachelor's degree or above and at least three years' teaching experience in related academic fields.

### III. Treatments

We implement a competitive salary and paid vacation.

**Contact:** Mr. Zhou  
**Tel:** +8610-68918577  
**E-mail:** bitzhaopin@bit.edu.cn



**東華大學**  
**DONGHUA UNIVERSITY**

## Donghua University Welcome Distinguished Scholars from Home and Abroad

### DHU Introduction

Donghua University, located in Shanghai, is one of the key universities under the direct administration of the Ministry of Education since 1960. It is a member of Project 211. Textile Science and Engineering is selected as world first-class discipline by the Ministry of Education in 2017.

Donghua University was founded in 1951 as East China Textile College. In 1985, it changed its name to China Textile University, and to its present name, Donghua University in 1999. It is one of the first universities accredited by the Ministry of Education for granting the doctor, master and bachelor degrees.

Currently Donghua University has developed into a distinctive multi-disciplinary university, with engineering as the predominant discipline alongside the coordinated development of engineering, science, management, and the liberal arts disciplines.

### Main Disciplines

- ◆ Textile Science and Engineering
- ◆ Materials science and Engineering
- ◆ Control Science and Engineering
- ◆ Environmental Science and Engineering
- ◆ Chemistry
- ◆ Management Science and Engineering
- ◆ Mechanical Engineering
- ◆ Design

### Recruitment Positions

- ◆ Donghua University Distinguished Research Fellow
  1. Under the age of 35 for researchers in natural science and engineering science, or under the age of 40 for researchers in humanities and social sciences.
  2. Applicant should get PhD degree and have post-doctor experience or obtained assistant professorship or above in prestigious overseas universities; or professors in domestic high-level universities or institutions.
- ◆ 1000 Plan Professorship for Young Talents
- ◆ 1000 Plan Professorship
- ◆ Chang Jiang Scholars Program

### Contact Information

**Email:** rcb@dhu.edu.cn **TEL:** +86-02167792043 More details available at <http://web.dhu.edu.cn/rcbdhu/>





## Zhengzhou University Is Recruiting Talent Globally

### I. Introduction to Zhengzhou University

Zhengzhou University, as one of the "211 Project" key universities in China, is included in the national scheme of 'Development of World-Class Universities and World-Class Disciplines' and co-developed by the Ministry of Education and Henan Provincial Government. We admit students nationwide (including Hong Kong, Macao and Taiwan) with over 54,000 full-time undergraduates, 19,000 postgraduates and over 2000 international students from more than 91 countries and regions currently studying with us.

There are 12 fields of study available at our university including Science, Engineering, Medicine, Literature, History, Philosophy, Law, Economics, Management Science, Pedagogy, Agriculture and Arts. Clinical Medicine, Material Science & Engineering and Chemistry are included as disciplines in the national scheme of 'Development of World-Class Universities and World-Class Disciplines'. Chemistry, Material Science, Clinical Medicine, Engineering, Pharmacology & Toxicology are ranked within the top 1% according to ESI.

We have ten national research platforms such as the National Engineering Research Center and the National Engineering Lab. With a total of over 5700 faculty members, there are 11 Academicians of the Chinese Academy of Sciences and the Chinese Academy of Engineering, two Fellows of the Chinese Academy of Social Sciences, three Academicians of overseas academies, eight Professors granted 'National Science Fund for Distinguished Young Scholars', two winners of the 'Chang Jiang Scholars Program' and seven professors included in the national 'Thousand Talents Plan'.

Medical education at Zhengzhou University started as early as 1928 at the Fifth National Sun Yat-sen University. In 1952, Henan Medical University was established as an independent Medical University that represents the beginning of medical education in Henan Province. The former Zhengzhou University was established in 1956 as the first comprehensive university after the founding of the People's Republic of China and was included as a national '211 Project' Key University. Zhengzhou University of Technology was founded in 1963 and this key university used to be under direct administration by the Ministry of Chemical Industry. In July 2000, the former Zhengzhou University, Zhengzhou University of Technology and Henan Medical University merged to become the current Zhengzhou University.

At a new historic starting point, Zhengzhou University has adopted the strategy of enhancing development through attracting high performers. At present, we have fostered a large team of talent led by Academicians and outstanding experts, with winners of the 'National Science Fund for Distinguished Young Scholars', 'Chang Jiang Scholars Program' and the 'Recruitment Program of Global Experts' as academic leaders, and young scientists with a doctoral degree as a backbone force. Zhengzhou University aims to be in the top tier in the world, which is in accordance with the social and economic development of the Central Plains, with the expectation of the hundreds of millions of Henan natives to enjoy quality higher education, with the trust of making this university better and stronger, and with the belief in the rising of Central China and the great

revival of the Chinese nation. Located in the Central Plains, we - by upholding the principle of truth-seeking - will assume the responsibility and strive to be recognized as a world-class, comprehensive, research-oriented university.

We sincerely invite outstanding scholars from home and abroad to join us to make Zhengzhou University a world leader.

### II. Positions

#### 1. Chief Scientist

Academicians of the Chinese Academy of Sciences or the Chinese Academy of Engineering; Fellow of the Chinese Academy of Social Sciences; Academicians of academies of developed countries; Nobel Laureates, Turing Award winners, Fields Medal winners or experts with equivalent academic achievements.

#### 2. Academic Leader

Experts included in the 'Thousand Talent Plan' and 'the Recruitment Program of Global Experts'; winners of 'Chang Jiang Scholars Program', 'National Science Fund for Distinguished Young Scholars' and 'National Special Support Program for High-level Personnel Recruitment'; Professors of world-renowned universities or experts with equivalent academic achievements

#### 3. Backbone Scientist

Experts included in the young scholar program of 'Thousand Talent Plan'; winners of 'Outstanding Youth Science Foundation' and young scholar program of 'Chang Jiang Scholars Program'; Associate Professors of world-renowned universities or experts with equivalent academic achievements

#### 4. Outstanding Young Talent

Young scientists with great potential to become leaders in relevant fields with outstanding achievements; under 35 years of age in principle.

### III. Salary Package

*Our university will provide excellent living and working facilities*

- 1) Competitive salaries
- 2) Start-up fund and funding for platform construction
- 3) Settlement allowance and special treatment for university housing
- 4) The employment of spouse and schooling of children
- 5) Allowance provided by Henan provincial government and Zhengzhou municipal government as rewards or for research funding and housing.

### Contact Information

Mr. Yang

E-mail: [rcb@zzu.edu.cn](mailto:rcb@zzu.edu.cn)

Phone: + 86-371-67781087

Fax: + 86-371-67739397

Web: <http://www5.zzu.edu.cn/rcb/>





Career Feature:

## Artificial Intelligence

Issue date: November 30

Book ad by November 15

Ads accepted until November 21 if space allows



# 129,562

subscribers in print  
every week

# 503,472

monthly unique browsers  
on ScienceCareers.org

# 56 %

of our weekly readers  
are Ph.D.s

Artificial Intelligence (AI) is impacting science in new and exciting ways as scientists are using it to better understand society to find solutions to problems across diverse disciplines. This feature will give an overview of AI, and explore the hotspots/centers of excellence and applications for AI. Typical career paths for those working in AI will be explored as well as the opportunities that exist for careers in AI.

Your organization can brand itself as a leader in AI by raising your visibility alongside relevant content while attracting potential candidates. Contact us for further details.

To book your ad:  
[advertise@sciencecareers.org](mailto:advertise@sciencecareers.org)

#### The Americas

+ 202 326 6577

#### Europe

+44 (0) 1223 326527

#### Japan

+81 3 6459 4174

#### China/Korea/Singapore/ Taiwan

+86 131 4114 0012

Produced by the Science/AAAS  
Custom Publishing Office.

### What makes Science the best choice for recruiting?

- Read and respected by 400,000 readers around the globe
- Your ad dollars support AAAS and its programs, which strengthens the global scientific community.

### Why choose this AI Feature for your advertisement?

- Relevant ads lead off the career section with a special "AI" banner.

### Expand your exposure by posting your print ad online:

- Link on the job board homepage directly to AI jobs
- Dedicated landing page for AI positions.



**ScienceCareers**  
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

FOR RECRUITMENT IN SCIENCE, THERE'S ONLY ONE SCIENCE.



## **FACULTY POSITION ANNOUNCEMENT Assistant/Associate Professor Of Quantitative Fisheries Ecologist**

The University of California at Davis is pleased to announce the recruitment for a tenure-track faculty position in quantitative fisheries ecology. The successful candidate will join the Department of Wildlife, Fish, and Conservation Biology in the College of Agricultural and Environmental Sciences at the rank of Assistant or Associate Professor; early career candidates are especially encouraged to apply. Criteria for appointment include: a Ph.D. or equivalent in fisheries science, ecology and evolutionary biology, statistics and applied mathematics, or a related field, a record of excellence in scholarly research, and demonstrable potential to establish a competitively-funded research program. The appointee will be responsible for teaching undergraduate and graduate courses in fisheries and marine science and management and/or quantitative methods, be actively involved in undergraduate advising, curricular development and department and university service. The appointee is also expected to guide and mentor graduate students.

Applicants should submit materials via the following website:  
**<http://apptrkr.com/1295757>**.

Additional inquiries can be directed to Andrew Rypel, Department of Wildlife, Fish, and Conservation Biology ([rypel@ucdavis.edu](mailto:rypel@ucdavis.edu)) or Marissa Baskett, Department of Environmental Science and Policy ([mlbaskett@ucdavis.edu](mailto:mlbaskett@ucdavis.edu)). The position will remain open until filled but to ensure consideration, applications should be received by November 1, 2018.

UC Davis is an affirmative action/equal employment opportunity employer and is dedicated to recruiting a diverse faculty community. We welcome all qualified applicants to apply, including women, minorities, veterans, and individuals with disabilities.

*The University of California is committed to creating and maintaining a community dedicated to the advancement, application, and transmission of knowledge and creative endeavors through academic excellence, where all individuals who participate in University programs and activities can work and learn together in a safe and secure environment, free of violence, harassment, discrimination, exploitation, or intimidation. With this commitment, UC Davis conducts a reference check on all finalists for tenured positions. The reference check involves contacting the administration of the applicant's previous institution(s) to ask whether there have been substantiated findings of misconduct that would violate the University's Faculty Code of Conduct. To implement this process, UC Davis requires all applicants for any open search for assistant/associate/full professor to complete, sign, and upload the form entitled "Authorization to Release Information" into RECRUIT as part of their application. If an applicant does not include the signed authorization with the application materials, the application will be considered incomplete, and as with any incomplete application, will not receive further consideration. Although all applicants for faculty recruitments must complete the entire application, only finalists considered for positions with tenure or security of employment will be subject to reference checks.*



## **Institut Curie is recruiting a Director for its Research Center**

The Research Center of Institut Curie in France has 1,200 employees, mainly located in the center of Paris and in Orsay, working in four thematic domains around biology, chemistry, physics and bioinformatics. It is organized into 12 mixed research units (with CNRS, Inserm and Universities), bringing together 89 research teams, one translational research department and 16 cutting edge technological platforms. The consolidated annual budget of the Research Center is about 100 M€.

### **Main responsibilities of the Director of the Research Center**

He/she supervises all the research and support activities of the Research Center (CDR), manages the CDR's scientific strategy, represents the CDR externally, and implements the MC21 project, including the medico-scientific project which he/she co-directs with the director of the Hospital. In addition to his/her institutional functions, the CDR director will have the opportunity to develop his/her own research.

### **Desired profiles and application methods**

The candidate must be an outstanding scientist in Life sciences with international recognition, an excellent track record in fundamental research and a proven interest in interdisciplinary approaches and biomedical applications. The applications received will be examined by an international ad hoc committee, by the international scientific board and by the scientific commission of Institut Curie.

Please send your application by email to the following address: **[thierry.philip@curie.fr](mailto:thierry.philip@curie.fr)**. These applications will include a full CV and a cover letter for the position. They will be treated confidentially. Deadline for reception of applications is **10th November 2018**.



**Tenure-Track Assistant Professorship in Physical Chemistry**  
**Department of Chemistry**  
**College of Science, Purdue University**

The Department of Chemistry in the College of Science at Purdue University invites applications for a tenure-track Assistant Professorship in **Experimental Physical Chemistry/Chemical Physics**.

Part of a large-scale interdisciplinary hiring effort across key strategic areas in the College of Science—Purdue's second-largest college, comprising the physical, life, and computing sciences—this position comes at a time when the College is under new leadership and with multiple commitments of significant investment.

Successful candidates will lead an experimental physical chemistry/chemical physics research program focused on advancing chemical innovation by performing measurements that address research questions of fundamental importance. Purdue has an outstanding tradition in physical chemistry, and the department is looking to integrate a creative scientist into the cutting-edge interdisciplinary environment provided by Purdue University. With 50 full-time faculty, 350 PhD students, and over 300 outstanding undergraduates, Chemistry at Purdue is one of the largest, most diverse, and most highly ranked departments in the country. The wide-ranging expertise in the Department enables fast and effective responses to interdisciplinary research opportunities, positioning the department as a key partner in many university-wide centers, institutes and initiatives, including an emerging College of Science program in quantum information science. Further, excellent facilities, including our unique research infrastructure in the Jonathan Amy Facility for Chemical Instrumentation (JAFICI), allow for on-site design and fabrication of next-generation instrumentation.

Candidates must have a PhD in physical chemistry, chemical physics, or related fields, with outstanding credentials in experimental research, an exceptional track record of publications and a strong commitment to excellence in teaching. Successful candidates are expected to develop a vibrant research program supported by extramural funding and teach physical chemistry courses at the undergraduate and/or graduate level. Applicants should submit a letter of application with curriculum vita, a summary of planned research, and a statement on teaching philosophy electronically at: [www.chem.purdue.edu/facultyopenings](http://www.chem.purdue.edu/facultyopenings). Applicants should also arrange for three letters of recommendation to be uploaded. Applications will be reviewed beginning **October 15, 2018**, and will remain in consideration until the position is filled.

Purdue University's Department of Chemistry is committed to advancing diversity in all areas of faculty effort, including scholarship, instruction, and engagement. Candidates should address at least one of these areas in their cover letter, indicating their past experiences, current interests or activities, and/or future goals to promote a climate that values diversity and inclusion.

A background check will be required for employment in this position. Purdue University is an ADVANCE institution.

*Purdue University is an EOE/AA Employer. All individuals, including minorities, women, individuals with disabilities, and veterans are encouraged to apply.*



**Tenure-Track Assistant Professorship in Biophysical Chemistry**  
**Departments of Chemistry and Physics & Astronomy**  
**College of Science, Purdue University**

The Departments of Chemistry and Physics & Astronomy in the College of Science at Purdue University invite applications for a tenure-track Assistant Professorship in **Biophysical Chemistry**. Successful candidates will be considered for a joint appointment in Chemistry (nominally 75%) and Physics (25%), and will be offered affiliation with Purdue's new Institute of Inflammation, Immunology and Infectious Disease (PI4D), established as part of the University's strategic investment in the life sciences.

Part of a large-scale interdisciplinary hiring effort across key strategic areas in the College of Science—Purdue's second-largest college, comprising the life, physical, and computing sciences—this position comes at a time when the College is under new leadership and with multiple commitments of significant investment.

Successful candidates will have research interests in biophysical chemistry as related to human health-relevant biological processes, especially those related to inflammation, immunology and infectious disease or similar fields. Ideal candidates will have appropriate training in both physical and biological chemistry and would use modern biophysical methods and integrative biochemical methods to study biological macromolecules. Chemistry and Physics at Purdue together have a strong tradition of excellence in biophysical chemistry and biochemistry, and are seeking to enhance this preeminence through strategic hiring of a creative scientist to be part of the cutting-edge interdisciplinary environment provided by Purdue University.

Candidates must have a PhD in biophysical chemistry, biophysics, physical chemistry, or a related field with outstanding credentials in biophysical chemistry research, an excellent track record of or potential for leading publications and a strong commitment to excellence in teaching. Successful candidates are expected to develop a vibrant research program supported by extramural funding and teach physical chemistry and biophysical chemistry courses at the undergraduate and/or graduate level. Applicants should submit a letter of application with curriculum vita, a summary of planned research, and a statement on teaching philosophy electronically at: [www.chem.purdue.edu/facultyopenings](http://www.chem.purdue.edu/facultyopenings). Applicants should also arrange for three letters of recommendation to be uploaded. Applications will be reviewed beginning **October 15, 2018**, and will remain in consideration until the position is filled.

Purdue University's Department of Chemistry and Department of Physics are committed to advancing diversity in all areas of faculty effort, including scholarship, instruction, and engagement. Candidates should address at least one of these areas in their cover letter, indicating their past experiences, current interests or activities, and/or future goals to promote a climate that values diversity and inclusion.

A background check will be required for employment in this position. Purdue University is an ADVANCE institution.

*Purdue University is an EOE/AA Employer. All individuals, including minorities, women, individuals with disabilities, and veterans are encouraged to apply.*



By Kate Furnell

# Breaking the silence

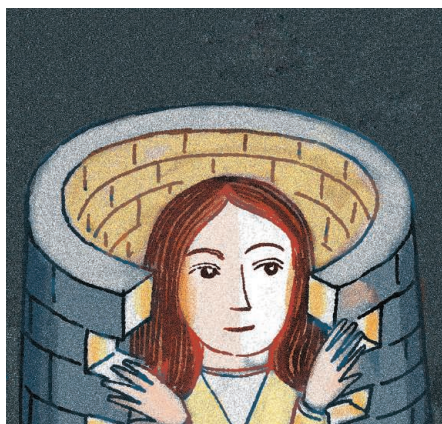
**M**y hands were shaking. My stomach began to churn as I looked out at the faces before me. I fought to calm myself. Facing a packed conference room of more than 150 people—including my doctoral supervisor, my collaborators, and world-leading researchers—I was about to let slip the secret I had worked desperately for years to keep from all but a few trusted colleagues: that I seriously struggle with mental illness.

I had decided to give this talk because I was tired of keeping this secret. But it wasn't just that. I also knew of many other graduate students who were going through similar challenges. For all of us, the stigma around mental health coupled with the intense competition of academia made it difficult to talk about our struggles with people who could help, or who simply ought to know.

A close friend had lost her way in her project and was mentally exhausted. Yet she was afraid that if she opened up to her supervisor and sought his guidance, he wouldn't understand or believe her problems and would start to treat her poorly. Another friend struggled with anxiety, which made her uncomfortable in the crowded office she shared with a dozen other Ph.D. students. But she could not bring herself to ask her department head for accommodations because she feared that doing so could have negative repercussions. Yet another person was in the midst of a breakdown because she wasn't able to get thesis feedback from her supervisor.

These are just three people suffering because of this culture of silence. There are many more.

I knew that I couldn't deliver a talk about mental health without including some of my personal story: I have borderline personality disorder, and I have also been diagnosed with depression, anxiety, and post-traumatic stress disorder. I worried that if I talked about my experiences with these conditions, some of my colleagues might not understand. I thought they would become uncomfortable around me or would stop taking my work seriously. I was also concerned that potential future employers might not be open to hiring someone such as myself, whose condition means that my moods can swing between apathy and despair in a matter of hours. But I wanted to start a conversation that was so desperately needed, and speaking publicly at a conference felt like the best possible chance to spark the discussion.



***“The stigma around mental health ... made it difficult to talk about.”***

Instead, I was met with hugs from my collaborators and colleagues, applause, and warm words. Many people came forward to share their own experiences. Many more thanked me or passed on testimony from others. People from my place of work told me how moved they were by my honesty. A number of people asked whether I'd be willing to deliver my talk again.

During my Ph.D. work, I have given talks across Europe and in South Africa, but I have never felt as proud of any of them as I did this one in my hometown of Liverpool, U.K. It is as significant an entry on my CV as my research work. Getting people to listen to and understand the perspectives of people struggling with mental illness is important because we are all far from infallible. Mental illness can strike anyone at any time. By discussing our experiences openly, we can reduce the isolation, shame, and resistance to seeking help that can come with it. We are all human, and that's OK. ■

*Kate Furnell is a doctoral candidate at Liverpool John Moores University in the United Kingdom. Send your career story to [SciCareerEditor@aaas.org](mailto:SciCareerEditor@aaas.org).*